

O. Syrotkina, M. Alekseyev, I. Udovyk

## GRAPHICAL AND ANALYTICAL METHODS FOR PROCESSING “BIG DATA” BASED ON THE ANALYSIS OF THEIR PROPERTIES

*Abstract. This article addresses the subject of creating mathematical methods in order to optimize time and computing resources when processing “big data.” One of the ways of solving this problem is the creation of NoSQL systems, an advantage of which is the flexibility of data models as well as the possibility of horizontal scaling, parallel processing and the speed of obtaining results. From the viewpoint of “big data” analysis, there have been other methods developed such as machine learning, artificial intelligence, distributed processing of streams and events, and visual data research technology.*

*Furthermore, the aim of the research is to develop mathematical methods for processing “big data” based on the system analysis of the data structure properties known as “m-tuples based on ordered sets of arbitrary cardinality (OSAC).”*

*The data structure “m-tuples based on OSAC” is the Boolean, which is ordered by right-side enumeration of the elements of the basis set with cardinality  $n$  from the lower boundary of the possible change of the index value for each element of the tuple to the upper one. We formulated certain properties for the data structure investigated. These properties result from rules of logic when forming this structure. We also described mathematical methods based on these properties. Boolean graphs are illustrated with drawings and the outlined vertices of the graph correspond to the declared properties of the given data structure. We derived analytical dependencies to determine these Boolean elements. These Boolean elements do not require the execution of algorithms that implement the particular operations of intersection, union, and membership because the desired result is already determined by these properties.*

*The properties of the data structure in question with regards to m-tuples based on OSAC allow us to determine some interdependencies between m-tuples by their location in the structure. Their location is determined by a pair of indices  $(j, m)$  without executing computing algorithms. In this case, the time estimate for obtaining results changes from a cubic  $O(n^3)$  to linear  $O(n)$  dependency.*

*Key words: “big data,” data structure, ordered set of arbitrary cardinality, m-tuples, Boolean graph, minimization of time and computing resources.*

**Problem statement.** At the present stage of information technology development, the task of processing, storing, analyzing and managing “big data” remains a serious problem. Recently, there have been many new and diverse “big data” generators such as social networks, web networks, cellular networks, multichamber video surveillance systems, distributed multi-level, multi-tasking and multi-user SCADA

systems, etc. At the same time, important problematic aspects of the uses of “big data” are the methods and tools to analyze them [1].

**Publication analysis regarding topic research.** At the present time, the storage and processing of “big data” in systems specifically based on relational databases is not always highly efficient. In order to store data in a relational database, we need to perform several steps: develop a data storage structure, filter and convert data to the required format, and load data into the database. Any of these stages can be a long and time-consuming process.

One of the ways of solving this problem is the creation of a new class of NoSQL (Not Only SQL) systems. The advantage of a NoSQL system is the flexibility of data models which include the use of horizontal scaling and parallel processing. Results from this system can be obtained rather quickly.

From the viewpoint of “big data” analysis, there have been other methods developed such as machine learning, artificial intelligence, distributed processing of streams and events, and visual data research technology.

Article [2] describes the models of interoperability and navigation through big data, allowing information to be extracted from large and complex collections of semi-structured data. The difference between the proposed models and the traditional relational data model is that the information about the structure of big data arrays is defined in terms of the domain identifier and business key. Using a domain/key mathematical model allows us to create a structure that can remove all restrictions, except for domain restrictions and key restrictions. However, the practical implementation of converting a collection of semi-structured data into a normal domain/key form remains an open question.

In article [3], the author considers mathematical methods and principles for constructing expert systems. The work describes such concepts as the system of alternatives, incompatibility modules, the module of the ban, implication modules, and production modules. It shows the implementation of the inference in the system of alternatives. According to the author, the method of alternative systems is designed to organize and process data when solving problems associated with the search for options. Another way of using this method is to create an expert system knowledge base in the form of a system of alternatives for solving recognition and classification problems. The paper also discusses methods for restoring regular and context-free languages. These methods are used by the author in the automated extraction of knowledge from the knowledge system of the expert system.

The principles of building an expert system for diagnosing an object by unstructured diagnostic features (UDF) are presented in article [4]. The paper shows that the values of the UDF can be determined by a set of patterns. The main idea of the diagnostic method is to determine the similarity coefficient of the diagnostic object compared to its pattern. The knowledge base of the expert system is determined by the graph. The vertices of the graph are the set of UDFs and its arcs are the set of patterns defining the values of UDFs. After calculating the similarity coefficients of the diagnostic object compared to its pattern, their values are assigned to the arcs. After that, the author calculates the probability of establishing a connection between the initial vertex of the graph with its terminal vertices. The paper concludes that if the degree of diagnostic discernibility corresponds to a given value, then the diagnostic process is complete. Otherwise, the process continues on to the next UDF.

In article [5], the authors consider the diagnostic concept of how information systems (IS) work based on LV (Latent Variable) models. The system determines the process parameters that exceed established monitoring levels based on the array of archived data collected by SCADA. These parameters are associated with the operation of equipment, the behavior of the technological process or the disruption of operating modes. The approach is based on the ability of personnel to formulate an expert assessment of the trends in the technological process and equipment operation. Knowledge-based decision support systems can be applied to automate the process of forming an expert assessment. At the same time, LV models must be unique and interpretable. It also requires processing large arrays of missing data and monitoring their integrity within the LV area as established by the training set. Restrictions on the use of LV models built on the archival data of the system consist of limiting the space of latent variables determined by those archived data. LV models also cannot be used to extrapolate the modes of information system operation for which there are no monitoring statistics.

**Formulation of the aim of the article.** The aim of the research is to develop mathematical methods for processing “big data.” It is based on the system analysis of the data organizational and structural properties known as “ $m$ -tuples based on ordered sets of arbitrary cardinality (OSAC).” The use of this data structure in the NoSQL database with the “key value” data model will allow optimizing the main characteristics of the “big data” such as reducing the amount of stored and processed information, an increase in the speed of searching, processing and transmitting data while preserving their value and reliability.

**Main part.** The description of the basic terms and definitions, as well as some properties and mathematical methods to work with “m-tuples based on OSAC” are given in the works [6 – 8].

In this paper, we consider mathematical methods for determining the truth of expression (1), whose operands are elements of Boolean  $2^X$ , represented by “m-tuples” based on an ordered basis set  $X$  of cardinality  $n$ .

$$\begin{cases} y_{m_1, j_1}^n \subset y_{m_2, j_2}^n, \\ m_1 < m_2 \end{cases}, \tag{1}$$

where  $y_{m, j}^n$  – m-tuple, Boolean element  $2^X$ ;  $n$  – cardinality of ordered basis set  $X$ ;  $m_1, m_2$  – tuple lengths;  $j_1, j_2$  – indices (index numbers) of m-tuples in ordered sets  $Y_{m_1}^n$  and  $Y_{m_2}^n$ .

As an example for our research, we can define a symbolic alphabetically ordered basis set  $X$  of cardinality  $n = 8$ . The decomposition of the Boolean  $2^X$  into sets is given in Table 1.

Table 1

Representation of a Boolean  $2^X$  by sets  $Y_m^n$  ( $n = 8$ )

$m$	$Y_m^n$	$k_m^n$
1	2	3
1	$Y_1^8 = \{a, b, c, d, e, f, g, h\}$	8
2	$Y_2^8 = \{(a, b), (a, c), (a, d), (a, e), (a, f), (a, g), (a, h), (b, c), (b, d), (b, e), (b, f), (b, g), (b, h), (c, d), (c, e), (c, f), (c, g), (c, h), (d, e), (d, f), (d, g), (d, h), (e, f), (e, g), (e, h), (f, g), (f, h), (g, h)\}$	28
3	$Y_3^8 = \{(a, b, c), (a, b, d), (a, b, e), (a, b, f), (a, b, g), (a, b, h), (a, c, d), (a, c, e), (a, c, f), (a, c, g), (a, c, h), (a, d, e), (a, d, f), (a, d, g), (a, d, h), (a, e, f), (a, e, g), (a, e, h), (a, f, g), (a, f, h), (a, g, h), (b, c, d), (b, c, e), (b, c, f), (b, c, g), (b, c, h), (b, d, e), (b, d, f), (b, d, g), (b, d, h), (b, e, f), (b, e, g), (b, e, h), (b, f, g), (b, f, h), (b, g, h), (c, d, e), (c, d, f), (c, d, g), (c, d, h), (c, e, f), (c, e, g), (c, e, h), (c, f, g), (c, f, h), (c, g, h), (d, e, f), (d, e, g), (d, e, h), (d, f, g), (d, f, h), (d, g, h), (e, f, g), (e, f, h), (e, g, h), (f, g, h)\}$	56
4	$Y_4^8 = \{(a, b, c, d), (a, b, c, e), (a, b, c, f), (a, b, c, g), (a, b, c, h), (a, b, d, e), (a, b, d, f), (a, b, d, g), (a, b, d, h), (a, b, e, f), (a, b, e, g), (a, b, e, h), (a, b, f, g), (a, b, f, h), (a, b, g, h), (a, c, d, e), (a, c, d, f), (a, c, d, g), (a, c, d, h), (a, c, e, f), (a, c, e, g), (a, c, e, h), (a, c, f, g), (a, c, f, h), (a, c, g, h), (a, d, e, f), (a, d, e, g), (a, d, e, h), (a, d, f, g), (a, d, f, h), (a, d, g, h), (a, e, f, g), (a, e, f, h), (a, e, g, h), (a, f, g, h), (b, c, d, e), (b, c, d, f), (b, c, d, g), (b, c, d, h), (b, c, e, f), (b, c, e, g), (b, c, e, h), (b, c, f, g), (b, c, f, h), (b, c, g, h), (b, d, e, f), (b, d, e, g), (b, d, e, h), (b, d, f, g), (b, d, f, h), (b, d, g, h), (b, e, f, g), (b, e, f, h), (b, e, g, h), (b, f, g, h)\}$	70

	$(b,e,f,h),(b,e,g,h),(b,f,g,h),(c,d,e,f),(c,d,e,g),(c,d,e,h),(c,d,f,g),(c,d,f,h),(c,d,g,h),$ $(c,e,f,g),(c,e,f,h),(c,e,g,h),(c,f,g,h),(d,e,f,g),(d,e,f,h),(d,e,g,h),(d,f,g,h),(e,f,g,h)\}$	
5	$Y_5^8 = \{(a,b,c,d,e),(a,b,c,d,f),(a,b,c,d,g),(a,b,c,d,h),(a,b,c,e,f),(a,b,c,e,g),(a,b,c,e,h),$ $(a,b,c,f,g),(a,b,c,f,h),(a,b,c,g,h),(a,b,d,e,f),(a,b,d,e,g),(a,b,d,e,h),(a,b,d,f,g),$ $(a,b,d,f,h),(a,b,d,g,h),(a,b,e,f,g),(a,b,e,f,h),(a,b,e,g,h),(a,b,f,g,h),(a,c,d,e,f),$ $(a,c,d,e,g),(a,c,d,e,h),(a,c,d,f,g),(a,c,d,f,h),(a,c,d,g,h),(a,c,e,f,g),(a,c,e,f,h),$ $(a,c,e,g,h),(a,c,f,g,h),(a,d,e,f,g),(a,d,e,f,h),(a,d,e,g,h),(a,d,f,g,h),(a,e,f,g,h),$ $(b,c,d,e,f),(b,c,d,e,g),(b,c,d,e,h),(b,c,d,f,g),(b,c,d,f,h),(b,c,d,g,h),(b,c,e,f,g),$ $(b,c,e,f,h),(b,c,e,g,h),(b,c,f,g,h),(b,d,e,f,g),(b,d,e,f,h),(b,d,e,g,h),(b,d,f,g,h),$ $(b,e,f,g,h),(c,d,e,f,g),(c,d,e,f,h),(c,d,e,g,h),(c,d,f,g,h),(c,e,f,g,h),(d,e,f,g,h)\}$	56
6	$Y_6^8 = \{(a,b,c,d,e,f),(a,b,c,d,e,g),(a,b,c,d,e,h),(a,b,c,d,f,g),(a,b,c,d,f,h),$ $(a,b,c,d,g,h),(a,b,c,e,f,g),(a,b,c,e,f,h),(a,b,c,e,g,h),(a,b,c,f,g,h),(a,b,d,e,f,g),$ $(a,b,d,e,f,h),(a,b,d,e,g,h),(a,b,d,f,g,h),(a,b,e,f,g,h),(a,c,d,e,f,g),(a,c,d,e,f,h),$ $(a,c,d,e,g,h),(a,c,d,f,g,h),(a,c,e,f,g,h),(a,d,e,f,g,h),(b,c,d,e,f,g),(b,c,d,e,f,h),$ $(b,c,d,e,g,h),(b,c,d,f,g,h),(b,c,e,f,g,h),(b,d,e,f,g,h),(c,d,e,f,g,h)\}$	28
7	$Y_7^8 = \{(a,b,c,d,e,f,g),(a,b,c,d,e,f,h),(a,b,c,d,e,g,h),(a,b,c,d,f,g,h),$ $(a,b,c,e,f,g,h),(a,b,d,e,f,g,h),(a,c,d,e,f,g,h),(b,c,d,e,f,g,h)\}$	8
8	$Y_8^8 = \{(a,b,c,d,e,f,g,h)\}$	1

where  $k_m^n$  – cardinality of set  $Y_m^n$ .

$$k_m^n = |Y_m^n| = \binom{n}{m} = \frac{n!}{(n-m)! \cdot m!} \tag{2}$$

We can represent Boolean  $2^X$  on the basis of an ordered basis set  $X$  of cardinality  $n = 8$  in the form of a graph (see Fig. 1), where each vertex of the graph corresponds to an element  $y_{m,j}^n$ .

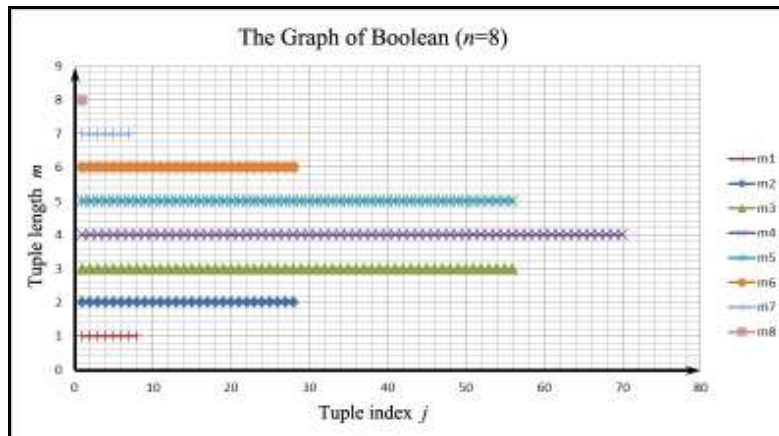


Figure 1 - The graph of Boolean  $2^X$  when  $n = 8$

where  $n$  – the number of parallel segments of the graph corresponding to the number of elements of basis set  $X$ ;  $m$  – the ordinal number of the segment that corresponds to the length of Boolean element  $y_{m,j}^n$ ;  $j$  – the ordinal number of the graph vertex that corresponds to the ordinal number of Boolean element  $y_{m,j}^n$  in ordered subset  $Y_m^n$  of Boolean elements of the same length;  $k_m^n$  – the number of vertices on the  $m^{\text{th}}$  segment of the graph, corresponding to the subset cardinality;  $(j, m)$  – coordinates of the graph vertex.

We define table 2 which shows if expression (1) is true for element  $y_{m_1,j_1}^n = y_{1,1}^8 = a$ . In other words, we define set  $J_{m_2}^n = \{j_2\}$ , where each of its elements  $j_2$  unequivocally determines tuple  $y_{m_2,j_2}^n$  for which expression (1) is true.

**Property of membership for Boolean elements 1.** Expression (1) is true for  $\binom{n-m_1}{m_2-m_1}$  elements of set  $Y_{m_2}^n$ .

**Property of membership for Boolean elements 2.** Expression (1) is true for  $\sum_{m_2=m_1+1}^n \binom{n-m_1}{m_2-m_1}$  elements of Boolean  $2^X$ .

**Property of membership for Boolean elements 3.** If  $j_1 = 1$ , then expression (1) is true for the first  $\binom{n-m_1}{m_2-m_1}$  elements of set  $Y_{m_2}^n$ .

Table 2

Determination of set  $J_{m_2}^8$  for Boolean element  $y_{m_1,j_1}^n = y_{1,1}^8$

$m_2$	$J_{m_2}^8$	$nj_{2_{m_2}}(y_{m_1,j_1}^n) = \binom{n-m_1}{m_2-m_1}$
2	$J_{2_2}^8 = 1 \div 7$	$nj_{2_2}(y_{1,1}^8) = \binom{7}{1} = 7$
3	$J_{3_3}^8 = 1 \div 21$	$nj_{3_3}(y_{1,1}^8) = \binom{7}{2} = 21$
4	$J_{4_4}^8 = 1 \div 35$	$nj_{4_4}(y_{1,1}^8) = \binom{7}{3} = 35$
5	$J_{5_5}^8 = 1 \div 35$	$nj_{5_5}(y_{1,1}^8) = \binom{7}{4} = 35$
6	$J_{6_6}^8 = 1 \div 21$	$nj_{6_6}(y_{1,1}^8) = \binom{7}{5} = 21$
7	$J_{7_7}^8 = 1 \div 7$	$nj_{7_7}(y_{1,1}^8) = \binom{7}{6} = 7$
8	$J_{8_8}^8 = 1$	$nj_{8_8}(y_{1,1}^8) = \binom{7}{7} = 1$

where  $nj2_{m_2}(y_{m_1,j_1}^n)$  – cardinality of set  $J_{m_2}^n$ .

$$nj2_{m_2}(y_{m_1,j_1}^n) = |J_{m_2}^n| = \binom{n - m_1}{m_2 - m_1}. \quad (3)$$

Figure 2 shows the Boolean graph with the cardinality of the basis set  $n = 8$ . Closed polylines outline the sets of Boolean elements represented by tuples of length  $m_2 > m_1$ , for which each element  $y_{m_1,1}^8$  is a subset of each set element  $J_{m_2}^8$ .

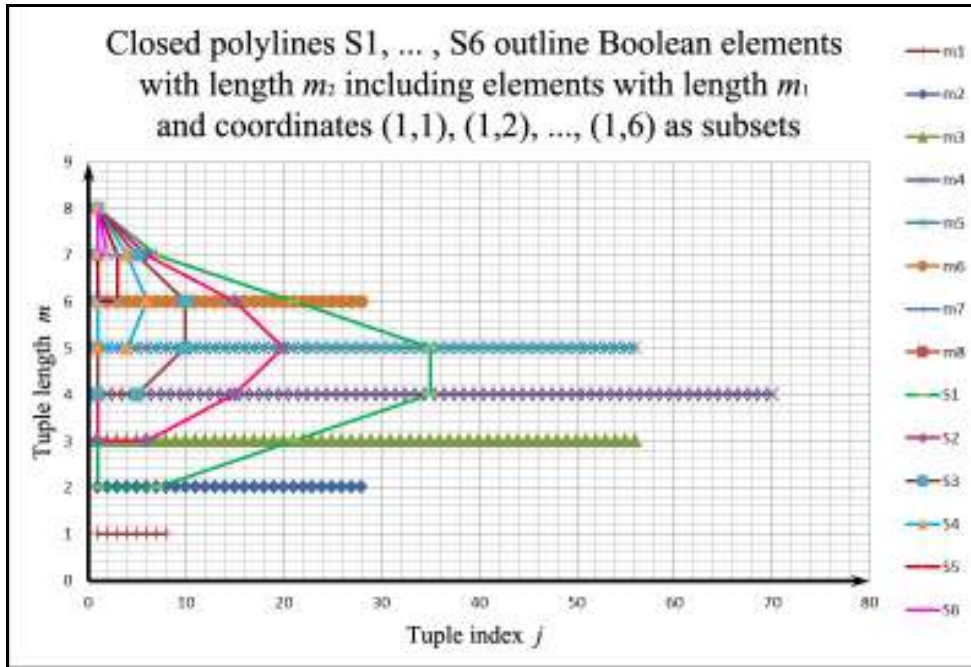


Figure 2 - Boolean graph with closed polylines for determining sets  $J_{m_2}^8$  to elements  $y_{m_1,1}^8$

Example 1.

We have  $n, m_1=1, j_1=1, m_2, j_2, m_1 < m_2$ .

Required to determine if expression (1) is true for element  $y_{m_2,j_2}^n$ .

Decision:

$$nj2_{m_2} = \binom{n - m_1}{m_2 - m_1}. \quad (4)$$

Answer:

Expression (1) is true for expression  $1 \leq j_2 \leq nj2_{m_2}$ .

We can define table 3 which shows if expression (1) is true for element  $y_{m_1,j_1}^n = y_{1,2}^8 = b$ .

Table 3

Determination of set  $J_{m_2}^8$  for Boolean element  $y_{m_1, j_1}^n = y_{1,2}^8$

$m_2$	$\eta$	$J_{m_2, \eta}^8$	$nj_{2, m_2, \eta}(y_{m_1, j_1}^n)$	$nj_{2, m_2}(y_{m_1, j_1}^n)$
2	2	1	1	7
	1	$8 \div 13$	6	
3	2	$1 \div 6$	6	21
	1	$22 \div 36$	15	
4	2	$1 \div 15$	15	35
	1	$36 \div 55$	20	
5	2	$1 \div 20$	20	35
	1	$36 \div 50$	15	
6	2	$1 \div 15$	15	21
	1	$22 \div 27$	6	
7	2	$1 \div 6$	6	7
	1	8	1	
8	2	1	1	1

where  $\eta$  – the ordinal number of element  $y_{m_1, j_1}^n$  in tuple  $y_{m_2, j_2}^n$ ;  $J_{m_2, \eta}^8$  – set of indices  $j_2$ , that unequivocally determines a set of elements  $y_{m_2, j_2}^n$  for which expression (1) is true. At the same time, element  $y_{m_1, j_1}^n$  is located in  $\eta$  in tuple  $y_{m_2, j_2}^n$ ;  $nj_{2, m_2, \eta}$  – cardinality of set  $J_{m_2, \eta}^8$ .

Figure 3 shows the Boolean graph with the cardinality of the basis set  $n = 8$ . Closed polylines outline the sets of Boolean elements represented by tuples of length  $m_2 > m_1$ , for which element  $y_{1,2}^8$  is a subset of each of the elements of set  $J_{m_2}^8$ .

On the basis of table 3, we can create table 4 which shows if expression (1) is true for element  $y_{1,2}^n$  when  $n$  has any value possible.

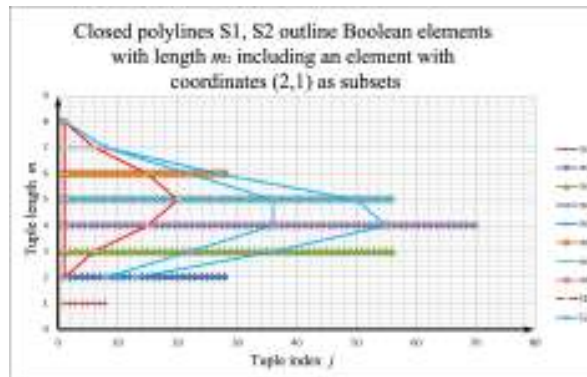


Figure 3 - Boolean graph with closed polylines for determining sets  $J_{m_2}^8$  to element  $y_{1,2}^8$



Table 4

Determination of set  $J2_{m_2}^n$  for Boolean element  $Y_{1,2}^n$

$m_2$	$\eta$	$J2_{m_2, \eta}^n = \{j2_{m_2, \eta, 1}, \dots, j2_{m_2, \eta, last}\}$	$nj2_{m_2, \eta}(y_{m_1, j_1}^n)$	$nj2_{m_2}(y_{m_1, j_1}^n)$
2	2	$j2_{2,2,1} = 1$ $j2_{2,2,last} = nj2_{2,2}$	$nj2_{2,2} = \binom{n-j_1}{m_2-m_1-(\eta-1)}$ $= \binom{n-j_1}{0} = 1$	$nj2_2 = \binom{n-m_1}{m_2-m_1} =$ $= \binom{n-1}{2-1} = n-1$
	1	$j2_{2,1,1} = nj2_2 + 1$ $j2_{2,1,last} = nj2_2 + nj2_{2,1}$	$nj2_{2,1} = \binom{n-j_1}{m_2-m_1-(\eta-1)} =$ $= n - j_1$	
3	2	$j2_{3,2,1} = 1$ $j2_{3,2,last} = nj2_{3,2}$	$nj2_{3,2} = \binom{n-j_1}{m_2-m_1-(\eta-1)} =$ $= \binom{n-j_1}{1} = n - j_1$	$nj2_3 = \binom{n-m_1}{m_2-m_1} =$ $= \binom{n-1}{3-1} =$ $= \binom{n-1}{2}$
	1	$j2_{3,1,1} = nj2_3 + 1$ $j2_{3,1,last} = nj2_3 + nj2_{3,1}$	$nj2_{3,1} = \binom{n-j_1}{m_2-m_1-(\eta-1)} =$ $= \binom{n-j_1}{2}$	
$m_2$	$\eta$	$J2_{m_2, \eta}^n = \{j2_{m_2, \eta, 1}, \dots, j2_{m_2, \eta, last}\}$	$nj2_{m_2, \eta}(y_{m_1, j_1}^n)$	$nj2_{m_2}(y_{m_1, j_1}^n)$
...	...	...	...	...
$\mu$	2	$j2_{\mu,2,1} = 1$ $j2_{\mu,2,last} = nj2_{\mu,2}$	$nj2_{\mu,2} = \binom{n-j_1}{\mu-m_1-(\eta-1)}$	$nj2_{\mu} = \binom{n-m_1}{\mu-m_1}$
	1	$j2_{\mu,1,1} = nj2_{\mu} + 1$ $j2_{\mu,1,last} = nj2_{\mu} + nj2_{\mu,1}$	$nj2_{\mu,1} = \binom{n-j_1}{\mu-m_1-(\eta-1)}$	
...	...	...	...	...
$n-1$	2	$j2_{n-1,2,1} = 1$ $j2_{n-1,2,last} = nj2_{n-1,2} = n-2$	$nj2_{n-1,2} =$ $= \binom{n-j_1}{n-1-m_1-(\eta-1)} =$ $= \binom{n-2}{n-3} = n-2$	$nj2_{n-1} = \binom{n-m_1}{n-1-m_1} =$ $= \binom{n-1}{n-2} = n-1$
	1	$j2_{n-1,1,1} = nj2_{n-1} + 1 = n$ $j2_{n-1,1,last} = nj2_{n-1} + nj2_{n-1,1} = n$	$nj2_{n-1,1} = \binom{n-j_1}{n-1-m_1-(\eta-1)}$ $= \binom{n-2}{n-2} = 1$	
$n$	2	$j2_{n,2,1} = 1$ $j2_{n,2,2} = nj2_{n,2} = 1$	$nj2_{n,2} = \binom{n-j_1}{n-m_1-(\eta-1)} =$ $= \binom{n-2}{n-2} = 1$	$nj2_n = \binom{n-1}{n-1} = 1$

where  $j2_{m_2, \eta, 1}$  – the minimum sequence number  $j_2$  of element  $y_{m_2, j_2}^n$  in set  $Y_{m_2}^n$ , for which expression (1) is true when element  $y_{m_1, j_1}^n$  is located in the  $\eta^{\text{th}}$  place in tuple  $y_{m_2, j_2}^n$ ;  $j2_{m_2, \eta, \text{last}}$  – the maximum sequence number  $j_2$  of element  $y_{m_2, j_2}^n$  in set  $Y_{m_2}^n$ , for which expression (1) is true when element  $y_{m_1, j_1}^n$  is located in the  $\eta^{\text{th}}$  place in tuple  $y_{m_2, j_2}^n$ .

Example 2.

We have  $n, m_1=1, j_1=2, m_2, j_2, m_1 < m_2$ .

Required to determine if expression (1) is true for element  $y_{m_2, j_2}^n$ .

Decision:

$$nj2_{m_2} = \binom{n - m_1}{m_2 - m_1}, \tag{5}$$

$$nj2_{m_2, \eta=2} = \binom{n - j_1}{m_2 - m_1 - 1}, \tag{6}$$

$$nj2_{m_2, \eta=1} = \binom{n - j_1}{m_2 - m_1}, \tag{7}$$

$$j2_{m_2, \eta=2, 1} = 1, \tag{8}$$

$$j2_{m_2, \eta=2, \text{last}} = nj2_{m_2, \eta=2}, \tag{9}$$

$$j2_{m_2, \eta=1, 1} = nj2_{m_2} + 1, \tag{10}$$

$$j2_{m_2, \eta=1, \text{last}} = nj2_{m_2} + nj2_{m_2, \eta=1} \tag{11}$$

Answer:

Expression (1) is true for expression  $(j2_{m_2, 2, 1} \leq j_2 \leq j2_{m_2, 2, \text{last}}) \vee (j2_{m_2, 1, 1} \leq j_2 \leq j2_{m_2, 1, \text{last}})$ .

The algorithms considered in example 1 and example 2 relate to algorithms of linear execution time  $O(n)$ .

**Conclusions.** In this article, we considered certain properties of data structure “m-tuples based on ordered sets of arbitrary cardinality.” These properties follow the rules of logic when forming ordered ascending elements of the given data structure. It allows us to determine interdependencies between m-tuples by their location in the structure defined by a pair of indices (j, m) without executing computing algorithms. At the same time, there are certain operations for Boolean elements. These are the operation of intersection, union and membership. For these operations the functional dependency of the time estimation to obtain the results using the meth-

ods described changes from cubic  $O(n^3)$  to linear  $O(n)$  depending on cardinality  $n$  of basis set  $X$ . This approach allows us to speed up data processing time significantly and minimize the time and computing resources to real time.

#### REFERENCES

1. Min Chen. Big Data. Related Technologies, Challenges, and Future Prospects. / Min Chen, Shiwen Mao, Yin Zhang, Victor C.M. Leung // Springer. – 2014. – 100 pp.
2. Petrova S. Ju. The Problem of Navigation in Big Data. / S. Ju. Petrova // Experimental and Theoretical Studies in Modern Science: Proceedings of the IIIrd International Scientific Conference, Novosibirsk: SibAK. – 2017. – P. 5-8. (in Russian).
3. Soloviov S. Ju. Mathematical Methods and Principles of Building Automated Knowledge Engineering Systems. / S. Ju. Soloviov // Dissertation of Doctor of Technical Sciences. – 1996. – 272 pp. (in Russian).
4. Totsenko V.G. Expert Diagnostic System by Unstructured Features / V.G. Totsenko, E.A. Petrova, A.A. Chernyavskaya // Registration, Storage and Data Processing. – 2005. – Issue 7(2). – P. 94-103. (in Russian).
5. MacGregor J. Monitoring, Fault Diagnosis, Fault-Tolerant Control and Optimization. / J. MacGregor, A. Cinar // Data Driven Methods. Computers & Chemical Engineering. – 2012. – Issue 47. – P. 111-120.
6. Syrotkina O. The Application of Specialized Data Structures for SCADA Diagnostics / O. Syrotkina // System Technologies. Regional Interuniversity Collection of Scientific Papers, Dnipropetrovsk – 2015. – Issue 4. – P. 72-81.
7. Syrotkina O. Evaluation to Determine the Efficiency for the Diagnosis Search Formation Method of Failures in Automated Systems / O. Syrotkina, M. Alekseyev, O. Aleksieiev // Eastern-European Journal of Enterprise Technologies. – 2017. – Vol. 4, Issue 9 (88). – P. 59-68.
8. Syrotkina O. Methods of Minimizing Computing Resources when Processing “Big Data”. / O. Syrotkina, I. Udovik, M. Alekseyev. // Fifth International Conference "High Performance Computing" HPC-UA 2018 (Ukraine, Kyiv, October 22-23, 2018). P. 151 – 157.

Received 04.03.2019.

Accepted 11.05.2019.

#### **Графоаналитические методы обработки «больших данных» на основе анализа свойств их структурной организации**

*В статье рассмотрены вопросы создания математических методов оптимизации временных и вычислительных ресурсов при обработке «больших данных». Одно из направлений для решения данной проблемы – создание NoSQL систем, достоинством которых является гибкость моделей данных, возможности горизонтального масштабирования и параллельной обработки, быстрота получения результатов. С точки зрения методов анализа «больших данных» дальнейшее развитие получили машинное обучение, искусственный интеллект, распределенная обработка потоков и событий, визуальные методы исследования данных.*

*Целью исследования является разработка математических методов обработки «больших данных» на основе системного анализа свойств структуры организации данных типа «т-арные кортежи на основе упорядоченных множеств произвольной мощности».*

СОД типа « $m$ -арные кортежи на основе УМПМ» - булеан, упорядоченный по правостороннему перебору элементов базового множества мощности  $n$  от нижней границы возможного изменения значения индекса для каждого элемента кортежа к верхней границе. Сформулированы некоторые свойства исследуемой СОД, которые являются следствием логических правил формирования данной структуры. Описаны основанные на этих свойствах математические методы. Графы булеана проиллюстрированы рисунками. Очерченные вершины графа отвечают задекларированным свойствам СОД. Для определения этих элементов булеана выведены аналитические зависимости. Для этих элементов булеана не требуется выполнение алгоритмов, реализующих исследуемые операции пересечения, объединения, принадлежности операндов, т.к. искомым результатом определен в самом свойстве СОД.

Рассмотренные в статье свойства СОД типа « $m$ -арные кортежи на основе УМПМ» позволяют определять некоторые взаимозависимости между  $m$ -арными кортежами по их местоположению в структуре, определяемому парой индексов  $(j, m)$ , без исполнения вычислительных алгоритмов. При этом оценка времени получения результатов изменяется с кубической на  $O(n^3)$  на линейную  $O(n)$ .

### **Графоаналітичні методи обробки «великих даних» на основі аналізу властивостей їх структурної організації**

У статті розглянуто питання створення математичних методів оптимізації тимчасових та обчислювальних ресурсів при обробці «великих даних». Один з напрямків для вирішення даної проблеми є створення NoSQL систем, перевагою яких є гнучкість моделей даних, можливості горизонтального масштабування та паралельної обробки, швидкість отримання результатів. З точки зору методів аналізу «великих даних» подальший розвиток отримали машинне навчання, штучний інтелект, розподілена обробка потоків і подій, візуальні методи дослідження даних.

Метою дослідження є розробка математичних методів обробки «великих даних» на основі системного аналізу властивостей структури організації даних типу « $m$ -арні кортежі на основі впорядкованих множин довільної потужності».

СОД типу « $m$ -арні кортежі на основі ВМДП» – булеан, упорядкований за правобічним перебором елементів базової множини потужності  $n$  від нижньої межі можливої зміни значення індексу для кожного елемента кортежу до верхньої межі. Сформульовано деякі властивості досліджуваної СОД, які є наслідком логічних правил формування даної структури. Описано засновані на цих властивостях математичні методи. Графи булеана проілюстровані рисунками. Окреслені вершини графа відповідають задекларованим властивостям СОД. Для визначення цих елементів булеана виведені аналітичні залежності. Для цих елементів булеана не потрібно виконання алгоритмів, що реалізують досліджувані операції перетину, об'єднання та належності операндів, тому що шуканий результат визначений у самій властивості СОД.

Розглянуті в статті властивості СОД типу « $m$ -арні кортежі на основі ВМДП» дозволяють визначати деякі взаємозалежності між  $m$ -арними кортежами за їх розташуванням у структурі, що визначається парою індексів  $(j, m)$ , без виконання обчислювальних алгоритмів. При цьому оцінка часу отримання результатів змінюється з кубічної на  $O(n^3)$  на лінійну  $O(n)$ .

**Сироткина Е.И.** - к.т.н., доцент кафедры программного обеспечения компьютерных систем НТУ «Днепропетровская политехника».

**Алексеев М.А.** - Декан факультета информационных технологий, д.т.н., профессор НТУ «Днепропетровская политехника».

**Удовик И.М.** - Заведующая кафедрой программного обеспечения компьютерных систем, к.т.н., доцент НТУ «Днепропетровская политехника».

**Сироткіна О.І.** - к.т.н., доцент кафедри програмного забезпечення комп'ютерних систем НТУ «Дніпровська політехніка».

**Алексеев М.А.** - Декан факультету інформаційних технологій, д.т.н., професор НТУ «Дніпровська політехніка».

**Удовик І.М.** - Завідувач кафедри програмного забезпечення комп'ютерних систем, к.т.н., доцент НТУ «Дніпровська політехніка».

**Syrotkina O.** - Associate Professor of the Software Engineering Department Dnipro University of Technology.

**Alekseyev M.** - Dean of the Faculty of Information Technology Dnipro University of Technology.

**Udovyk I.** - Head of the Software Engineering Department Dnipro University of Technology.