Petrov T.G.,
Institute of Earth Sciences, Saint Petersburg University, Saint Petersburg, Russia
Moshkin *S.* **V.** *Open Company «Infocomm», Saint Petersburg, Russia*

METHOD RHAT AND ITS IMPLEMENTATION IN THE SOFTWARE PACKAGE PETROS-3

ABSTRACT: *Information language-method* ***RHAT*** *is designed for coding, ordering, and representation of composition changes in objects of any nature. Here R, rank formula of composition, is a rating of components by content decrease, H - C. Shannon information entropy as a measure of composition complexity, A - anentropy as a measure of purity, T-tolerance as a measure of high purity - sterility. Linguistic ordering of* ***RHAT*** *codes of object composition generates hierarchical, periodic systems of compositions. HA and HT diagrams adequately reflect the processes of separation and mixing of compositions. The method is implemented in PETROS-3 software package that also supports a wide range of methods for petrochemical data processing accepted in geology.*

Keywords: information language, Petros3 software package, coding of compositions, rank formula, information entropy, anentropy, tolerance, hierarchic system, periodic system, system of deposits, composistics.

1. Introduction.

Impetus for the creation of ***RHA*** *method and Standard software, out of which the current software package PETROS-3 and information language* ***RHAT*** *has grown, was the understanding that the development of many branches of knowledge is substantially hindered by the lack of a unifying idea for description of compositions of the studied objects as collections, mixtures, cenoses of their components: chemical elements, minerals, biological species, nationalities, age intervals, heights, masses, which creates great difficulties in the primary documentation of analyses and statistical information, during ordering, searching for information both about compositions of multicomponent objects, and processes of their changes under existing in the world conditions of its registration, storage and publication. Consequences are diverse, among them:*

1. Deep dismembering in natural sciences and technology, in particular, under the unity of the chemical space and presence of genetic, spatial-temporal relationships between the compositions of all natural and artificial products. 2. Contradiction between continuous processes of composition change and discrete nomenclature of objects (e.g., rocks, ores, waters, minerals, gases, biological and space objects). 3. Uncertainty of relationships between detailed studies and distinguishing of object varieties by their

composition. 4. Unlimited growth of the number of new object names, not focused on their essence, which is particularly pronounced in mineralogy. 5. Lack of developed methods for assessing the source data quality. 6. Loss of not aging analytical data related to out-of-date names. 7. Lack of effective means for uniform organization of materials to create factual databases. 8. Difficulties in identifying the general orientation of the processes of composition change in objects of different nature, while such a generality may indicate a generality of causes and mechanisms of change.

Solution has been found in the way of iconic (after Ch.S.Pierce) encoding, alphabetic ordering of information on compositions and diagram display of compositions using ***RHA*** method (Petrov 1971; Petrov, Farafonova 2005), later *RHAT* (Petrov 2007; Petrov, Moshkin 2012).

Purpose of the article is to give an idea about the language-method ***RHAT,*** some of its results and method implementation using the software package PETROS-3.

2. Language-method *RHAT*

2.1 Object - composition - component - element - alphabet

Study objects of geologists, demographers, biocenologists, financiers, and experts in many other branches of knowledge are considered as sets (populations, cenoses, mixtures) of indivisible and indistinguishable (at this stage of study) elements combined into components. Elements in the component are indistinguishable (chemical elements), or are not distinguished for some purpose (altitude ranges in a map). Objects are multicomponent, which also applies to those called the «purest» ones or «substance of high purity». Each component has two characteristics: the first - name, the second - content expressed by a number: share, percentage, significance, grade, frequency of occurrence. Set of two rows or columns with names and corresponding numbers is the **object composition.** Analysis length ***n*** characterizes ***detail*** of object study. For component ordering, alphabets called intensional (Chebanov, Petrov 2013) providing semantic links between words during their alphabetical ordering are preferred. In such alphabets letters are meaningful and neighbouring ones have links with each other (perfect examples: the Periodic System of Elements and positive integers).

2.2 Rank formula

The first member in code ***R***, rank formula (Petrov 1971, 2007; Petrov, Moshkin 2012), a sequence of component characters ordered by a nonascending succession of their contents $\boldsymbol{p}_i$. In case of equal $\boldsymbol{p}_i$, symbols are ordered by accepted alphabet and equal sign is put between them. Rank formula is taken for a «word», in which «letters» are symbols of components.

Use of the Periodic Table of Elements as an alphabet enabled to order the ***R***-dictionary of chemical compositions of minerals (Petrov, Krasnova 2010), resulting in a linear system of mineral compositions. It could incorporate unnamed minerals from a large collection (Smith, Nickel 2007; Gordienko 2011). The system is hierarchic, since each ***R*** with length ***n*** falls in a single group of ***R****'s* with rank ***n***-l. ***R****'s* themselves are

constructed hierarchically, as they arrange components by decrease of value. Such a system is periodic, since close compositions are arranged in groups, between which there may be compositions that substantially differ from them. One ***R*** may belong to very different compositions. Therefore, the second parameter is introduced, ***H***.

2.3. Information entropy

C. Shannon information entropy $\boldsymbol{H} = -\sum \boldsymbol{p}_i * \ln \boldsymbol{p}_i$, where pi is frequency of the i-th event, is a measure of composition complexity, characteristic of the degree of uniform component distribution in a composition. When $\boldsymbol{p}_1=\boldsymbol{p}_2=\boldsymbol{p}_3=\ldots=\boldsymbol{p}_n = 1/\boldsymbol{n}$, $\boldsymbol{H}\max=\ln\boldsymbol{n}$. If the object is represented by a single component, $\boldsymbol{H}\min=0$. Up to the gas constant, ***H*** equals to the thermodynamic ***entropy of mixing*** (Moelwyn-Hudges 1961) that is significant in genetic studies (Petrov 2012, 2014). ***H*** value depends on ***n***, therefore, at varying length of the initial analyses ***n*** is standardized at length of the shortest analysis in the sample. Entropy is reduced to the interval 0÷1 by the formula $\boldsymbol{En} = \boldsymbol{H}/\ln\boldsymbol{n}$. Since contribution to the entropy ($-\boldsymbol{p}*\ln\boldsymbol{p}$) is maximum at $\boldsymbol{p}=0.368$ and decreases with decreasing and increasing ***p*** (Fig. 1), the use of entropy at wide content variations (4-5 orders of magnitude or more) becomes ineffective. Furthermore, for $\boldsymbol{n}>3$ identical entropy values may belong (as in the case of rank formulas) to substantially different compositions. Therefore, one more composition characteristic, ***A***, has been introduced.

2.4. Anentropy

Anentropy is calculated by the formula $\boldsymbol{A} = -(\sum \ln\boldsymbol{p}_i)/\boldsymbol{n} - \ln\boldsymbol{n}$. Contributions to anentropy ($-\ln\boldsymbol{p}$) are shown in Fig. 1. $\boldsymbol{A}\min= 0$ at $\boldsymbol{p}_1=\boldsymbol{p}_2=\boldsymbol{p}_3= =\boldsymbol{p}_n$, that is when entropy is maximum. ***Amax*** value is determined *constructively* - for composition of an "***analytically perfectly pure***" one-component system, since zero ***p*** values of a "perfectly pure" one make contributions equal to infinity. For such composition, $\boldsymbol{p}_1=\boldsymbol{p}_2=\boldsymbol{p}_3=\ldots=\boldsymbol{p}_{n-1}=0.5\delta$, $\boldsymbol{p}_n=1-[(\boldsymbol{n}-1)*0.5\delta]$, where δ is analysis method sensitivity. In classifying, indexing of terrestrial objects, and genetic constructs δ is taken for 0.0001, $\boldsymbol{n}=10$. This enables inclusion in the calculation 99.5% of the substance mass. Reducing ***A*** to the interval 0÷1 is made by formula $\boldsymbol{An} = \boldsymbol{A}/\boldsymbol{A}\max$; anentropy the has sense of measure of composition purity, or smallness of small contents. As shown in (Petrov 2012, 2014), ***A*** can play the role of ***entropy of separation*** in physical chemistry, and in computer science it reduces the uncertainty of intuitively assessed type of ranked distribution.

When calculating anentropy, the same feature is more clearly pronounced as in calculations of information entropy, namely the dependence of these values on the completeness of considered information on composition. However, if entropy depends weakly on small contents, and therefore loss of information on small components can be neglected (which is often done), the same loss in ordinary neglecting of statistically small, lack of consideration on the number of components in the analysis, absence of consideration on the analysis method sensitivity, will be significant when calculating anentropy. This is a qualitative difference of anentropy from entropy. Inclusion of anentropy in use of information science, «composistics» (Petrov 1996) as a supra-research tool for investigating

both «complex» and «pure» systems (limited to conventional chemical formulas) opens up new possibilities for ***studying the evolution*** of compositions of natural objects (Petrov 2012, 2014), which do not have major and minor components, that live together and can testify to the state and behaviour of the whole object.

2.5 Tolerance

In the case of extra pure substances, in which «micro-nanocontents» much less than unity are found, anentropy appears little sensitive. To account for such components, value $\boldsymbol{T}= \ln(1/\boldsymbol{n}\sum 1/\boldsymbol{p}_i) - \ln\boldsymbol{n}$ has been introduced. The dependence of contributions in ***T*** *on* ***p*** (i.e.,1/ ***p***) and their comparison with others are shown in Fig. 1.

2.6 Contributions in H,A,T.

Dependences of component contributions in ***HAT*** *on p* are shown in Fig.l.

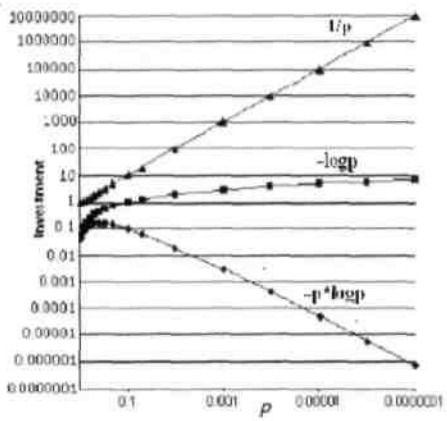

Fig.l. Contributions in ***H,A,T***

As one can see, contributions in tolerance, as contributions in anentropy are monotonically dependent on ***p*** changes. Contributions to ***H, A,*** and ***T*** are clearly and significantly related in terms of sense and mathematically. They relate as path, speed, and acceleration. Thus, the contribution in anentropy in contents (without unity) equals to the first derivative of the contribution in entropy. Contribution in tolerance is the second derivative of the contribution in entropy and the first derivative of the contribution in anentropy. Thereby the introduced characteristics ***A*** and ***T*** preserve an organic link with the fundamental C. Shannon information entropy.

2.7 Entropy diagrams HA, HT

Diagrams are used for two purposes. The first one is representation of analysis groups to identify the point field shape, its uniformity, density, orientation, presence of anomalous points. Second, representation of processes of composition change in time and space (Petrov 2012). The last utilization option is of special value as the most appropriate means for visualizing and comparing the processes of separation and mixing, which, according to Herbert Spencer, are the fundamental processes of composition evolution in nature (Petrov 2014). According to (Shurubor 1972), when systems are separated, a statistically significant entropy reduction in the resulting systems occurs, when mixed, an increase as compared with the initial. Entropy and anentropy changes are usually opposed. Typical areas of change: entropy decrease and anentropy increase, as well as entropy increase and anentropy decrease. Other areas are transitional between the typical ones. Restricted types of trajectories of unidirectional processes in the used diagram enable to identify the individual steps in the evolutionary sequence of compositions. Minerals are generally purer than rocks. Accordingly, magma enriched in crystals (cumulates) is purer than much more liquid, magmatic (Mukhamedzhanov, Petrov 2006). When crystallization process intensity is reduced (Petrov, Farafonova, Sokolov 2003), there is an increase in mineral purity. Thus, more than 80 deposits and occurrences of sphalerite mineral (ZnS) have been discovered in the Polar Urals (Pai-Khoi) and on Vaigach island. Calculations of ***H*** and ***A*** of compositions comprising 15-17 impurity elements revealed the area of minimum ***A*** values (the highest admixture contents) as an indication of the maximum crystallization process intensity in the region and the epicenter of solution inflow for the entire area with dimensions of about 450x60 km. ***A*** values increased in different directions from it, mainly to the northwest (Yushkin et al. 1978). Thus the system of mineral deposits has been distinguished. Let us note as an important fact that at the very substantial ***A*** changes, ***H*** values within the system varied slightly (Fig. 2).

Thus, the range of anentropy values for the mineral, in this case sphalerite, exceeds the range of entropy values in several times, that is a direct evidence of anentropy use effectiveness when studying systems with wide content ranges.

HT diagram given in (Petrov 2014) is used for highly pure substances not found in nature, as well as to improve the legibility of points at their high densities in diagrams.

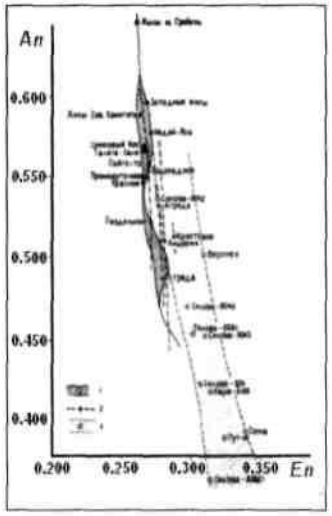

Fig. 2 Entropy characteristics of sphalerite compositions in the system of deposits in Pai-Khoi - Vaigach area. Different filling distinguishes different fault zones.

3. *RHAT* codes ordering

In has been mentioned above on ***R***'s ordering. Ordering of similar ***R*** is made by entropy decrease, and in case of coincidence, by anentropy (and tolerance) increase.

Used method of code ordering, in particular, has generated the Hierarchic Periodic System of Chemical ***Compositions*** (Petrov 2009) and made it possible to organize a Database including over 80,000 records of codes of different composition types (see Table). In lines with characters «*», for comparability with the others, ***EnAn*** are calculated for 10 elements by substituting ***p***= 0.00005. In lines with «XX» calculations are made only for the reduced elements. ***An***> 1 values correspond to purer composition than it is accepted for most of the rest.

Table. Sample from the «Chemical Compositions of the Universe» database

Rank formulae										En	An	Description
H	He	Li	Be	B	C	N	O	F	Ne			First RF
H	He	O	C	Ne	N	Mg=	Si=	Fe	S	0.135	0.837	Sun
H	O	*	*	*	*	*	*	*	*	0.278	0.873	water, pure
H	O	C	N	Ca	P	Na	K=	S	Mg	0.415	0.441	human body
H	O	C	Ca	Na	Cl=	Si	Mg	S	K	0.231	1.217	river water
H	O	N	Cl	Si	Li	B=	S	C	Ca	0.278	0.980	geyser water
H=	O=	K	*	*	*	*	*	*	*	0.479	0.750	potassium hydroxide pure
H	Cl	S	B	Ca	Si	Na	Mg	K	F	0.412	0.425	gas fumarolic
C=	N=	K	*	*	*	*	*	*	*	0.479	0.750	potassium cyanide pure
N	O	Ar	Ca	Ne	H	He	Kr	Cl	Xe	0.235	1.165	atmosphere
O	Mg	Si	Fe	Al	Ca	Na	K=	Cr	Ti	0.542	0.301	Mars
O	Mg	Si	Fe	Al	Ca	Na	Cr	Na	Mg	0.511	0.305	Earth mantle+crust
O	Si	*	*	*	*	*	*	*	*	0.278	0.873	quartz pure
O	Si	H	Al	C=	Ca=	Mg=	Fe=	K	Na	0.578	0.166	clay, Quaternary
O	Si=	H	Al	Na	Ca=	Fe=	Mg	K	C	0.604	0.178	Earth clarke
O	Si	Na	Mg	Al=	Ca	Fe	Mn	W	Ti	0.286	0.804	quartz, Transbaykalia
O	Si	Na	Zr	Ca	Al	K	Mg	Ti	Fe	0.491	0.506	glass Pat.GB
O	Si	Mg	Fe	Al	Ca	Na	Mn	S	K	0.567	0.193	meteorite Zhmerinka
O	Si	Al	Na	K	H	Fe	Ca	Mg	Ti	0.488	0.247	granite, average
O	Si	Al	Ca	Na	Fe	Mg	K	Ti	Mn	0.476	0.494	glass volc. Moon
O	K=	Br	*	*	*	*	*	*	*	0.414	0.757	potassium bromate pure
O	Ca=	H	C	P	Na	Si=	Mg	S	K	0.646	0.220	bones fish
O	Ca	Si	Mg	Al	Fe	Na	K=	Ti	Mn	0.535	0.362	portlandcement
Na	K	P	Fe	Ca=	Mg	Si	Zn	Al	Cu	0.508	0.359	venous blood
Al	Si	Zn	Fe	Mg	Ni=	Mn=	Cu	Pb	Ti	0.227	0.335	alloy AlCu 21000_D
Al	Cu	Mg	Mn	Sc=	Cr	Fe	Se	Ni=	Ga	<0.001	1.977	Al high pure
Cl	Na	Mg	S	Ca=	K	H	C	Br	B	0.466	0.334	sea-salt
K=	Br	*	*	*	*	*	*	*	*	0.303	0.871	potassium bromide pure
Fe	C	Cr	V	Mo	W	Si	Mn	XX	XX	0.391	0.216	steel M3 USA
Ni	Fe	Cr	Si	Mn	C=	Al	Ti	P	S	0.524	0.303	steel X15H60 Ru
Cu	Fe	Ag	Sn	As	Ni	XX	XX	XX	XX	0.021	0.722	knife, aerchaeological
Cu	Sn	Pb	Fe	As=	Mn	XX	XX	XX	XX	0.122	0.838	prick aerchaeological
Au	Cu	Ag	=Fe	Pb	Bi	XX	XX	XX	XX	0.003	0.947	gold 99.99
Lr	No	Md	Fm	Es	Cf	Bk	Cm	Am	Pu			Last RF

There is no chemical composition that does not find a place in the system; this fact enabled to give the base such a name. Rank formula corresponding to the order in the Periodic Table of Elements is given at the table beginning, the latter one is inverse with respect to the first one. All real, theoretical, and discovered compositions will be placed between them in the filled system.

Databases of mineral compositions of individual rock types (Petrov, Krasnova 2012); crystal chemical compositions of tourmaline (Petrov, Andriyanets-Buyko, Moshkin 2012) as a prototype of a general crystal chemical classification of minerals

and other compounds are organized similarly. Method for ordering codes of halftone rock patterns is provided in the same logical framework (Petrov, Shuisky 2013).

4. *PETROS-3* software package

4.1 Shared about the program

Program reserves in the database both the input information in original form, and the processing results of each object, which enables searching and constructing of samples not only by compositions and descriptions of objects, but also by their estimated characteristics, in particular, ***R***'s and ***A, H***, and ***T*** values. An important feature of PETROS-3 SP is the opportunity to work with many alphabets. Up to 100 alphabets of 100 characters each can be inputted in the described version.

Software feature is the ability to assess the representativeness and novelty of input data with respect to a random sample. The presence of «Bibliography» block provides communication of objects with the primary sources of information.

4.2 Other tools of PETROS-3 SP

In addition to the block provided for implementation of the described method of analytical data processing, the software includes a variety of methods used in geochemical data processing.

4.2.1 Block for standard petrochemical index computation. Supports the ability to add new calculation formulas.

4.2.2 Block of regulatory petrochemical revaluation. Designed to handle data using rather complex algorithms (e.g., CIPW). Supports the ability to add new algorithms. The block contains a user friendly graphic interface for editing algorithms. Algorithm created with its help is converted into a program code and stored in the database. If necessary, this code is executed by the interpreter integrated in PETROS-3.

4.2.3 Block for construction of binary and ternary diagrams enables construction of an indefinitely large number of diagrams, along the axes of which arbitrary functions of element contents, standard petrochemical indices or variables derived from the revaluation can be plotted. Computation of indices of linear and polynomial regressions of variables plotted on the diagram axes is possible.

4.2.4 Block for computation of pair correlation indices between the individual components, to which integral characteristics ***EnAn*** are added.

4.2.5 Block for computation of the basic statistic parameters (mean, median, dispersion, asymmetry, kurtosis, approximating Gaussian distribution indices.

4.2.6 Block for spider diagrams construction, i.e. diagrams of relative analysis. Standard is user selectable.

4.2.7 Block for construction and use of classification diagrams, i.e. diagrams con taining fields (or intervals), getting into which enables object attributing to a certain class. Classification is carried out automatically.

4.2.8 Block for calculation of distances between analyses. Distances in the data array are computed in 4 versions:

Euclidean: $Eu = \sqrt{\Sigma (p_i - q_i)^2}$
Minkowski: $Mi = \Sigma | p_i - q_i |$
Entropy: $De = [\Sigma (p_i - q_i) * \ln(p_i / q_i)]/n$,
Anentropy: $Da = [\Sigma |\ln(p_i/q_i)|]/n$.

The first two are applicable to high-entropy compositions, entropy for intermediate ones, and De for low-entropy, the purest compositions. Tables include data for calculating average distances enabling to measure the analysis position with respect to the center of point distribution in the composition cloud.

Results of all computations and constructed diagrams can be printed or exported to MS Excel or MS Word.

4.3 Work with the software

Software package has a user friendly interface with main menu and toolbars changing depending on the current software state.

Data entry is made in a convenient tabular interface. In addition, data import from files in text format, and import from MS Excel are possible. Correctness of the initial data is monitored during entering. Initial data and results of ***RHA*** characteristics computation are stored in the database. Access to the data is possible both by the dataset name and by creating a sample (inquiry) on arbitrary analysis attributes (including descriptive features, component content values, values of ***RHA*** characteristics etc.) and their arbitrary combinations. A simple, intuitive interface is used for query forming. Formulated queries are stored in the database and can be used in all methods implemented in PETROS-3.

4.4 Technical specifications

Software package PETROS-3 runs on Windows (Windows XP SP3 and above), uses MS SQL Server (2000 or higher) as a database server. Standard distribution includes a free version of MS SQL Server 2008 Express

System requirements:

Client side: Celeron 800 MHz CPU, 512 MB RAM, 20 MB hard disk space;

Server side: standard requirements for MS SQL Server 2008 + 50 MB of disk space for every thousand analyses in the database.

5. Conclusion

RHA *T* designation in the context of composition representation can be considered gross formula. Hill system (1902) is the most well known of them.

Removal of the following oppositions of semantic primitives (Chebanov, Petrov 2005) using ***RHAT*** method can be considered an important methodological achievement:

«Discrete and continuous»
«Major and minor»
«Large and small»
«Structure-forming and admixture»

«Immovable and movable»
«Frequent and rare»
«Typical and variable».

The first members are mainly reflected in rank formula and entropy when solving the problems of material registration. The second ones, mainly in entropy and anentropy when studying movement and relations between compositions.

RH*AT* method and PETROS-3 SP that enabled to solve the issue of discretization of the continuous chemical space can be used for multi-aspect treating of a variety of analytical information, in geology, technology dealing with matter, and in many other fields of knowledge.

Acknowledgements. The authors thank S.V. Chebanov for consultations on the semiotic system issues.

Bibliography

1. Chebanov S.V., Petrov T.G. Language-method RHA for describing compositions of multicomponent objects and processes of their change.// In: Man-made self-organization and mathematical apparatus of cenological investigations. M CSR, Iss. 28. 2005 Pp. 345-370. (In Russian)
2. Chebanov S.V., Petrov T.G. Intensionality, intensional alphabets, intensional words and dictionaries. In: Actual problems of modern cognitive science. Ivanovo. 2013. Pp.239-266. (In Russian)
3. Gordienko V. V. Two codes and two databases of chemical compositions of minerals - which one is better? //Zapiski RMO, N 3, 2011(In Russian)
4. Hill E.A. On a system of indexing chemical literature; Adopted by the Classification Division of the U.S. Patent Office // J. Am. Chem. Soc, 1900, 22 (8). Pp. 478-494.
5. Moelwyn-Hudges, E.A., Physical Chemistry, London: Pergamon Press, 1961
6. Mukhamedzhanov A.F., Petrov T.G. Consideration of differentiated intrusions using chemical-information RHA method. // Structure, matter, history of the Timan-Northern Urals lithosphere segment. Ur. B. RAS. Syktyvkar. 2006. Iss. 15. (In Russian)
7. Petrov T.G. Substantiation of a version of general classification of geochemical systems. // LSU Bulletin. No. 18. 1971. P. 30-38. (In Russian)
8. Petrov T.G. Language-method RHA for describing, systematizing, and studying changes in compositions of any nature. // In: Cenological studies. Ed. B.I. Kudrin. Iss. 1 Mathematical description of cenoses and technetical laws. Abakan: SRC 1996. Pp.256-273. (In Russian)
9. Petrov T.G. Rank-entropy approach to description of compositions of geological objects and their changes (exemplified by geological cenology). // General and applied cenology. 2007. No.5. - Pp.27-33. (In Russian)
10. Petrov T.G. Hierarchical periodical system of chemical compositions and its association with the periodical system of elements. //SPb SU Bulletin. Ser. 7. 2009. - Iss. 2. Pp. 21-28. (In Russian)

11. Petrov T. G. Graphic Representation of the Evolutionary Processes of the Compositions of Multicomponent Objects of Any Nature. //Automatic Documentation and Mathematical Linguistics, 2012, Vol. 46, No. 2, pp. 79-93. © Allerton Press, Inc., DOI: 10.3103/S0005105512020045
12. Petrov T. G. Separation-Mixing as a Model of Composition Evolution of any Nature.// J. Systemics, Cybernetics and Informatics V. 12, N 1 2014, pp. 76-81. ISSN: 1690-4524
13. Petrov T. G., A. A. Andriyanets-Buyko, and S. V. Moshkin. A Two-Parameter Alphabet for Coding Structural-Chemical Information and its Systematization (Using the Example of Tourmaline).// Automatic Documentation and Mathematical Linguistics, 2012, Vol. 46, No. 1, pp. 40-49. © Allerton Press, Inc., 2012. SpringerLink Online T. G. Petrov, A. A. Andriyanets-Buyko, and S. V. Moshkin. Automatic Documentation and Mathematical Linguistics, 2012, DOI 10.3103/S0005105512010086 (Author, Journal Title, Year, DOI)
14. Petrov T.G., Krasnova N.I. R-dictionary-catalogue of mineral chemical compositions. St. Petersburg, «Nauka», 2010. 150 p. (In Russian)
15. Petrov T.G., Krasnova N.I. RHA-coding of mineral composition of crystalline rocks and method of their systematization. Modern problems of magmatism and metamorphism. Vol.2 - St. Petersburg 2012. Pp. 120-123. (In Russian)
16. Petrov T.G., Farafonova O.I. Information-component analysis. RHA method. (Study guide). St. Petersburg. 2005. 168 p. (In Russian)
17. Petrov T.G., Farafonova O.I., Sokolov P.B. Informational-entropic characteristics of composition of minerals and rocks as a reflection of crystallization process stress.// RMS Proceedings 2003. No.2. Pp. 33-40. (In Russian)
18. Petrov T. G., Moshkin S.V. RHA(T)-System for Coding of Discrete Distributions and Their Alteration Processes.// Proc. The 3rd International Multi-Conference on Complexity, Informatics and Cybernetics IMCIC 2012. 2012 pp. 12-16
19. Petrov T.G., Shuisky A.V. Parametric description of gemstone pattern. // Modern issues of science and education 2013, No. 5. (In Russian)
20. Shurubor Yu.V. On one property of complexity measure of geochemical systems.// Proceedings of the AS USSR. 1972. V. 205. No.2. Pp. 453-456. (In Russian)
21. Smith D. G. W. and Nickel E. H. A system of codification for unnamed minerals: report of the subcommittee for unnamed minerals of the IMA commission on new minerals, nomenclature and classification// Canadian Mineralogist, 2007, vol. 45. Pp. 983-1055.
22. Yushkin N.P., Eremin N.I., Makeev A.B., Petrov T.G. Sphalerite from the Pai-Khoi-Yuzhnaya Zemlya Province (topomineralogy and typomorphism).// Pr. Inst. Of Geology, Komi Branch of the AS USSR. Iss.24. 1978. Pp. 23-52. (In Russian)