ECONTECHMOD. AN INTERNATIONAL QUARTERLY JOURNAL – 2016. Vol. 5. No. 1. 35–41

The technology of hierarchical agglomerative cluster analysis in library research

N. Kunenets

Informative systems and networks department
Lviv Polytechnic National University
S. Bandery str., 12, Lviv, 79013, Ukraine;
e-mail: nek.lviv@gmail.com

Received November 16.2015: accepted January 22.2016

Abstract. This article describes the simple technology of the hierarchical agglomerative cluster analysis of 20 different libraries, presented by the samples of classification attributes of the same volumes. It is necessary to construct a proximity matrix for an effective process of cluster analysis, using the data from the table "library-classification features". For separating a set of selected objects into clusters, so that each of them has objects, the most appropriate for its type, it is necessary to create a table "object-property", where libraries are objects and individual and equal dimension vectors (sets) of characteristic classification features are properties. To do this you should: form the set of libraries that are the objects of clustering; define for each library the set of classification features and its power (volume) in the same nominal scale; choose the value scale of classification features; form a table object-property. This technology is implemented in the environment of MsExcel-2003. It includes the transformation of one-dimensional data into multi-dimensional indexes, using the descriptive statistics and distributions of individual parameters, the creation of the "object-property" table, the building of the proximity matrix, the definition of the dendrogram structure and the cluster interpretation. The application of clustering method can further focus on creating algorithms of effective information search, and also building the scientifically-reasonable classification systems orientated on the library science. The method of hierarchical agglomerative cluster analysis can be used with typological or semantic distribution of library funds, or studying of their thematic and specific composition. This method can be considered as universal, that gives an opportunity to formalize the typology division of any objects of librarianship.

Key words: the cluster analysis , the object-property table, the proximity matrix, the dendrogram, clusters.

INTRODUCTION

An important role in the realization of scientific research plays the optimization of methods which are used as basic tools, the determination of possible sphere of the use of every method and the selection of most effective one in each case. The methodology of scientific activity is a process, that foresees the application of integral totality of certain ways, approaches, methods, actions, aimed to the obtaining of new scientific results, the achievement of the imposed aim and the implementation of the pre-arranged tasks. A logical construction of research work of the library specialists must base on the interdisciplinary approach of library studying in the context of social and communicative processes.

A library is an integrative social institute, as well as the library science must be diverse, that caused a choice of research methods not only from the arsenal of librarianship, but also from information technologies. And it is fully naturally, as nowadays a methodology of scientific research requires new paradigms and conceptual principles along with the use of time-tested traditional methods and approaches.

It would be impossible to have the deep modern theoretical developments without the mastering of historical experience. The working with the considerable array of works of leading scientists, both library specialists and researchers from contiguous fields of sciences, allowed to formulate a number of characteristic features of libraries with the aim to realize their typification, and more exactly carry out their division into homogeneous classes in this way, using the method of hierarchical agglomerative cluster division.

THE ANALYSIS OF RECENT RESEARCHES AND PUBLICFTIONS

The substateliness of such research is confirmed that in early 70th of the XX century the American scientist J. Solton suggested using the method of cluster analysis for distribution of library fund [1]. The method used in psychology [2,3, 4,5], for ecological data [6], bioinformatics [7], biology [8], Statistics [9], sociology [10], economy [11, 12, 13, 14, 15], the analysis results search [16, 17, 18, 19]. The tool of cluster analysis during the library research did not find implementation in Ukrainian librarianship.

OBJECTIVES

The aim of this article is to analyse the possibilities of implementation of hierarchical cluster analysis method in library science, particularly in the tasks of comprehensive research of libraries and to show its realization by means of Ms Excel in order to divide the set of libraries into homogeneous groups on the principle of similarity based on their classification characteristics given with a table “object-property”.

THE MAIN RESULTS OF THE RESERCH

The basic nature of cluster analysis

The first application as a method the cluster analysis found in the field of sociology. The subject of cluster analysis was defined firstly in 1939 and its description was given by a researcher R. Trion [20]. A cluster analysis is a statistical procedure[21], that executes multidimensional classification of data containing information about the selection of objects, that organizes the objects in a relatively homogeneous groups. Thus, the task of data classification is solved with the use of formed mathematical device.

The essence of this procedure is that the objects that must be classified are given by the same vector (by a set) of individual characteristics of these objects in the form of a table “object-property”, on the basis of which the matrix of distances (similarities) is constructed, after that the clustering is exercised. The content of clustering is that objects that are coming together can form a separate group – cluster [22]. The main criterion for classifying objects to a particular group, in other words the realization of clustering is definitely given and measured “distance” between objects.

The methods of cluster analysis [23] can be applied in different cases, including, when the question is only about a simple groupment where everything is reduced to the formation of groups for quantitative feature similarity. Depending on the certain applied task, the aim of cluster analysis can be different, for example:

- to understand the structure of the set of objects, breaking them into homogeneous groups, in any sense, and at the same time, to simplify further data processing for the decision making, working with every cluster separately;
- to distinguish unusual objects that do not belong to any of these clusters. This task is named a single class classification of exposure of non-typicality or novelty;
- to decrease, in the case of very large scale samples X, the data volume for maintenance, leaving one of the most characteristic representative from each cluster;
- to investigate the object dynamics in the process of their operation by changing the distances between classes and within classes.

The methods of cluster analysis can be applied when it is necessary to carry out the usual division of object set into groups only with a quantitative similarity of the selected attributes.

The cluster analysis algorithms are developed by software implementation that allows effectively to solve the problems of large dimension, but such software is licensed, and thus required some payment and is a little accessible. Therefore, its realization among Ms Excel is a very actual task.

The method of hierarchical agglomerative cluster analysis (is sometimes called numeral taxonomy) of existent data carries out the classification that did not exist before, or looking over data again creates new one, ignoring previous.

The result of realization of cluster analysis is given by a special chart of dendrogram that in a natural scale represents the closeness or similarity between objects that are classified.

It is possible to distinguish different on a volume, clusters on the basis of the dendrogram, thus distances between objects in a cluster will always be less than the least distance between objects from two different clusters.

In this research with the use of method of hierarchical agglomerative cluster analysis it is conducted the classification of libraries on selected features and criteria. The main purpose of using the cluster analysis in our research is to confirm the correctness and accuracy of the classical approach to distribution of libraries by their types. The main feature of these types is that libraries as a result of cluster analysis belong to the same cluster, on certain features are more similar to each other than with the objects attributed to the other clusters.

For a cluster analysis, collected data in the process of research or obtained experimentally is given in the form table “object-property”, in a way that the first column has the names of objects to be grouping, and the other columns correspond to certain properties (attributes, descriptions, indexes, etc.) and contain their certain values.

Its task is, that on the basis of the data provided by a table “object-property”, to break a set:

$$G=\{ g_i : g \in G,\ i=1, 2, \mathbf{K}, m\}, \quad (1)$$

of these objects to k (k – integer) clusters – subsets that are not crossed $Q_1, Q_2, \mathbf{K}, Q_k$ so that their association answer the whole set:

$$\bigcup_{j=1}^{k} Q_j = G, \quad (2)$$

and their crossing will be an empty set:

$$\bigcap_{j=1}^{k} Q_j = \varnothing. \quad (3)$$

Thus:

- every object g_i must belong to one and only to one subset of partition,
- objects that belong to one and same cluster must be similar,
- objects that belong to the different clusters must be different.

According to the received information the dendrogram should be built and its interpretation regarding the choice of a cluster number should be shown.

The implementation of methodology of this kind of cluster analysis is performed directly in the tabular processor Ms Excel without involving any additional subroutines and libraries for typological division should involve the following steps.

The formation of the table “object-property”

For separating a set of selected objects into clusters, so that each of them has objects, the most appropriate for its type, it is necessary to create a table “object-property”, where libraries are objects and individual and equal dimension vectors (sets) of characteristic classification features are properties. To do this you should:

- form the set of libraries that are the objects of clustering,
- define for each library the set of classification features and its power (volume) in the same nominal scale,
- choose the value scale of classification features,
- form a table object-property.

On the basis of classic criteria of typological division, formed set of objects-libraries, with the strong typological characteristics is given in Table 1.

The library average is implemented with a number of classifications. This cluster analysis does not impose any restrictions on the type of objects, but each object has to be submitted by the same set of features. The formation of the feature set of libraries is one of the most important objectives of the study.

Table 1. The set of selected libraries

№	Name of library	Note
1	Pochaiv lavra	
2	Stauropegion brotherhood	
3	Scientific society named by T. Shevchenko	
4	Prosvita	Including reading rooms
5	Scientific theological society	
6	The union of Ukrainians	
7	Native school	Including school
8.	National museum	
9	The museum of Didushytskiy	
10	Studion	
11	National house	
12	Stauropegion institute	
13	Institution of Ossolinski	
14	Municipal public	
15	Military	
16	Archive	
17	Lviv university	Including departments
18	Lviv Polytechnic	Including departments
19	Agrarian Academy	Including departments
20	Theological Academy	Including departments

The implementation of this step is to determine the feature set that reflects the peculiarities of libraries in the best and the most complete way. As a result of selection on the basis of the analysis and processing of available primary information material, this set is the set of features:

$$X=\{ x_1, x_2, \mathbf{K}, x_m\}, \quad (4)$$

where: $m=8$. In the Table 2, there are denominations of classification features, their characteristics and the individual value scales of quantitative assessment classifications.

Selected features of objects for clustering are given ith quantitative values, obtained from the expert assessment, taking into account specified scales that are in the table. Their values are given in points. Scores were determined based on the analysis of sources on the library history [24].

Two of the following attributes, namely “Origin of funds” and “Acquisition criteria” are presented in binary terms, the rest of them with defined and established scale on the basis of research.

Assessment values of attributes selected by our libraries are presented in Table 3 in the form of a table “object-property”, more concrete in the table “library-classification features”.

Table 2. Scales of library attributes

№	Attribute Name	Charac-teristic	Scale	Note
1	Motivation of foundation – x_1	Number	10	Number of reasons (justifica-tions)
2	Functional tasks – x_2	Number	10	
3	Origin of funds – x_3	Number	2	At their own expense 0 or action outside 1
4	Rooms for funds – x_4	Number	8	Geogra-phically separated rooms
5	Acquisition criteria – x_5	Number	2	Available 1 or no 0
6	Theme content – x_6	Number	12	Type of content
7	Communi-cation with persona-lities – x_7	Number	10	
8	Reader cate-gories – x_8	Number	10	

The construction of proximity matrix

The similarity or the difference between the objects, which are classified, is determined depending on the metric distance between them. If each object is described by “k” characteristics, it can be submitted as a point of k-dimensional space, and the similarity with the other objects will be defined as an appropriate distance, which is calculated by these “k” characteristics. The quantitative evaluation of similarity is based on the concept of metrics.

It is necessary to construct a proximity matrix for an effective process of cluster analysis, using the data from the table 3 “library-classification features”.

A proximity matrix is constructed using the appropriate metrics, which define the distance between the objects, using the value of their features set. This procedure needs to choose the appropriate metrics to calculate the distances of closeness between each of the libraries.

One of the most known metrics of cluster analysis is Euclidean distance [25], as it corresponds to intuitive notions about proximity and describes the classical statistical designs by own quadratic form.

However, the Euclidean distance uses the quantitative values of features or preliminary fixing of their values. Therefore, the Manhattan metric is used for the construction of the proximity matrix in this study that is the best up to a procedure of determination of distances between the objects, which are lodged by the point scale.

Table 3. “Library-classification features”

№	Name of the library	Non names of classifications							
		1.	**2.**	**3.**	**4.**	**5.**	**6.**	**7.**	**8.**
1	Monasteries	6	5	0	3	0	4	2	2
2	Stauropegion brotherhood	3	6	0	4	0	3	2	3
3	Scientific society named by T. Shevchenko	3	4	1	2	1	7	5	4
4	Prosvita	5	4	0	1	1	6	8	4
5	Scientific theological society	4	4	0	3	1	5	5	3
6	The union of Ukrainians	4	6	1	2	0	4	9	4
7	Native school	7	9	1	5	1	10	6	6
8	National museum	6	5	1	4	1	6	7	7
9	The museum of Didushytskiy	8	4	0	3	1	7	6	5
10	Studion	3	3	0	2	0	4	4	2
11	National house	9	6	1	3	1	6	5	3
12	Stauropegion of institute	4	2	0	2	0	6	3	3
13	Institution of Ossolinski	5	4	0	6	1	11	7	8
14	Municipal public	10	7	1	8	1	12	5	8
15	Military	4	6	1	4	1	10	3	3
16	Archive	3	8	1	6	1	9	2	5
17	Lviv university	10	10	1	8	1	12	8	5
18	Lviv Polytechnic	10	10	1	8	1	7	5	5
19	Agrarian Academy	7	5	1	2	1	3	2	5
20	Theological Academy	5	3	1	1	1	2	3	2

The distance (metric) between the objects in a space of parameters is assumed as d_{ab} value, which meets the requirements of following axioms:

$$\text{A1. } d_{ab} > 0,\ d_{ab} = 0, \tag{5}$$

$$\text{A2. } d_{ab} = d_{ba}\ , \tag{6}$$

$$\text{A3. } d_{ab} + d_{bc} \geq d_{ac}. \tag{7}$$

The Manhattan metric is chosen for matrix construction in our study, that corresponds to the Euclidean metric of the order one. The Manhattan metric meets these axioms in accordance with a physical meaning of objects, which are classifying. Thus, the distance d_{ij} between i and j objects is defined as the sum of the absolute values of differences between the two features of these objects. Analytically it looks like:

$$d_{ij} = \left|x_{1i} - x_{1j}\right| + \left|x_{2i} - x_{2j}\right| + \ldots$$

$$+ \left|x_{8i} - x_{8j}\right| = \sum_{k=1}^{8} \left|x_{ki} - x_{kj}\right| . \quad (8)$$

We get the proximity matrix in the table 4 using the computing capabilities of Ms Excel, namely the tabulation functions by AutoComplete and fixing the column name and the line number, with use of simultaneous reading of classification features from the table "library-classification features" and its copy for $N = 20$ iterations(Table 4).

This matrix is symmetric to diagonal, where the values are zero. The equality of zero means the distance between the same object, which of course is zero.

The construction of dendrogram

The final procedure of agglomerative hierarchical cluster analysis is the determining of the dendrogram parameters, its construction and interpretation on choosing clusters and their numbers. A dendrogram based on the proximity table and the chosen strategy of association, which consists of the combining of objects from the table in group-clusters. The hierarchical agglomerative cluster analysis uses the following strategies for work with proximity matrix. They are the nearest neighbour, the distant neighbour, the group average and so on.

The initial steps in the procedures of hierarchical agglomerative cluster analysis are identical. First of all, looking for a pair of the objects with the least distance among all objects of proximity matrix and combining them into one group. The columns and rows of these objects are eliminated, and a new column and row with listed values of the features are inserted in their place, so as not to disrupt the diagonal of zeros. As a result, the proximity matrix size decreases by one and the smallest value becomes a parameter of dendrogram because it determines the distance between objects, and indicates the group number $n+1$. Each next step has a merge between two objects or between an object and a group or between the two groups for which the degree of proximity is minimum, then it's carried a similar recalculation and the merged groups are denoted as:

$$n+2,\ n+3,\ \ldots,\ n+(n-1) . \quad (9)$$

The procedure is completed when the dimension of the proximity matrix is 2×2. The mathematical basis for the merging in the groups is expression of Williams-Lance [21]:

$$d_{nk} = a_i \cdot d_{hi} + a_j \cdot d_{hj} + b \cdot d_{ij} + g \cdot \left|d_{hi} - dhj\right|, \quad (10)$$

where: a_i, a_j, b, g – are determined by the type of strategy, d_{hi} and d_{hj} columns, which are combined, d_{hk} – the column as a result of the merging.

This study uses a flexible strategy, which is applied to any measures of proximity and determined by the following parameters $\alpha_i = \alpha_j = 0.625$, $\beta = -0.25$ i $\gamma = 0$. The use combining for this strategy is implemented by the following algorithm:

$$d_{rs} = 0{,}625 \cdot \left(x_{rs} + x_{(r+1)s}\right) - 0{,}25\left(\min\left(x_{r(s-1)}\right)\right), \quad (11)$$

where: r – column number, $r = \overline{1, n}$, and s – line number, $s = \overline{1, n}$.

The combining strategy is that we find the proximity matrices with the least value d_{rs}, which is equal to 6. It gives an opportunity to merge the objects 3 and 5 in one group and to appropriate a number 21 to this group, then it is recounted column and row by the expression (2) for this group. The recount comes during all work of algorithm for every new group according to the value of groups, which are merging, so it acquires a new value.

Then we look for the smallest value again – it corresponds to the distance between the groups 10 and 12 and is equal to 6. Similarly, we make a recount of features values and mark the merging by number 22, and so on. The results of procedure are given in Table 5.

Table 5. Step merging of clusters

Value of dendrogram		
Merging of groups	Distances between groups, which are merging	Newly formed group
d 3-5 =	6	21
d 10-12 =	6	22
d 1-2 =	7	23
d 17-18 =	8	24
d 21-4 =	8,5	25
d 8-9 =	9	26
d 15-16 =	9	27
d 26-11 =	9,63	28
d 22-20 =	9,75	29
d 23-19 =	10,13	30
d 14-24 =	10,5	31
d 25-6 =	11,16	32
d 7-13 =	13	33
d 30-29 =	13,89	34
d 32-28 =	15,43	35
d 33-27 =	16,59	36
d 34-35 =	22,51	37
d 36-31 =	24,94	38
d 37-38 =	54,67	39

Unfortunately, the graphic providing of Microsoft Excel does not have the opportunity directly to build a dendrogram, as her construction conjugates not only with the scale of distances but also with the place of objects, which belong to the certain clusters now. That's why, the first step of a construction is manual, and the process of construction begins from the root of tree. For example, the number 39 indicates the merging of two groups 37 and 38 that mean we can build in any scale two rectangular branches. In turn, the number 37 is a group of two primary objects 34 and 35. Instead, the number 38 combines groups 36 and groups 31, moreover group 31 is a merging of 14th group and 24th group. Similarly we analyze the rest of the group. The procedure is repeated to complete full construction of the tree, until we reach the most non-merged objects. After the construction of the tree, it is clear the hierarchical order of objects clusters, which means: the higher level (bigger distance between clusters – vertical axis), the more objects are included by clusters. The next is resulting of vertical segments in accordance with the scale of the distances on the vertical axis. As a result, the dendrogram has the form shown in Figure 1.

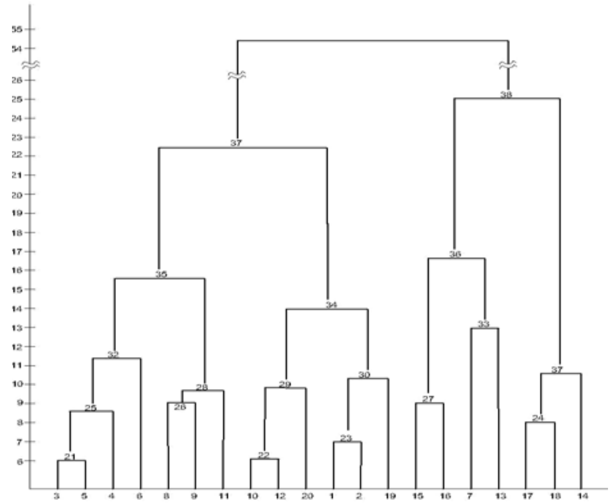

Fig. 1. Dendrogram of library clustering

Table 4. Proximity matrix

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII	XIX	XX
I	0	7	15	15	9	14	23	15	14	8	12	10	24	30	14	19	33	25	8	10
II	7	0	14	18	10	13	24	18	19	9	15	11	25	31	11	14	34	26	11	13
III	15	14	0	8	6	11	18	12	9	9	11	9	17	25	11	14	28	20	13	13
IV	15	18	8	0	8	9	20	10	9	13	13	11	15	29	17	22	26	24	15	13
V	9	10	6	8	0	11	20	12	9	7	9	7	17	27	11	18	30	22	13	11
VI	14	13	11	9	11	0	21	13	16	12	14	14	22	30	16	21	27	25	14	16
VII	23	24	18	20	20	21	0	12	13	27	15	25	13	13	13	12	12	12	19	27
VIII	15	18	12	10	12	13	12	0	9	19	11	17	11	19	15	18	22	18	13	19
IX	14	19	9	9	9	16	13	9	0	16	8	14	14	20	16	19	21	15	12	18
X	8	9	9	13	7	12	27	19	16	0	16	6	24	34	16	21	37	29	14	8
XI	12	15	11	13	9	14	15	11	8	16	0	14	22	18	12	19	21	13	12	16
XII	10	11	9	11	7	14	25	17	14	6	14	0	22	32	12	19	35	27	14	10
XIII	24	25	17	15	17	22	13	11	14	24	22	22	0	14	16	17	19	23	24	26
XIV	30	31	25	29	27	30	13	19	20	34	18	32	14	0	20	19	9	11	26	34
XV	14	11	11	17	11	16	13	15	16	16	12	12	16	20	0	9	23	21	16	16
XVI	19	14	14	22	18	21	12	18	19	21	19	19	17	19	9	0	20	16	17	23
XVII	33	34	28	26	30	27	12	22	21	37	21	35	19	9	23	20	0	8	29	37
XVIII	25	26	20	24	22	25	12	18	15	29	13	27	23	11	21	16	8	0	21	29
XIX	8	11	13	15	13	14	19	13	12	14	12	14	24	26	16	17	29	21	0	10
XX	10	13	13	13	11	16	27	19	18	8	16	10	26	34	16	23	37	29	10	0

CONCLUSIONS

It was conducted in the “manual mode” the hierarchical agglomerative cluster analysis only for the small selection of libraries. In case of their large number (thousands and tens of thousands of elements) manually to realize such technology is impractical and impossible. Appropriate software must be used for this purpose. The deep analysis of clustering results of plenty of libraries can serve as a basis for the construction of relevant ontology in the field of library science.

The application of clustering method can further focus on creating algorithms of effective information search, and also building the scientifically-reasonable classification systems orientated on the library science. The method of hierarchical agglomerative cluster analysis can be used with typological or semantic distribution of library funds, or studying of their thematic and specific composition. This method can be considered as universal, that gives an opportunity to formalize the typology division of any objects of librarianship.

The received results of cluster analysis do not only confirm the classical classification scheme of the investigated libraries, they simplify the procedure of division and also can be used for further theoretical developments of classification schemes of libraries.

REFERENCES

1. **Solton D. 1979.** Dynamic library – information systems. Moscow. 9, 35, 369 (in Russian).
2. **Odilia Yim and Kylee T. Ramdeen. 2015.** Hierarchical Cluster Analysis: Comparison of Three Linkage Measures and Application to Psychological Data The Quantitative Methods for Psychology. – Access mode: http://www.tqmp.org/Regular Articles/vol11-1/p008/p008.pdf.
3. **Alexander M.J. and Spaans Willem A. van der Kloot. 2005.** The rules of SPSS' Hierarchical Cluster Analysisfor processing ties.– Access mode: http://www.leidenuniv.nl/fsw/mtlab/software/SPSS%20rules.pdf.
4. **Klimchuk V.O. 2006.** Cluster analysis: the use of psychological studies Proc. Practical psychology and social work. №4, 30–36.(in Ukrainian)
5. **Nasledov A. D. 2004.** SPSS: Computer analysis of data in psychology and social sciences. – SPb. – 416. (in Russian).
6. **Warsha Singh. 2008.** Robustness of three hierarchical agglomerative clustering techniques for ecological data. – Access mode: http://www.unuftp.is/static/files/rannsoknarritegrdir/WarshaSingh_MastersThesis.pdf.
7. **Giovanni Bottegoni, Walter Rocchia, Maurizio Recanatini and Andrea Cavalli. 2006.** AClAP, Autonomous hierarchical agglomerative Cluster Analysis based protocol to partition conformational datasets. – Access mode: http://bioinformatics.oxfordjournals.org/content/22/14/e58.abstract.
8. **I. Crawford, S. Ruske, D. O. Topping and M. W. Gallaghe. 2015.** Evaluation of hierarchical agglomerative cluster analysis methods for discrimination of primary biological aerosol Atmos. Meas. Tech. Discuss., 8, 7303–7333, 201. – Access mode:http://www.atmos-meas-tech – personal /discuss.net/8/7303/2015/amtd-8-7303-2015-print.pdf.
9. **Janette Walde. 2002.** Cluster Analys. – Access mode:http://www.uibk.ac.at/statistics/janettewalde/lehre/phd_biology/clusteranalyse.pdf
10. **Volkova N.A. 2005.** Cluster analysis the results of a sociological survey of employees. – Access mode: http://ieee.tusur.ru/nauka/cluster.pdf. (in Russian).
11. 11**.Sokolova L.V., Veryasova G.M. and Sokolov O.E. 2011.** Application cluster analysis methods in practice enterprises In Proceedings of the N U "Lviv Polytechnic". – №720: Management, 240–246 (in Ukrainian).
12. **Voronin A.V. 2001.** Using cluster analysis to select local strategies In Problems and prospects of economic management and marketing in organizations. –Access mode: http://perspectives.utmn.ru/No1/text02.shtml. (in Russian).
13. **Lavrenko A.S. 2011.** Clustering in the tasks of assortment planning in retail networks In Management of economic systems: electronic scientific journal, № 5 (29). – Access mode: http://uecs.mcnip.ru. (in Russian).
14. **Even G.M. 2009.** Clustered analysis in the study of the motivational structure of personnel of the enterprise [electronic resource] In Actual problems of economy. – № (93). – Access mode: /http:// www.nbuv.gov.ua / Portal/Soc_Gum/APE/2009_3/226-230.pdf. (in Russian).
15. **Kurylo O. 2013.** Structuring expenses of industrial enterprises in the evaluation process of its production and sales potential / A. Emelyanov, O. Kurylo, A. Vysotskij // Ekontechmod. An international quarterly jornal. – Poland, Lublin – Rzeszow. Vol. 02. No. 4, 11–16.
16. **Pasichnyk V.V. and Savchuk V.V. 2015.** Mobile information technologies for tourism domain. Econtechmod: an international quarterly journal on economics in technology, new technologies and modelling processes. – Lublin–Rzeszow, Vol. 04, No. 2, 25–32.
17. **Grabmeier J. and Rudolph A. 2002**.Techniques of cluster algorithms in data mining // Data Mining and Knowledge Discovery. – October, Vol. 6, № 4, 303–360.
18. **Runkler T.A., Bezdek J.C**. **2003.** Web mining with relational clustering In International Journal of Approximate Reasoning. – February, Vol. 32, №. 2–3, 217–236.
19. **Wagstaff K**. **2001.** Constrained K-means Clustreing with Background Knowledge. In Proceedings of the Eighteen International Conference on Mahine Learning . Cardie, 577–584.
20. **Tryon R.C. 1939.** Cluster analysis. London: Ann Arbor Edwards Bros, 139.
21. 3. Statistical method for computers. – Moscow, 1986, 269–301(in Russian).
22. **Blei David M. 2008.** Hierarchical clustering. COS424 Princeton University.February 28. – Access mode http://www.cs.princeton.edu/courses/archive/spr08/cos424/slides/clustering-2.pdf.
23. **Fionn Murtagh and Pierre Legendre. 2011.** Ward's Hierarchical Clustering Method: Clustering Criterion and Agglomerative Algorithm. – Access mode: http://arxiv.org/pdf/1111.6285.pdf.
24. **Kunanets N. E. 2011.** Libraries in Western Ukraine (XIII. – 1939): consolidated information resource monograph. Lviv. 448. (in Ukrainian).
25. **Ana Fred** Hierarchical Clustering: Agglomerative Methods. – Access mode: http://www.lx.it.pt/~afred/tutorials/.pdf.