Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 1 (2012) 2311–2317
Procedia Computer Science
www.elsevier.com/locate/procedia

International Conference on Computational Science, ICCS 2010

On-line analytical processing based on formal concept analysis

A.V. Korobko[a], T.G. Penkova[a]

[a]*Institute of Computational Modelling of the Siberian Branch of the Russian Academy of Sciences, Akademgorodok, 50/44, 660036, Krasnoyarsk, Russia*

Abstract

An approach to analytical decision making support based on integration of OLAP technology and Formal Concept Analysis is suggested in this paper. The base of domain OLAP modeling is considered. The construction of an analytical model with expert knowledge as a lattice of formal cube-concepts is described formally. The use of the suggested approach for constructing an analytical model of municipal procurement procedures is represented. The constructed integral analytical model of domain includes all possible combinations of analyzed objects and gives opportunity ad-hoc manipulation of them.

Keywords: Decision making support; On-line analytical processing; Formal Concept Analysis; municipal procurement; OLAP.

1. Introduction

The effectiveness of administrative resources government depends on the way how opportunely analytical information will be given. Analytical systems based on On-Line Analytical Processing (OLAP) are widely used for decision making support [1, 2, 3, 4]. OLAP-system “Analytic” has been developed for on-line analytical support of governmental decision making by the Applied Informatics Department of the Institute of Computational Modeling of the Siberian Branch of Russian Academy of Science [5]. “Analytic” system is used by many regional governmental organizations of Krasnoyarsk region. For example, in Krasnoyarsk Region it is used by the Ministry of Health to control health care and drug provision, by the Ministry of Social Policy to govern in the field of social services and social support and by the Department of Emergency for the purpose of crisis management. In Krasnoyarsk city it is used by the Department of Economics to plan the social economic development of territories and to estimate the efficiency of budget implementation through the Department of Municipal Procurement [6, 7, 8, 9].

The maxims of OLAP technology have been formulated by E. Codd in 1993 [10]. The realization of those maxims depends on current level of information technology development. Using OLAP means data organized as easy-to-understand and easy-to-use data model, which consists of multidimensional cubes. The OLAP-cube is an abstract representation of a projection of the RDBMS (Relational Database Management System) relation [11]. Obviously, it is a result of query executions for the needed tables of the database. For improving analytical processing, data warehouse is usually used [12, 13]. To construct an OLAP-cube by himself the user needs to know the schema of the database tables and fields compatible to the domain terms and algorithm of complex measures calculation. Thus, the decision maker has to write a specification for the required information and to send it to the

1877-0509 © 2012 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
doi:10.1016/j.procs.2010.04.259

engineer. As a result, this engineer has to construct an OLAP-cube for solving each problem.

The most prevalent way of decision making support is preliminary preparations of analytical cubes [1, 5, 14, 15]. It allows the decision maker to focus his mind to an analytical process with the cube: pivoting, drill-down, roll-up. But, the necessity for additional measures or dimensions in the cube leads to reconstructing the cube and spending time.

A great number of cubes are accumulated during the use of an OLAP application. Such a situation can be represented by a fragmentary analytical model of the domain. The union of the prepared cubes can contain all measures and dimensions of the domain, but it cannot satisfy all the needed queries of the decision maker.

2. OLAP-modeling of domain

OLAP-modeling of the domain is constructed based on the integral schema of multidimensional data. The model score is a vocabulary of special terms. To find all the used terms of the field we should interview the end user, should determine his common requirements and investigate reports. The quantity of the determined terms is a set of analyzed objects, which can be used for OLAP-cube construction. In common case, we can characterize the cube as a function.

The offered analytical model of the OLAP-cube can be represented as a pair:

$G = (D, F)$, where

$D = \{d1, d2, \ldots, dn\}$ is a set of the cube dimensions and

$F = \{f1, f2, \ldots, fm\}$ is a set of the cube measures (facts).

The measure is a numerical characteristic of the analyzed process and a dimension is an array of values which belongs to the same data type and characterizes a structural property of the domain. A set of dimensions forms an axis of the cube. There are analytical measures in a cell of the cube.

The quantity and content of cubes depend on the quantity of solving problems in the fragmental analytical model. This approach leads to irrational grouping measures and dimensions. To solve a new problem we have to construct a new OLAP-cube. It means frequent model changing. To avoid these problems the analytical model needs to include all existent measures and dimensions and all their possible combinations. In this paper the possibility of combination is defined as the comparability of each pair of measure and dimension (they can be processed together). We suggest the integral approach to analytical model construction, which gives an opportunity to identify all the compatible dimensions for each measure and to use them for ad-hoc requests.

To obtain information about objects of the domain and their properties, about algorithms of analytical measures calculation and relationships between object of analysis we should use an expertise. Hereby, the urgent issue for us is to determine the structural features of the analyzed object, to group them by their comparability and to find special relations between these groups. In this way we can summarize that these groups are more appropriate to execute our request and to process data about a current problem, being solved.

3. An analytical model based on formal concept analysis

The terms of domain represent a set of objects of two types: some objects are the numeric characteristics of the considered process, other objects are the analysis aspects. To construct an integral analytical model let's consider a set of all existent measures $F = \{f1, f2, \ldots, fm\}$ of the object domain and a set of all possible dimensions $D = \{d1, d2, \ldots, dn\}$. The relation of comparability can be identified as R. Strictly, we can say that $(di, fj) \in R$ if the i-th measure can be processed with the j-th dimension.

To provide simultaneous measures processing with the maximum number of dimensions it is needed to determine groups of equidimensional measures, which have common structural features relative to comparability R. In this case, it is advisable to use the methods of binary clustering, where one cluster objects have common properties description [16]. One of the most appropriate methods for this is the method of data analysis based on formal concepts and concept lattices.

Formal Concept Analysis (FCA) is introduced by R. Wille in 1981 [17, 18]. The method is based on understanding the world in terms of objects and attributes. The formal context is a triplet $K = (G, M, I)$, which consists of set G, set M and relation $I \subseteq G \times M$. The elements of set G are objects of the context; the elements of set M are attributes of the context. A binary relation I between G and M (described by gIm) indicates when object $g \in$

G, has attribute $m \in M$. A formal context is often described by a cross table.

For set $A \subseteq G$ and for set $B \subseteq M$ it is defined that:

$A' = \{m \in M \mid gIm \text{ for all } g \in A\}$ (all attributes in M shared by the objects of A);

$B' = \{g \in G \mid gIm \text{ for all } m \in B\}$ (all objects in G that have all the attributes of B).

The formal concept of the formal context is defined by derivation operators as pair (A, B) with $A \subseteq G$, $B \subseteq M$, $A = B'$, $B = A'$. A is called an extent, and B is called an intent of concept (A, B).

For creating domain analytical model based on formal concept analysis, the formal context is defined as triplet (F, D, R). The elements of set F are measures; the elements of set D are dimensions. Binary relation R between F and D is the relation of comparability. Thus, pair (A, B) is a formal concept, where A is a set of equidimensional measures, which are processed with all dimensions of B. In terms of OLAP technology, the formal concept is an analytical multidimensional cube, which is complete to add equidimensional measures and compatible dimensions.

The set of all formal concepts of the formal context is ordered by the subconcept-superconcept relation. For two concepts (A1, B1) and (A2, B2) this order is formalized as: $(A1, B1) \leq (A2, B2)$: $\Leftrightarrow A1 \subseteq A2$ ($\Leftrightarrow B2 \subseteq B1$). (A1, B1) is called a subconcept of (A2, B2), and (A2, B2) is called a superconcept of (A1, B1). The set of all concepts together with subconcept-superconcept relations forms a complete lattice, which is called a concept lattice [19]. In terms of OLAP technology, the subconcept-superconcept relation for domain analytical model is a subcube-supercube relation. It means that a parent cube set of measures includes a child cube set of measures and a child cube set of dimensions includes a parent cube set of dimensions. These concept lattice features (cube lattice features) enable adaptive manipulations with all existent domain objects of analysis for an analytical experiment made by the decision maker.

4. Analytical model of municipal procurement

Let's consider constructing the analytical model based on FCA for the municipal procurement.

The municipal procurement is a procedure of goods, works and services acquisition. This procedure is organized by the municipal government and is financed by the municipal budget for meeting the municipal needs [20]. The municipal procurement is the most important administrative form of territorial management.

Municipal procurement includes three stages:

Planning is a process of distribution of budget and requests.

Conduct procurement is a process of auctions preparation and realization.

Control is a process of contraction and controlling contract execution.

Estimation of the municipal procurement efficiency at every stage is connected with solving different analytical problems, such as:

- Analysis of municipal needs register;
- Analysis of requests;
- Analysis of auctions results;
- Repayment Monitoring;
- Analysis of the supplier activity;
- Budget implementation efficiency estimation;
- Contract execution monitoring;
- Government activity efficiency estimation;
- other problems.
- Solving the analytical problems is connected with the following queries realization:
- Budget was pre-arranged to meet the municipal needs;
- Procurements (competitions, quotations, auctions) were conducted in a year (quarter, month);
- Procurements (competitions, quotations, auctions) were conducted for goods, works and services acquisition;
- Threshold price and contract price for goods, works and services acquisition were presented;
- Lots were frustrated for various reasons;
- Proposals were submitted by suppliers, who have a privilege category;
- Proposals were submitted for goods, works and services acquisition;
- Contracts were made by the suppliers in a year (quarter, month);
- Contracts were made by the suppliers, who have the first place (second place, third place, etc.);

- Non-execution contracts and the budget of these contracts were presented;
- other queries.

By analyzing the queries we can find a set of domain terms. Then, we can define the municipal procurement analysis objects based on the found domain terms: a set of measures (procurement count, lot count, proposal count, contract count, etc.) and a set of dimensions (procurement activity, product category, central purchasing body, lot state, proposal state, contract status, month, year and etc.).

The formal context is established from the comparability relation between measures and dimensions in expert opinion. Fig. 1 shows the context of municipal procurement. The described context is given by measures set F = {*Procurement count*, *Lot count*, *Proposal coun*t, *Contract count*, *Threshold price*, *Contract price*, *Budget*, *Proposal guarantee*, *Contract guarantee*} and dimensions set D = {*Procurement activity*, *Procurement state*, *Product category*, *Procurement privilege category*, *Central purchasing body*, *Contracting authority name*, *Contracting authority type*, *Lot state*, *Proposal status*, *Proposal state*, *Place*, *Supplier name*, *Supplier category*, *Ownership type*, *Sponsor type*, *Term of payment*, *Procurement registration number*, *Procurement name*, *Lot name*, *Contract status*, *Record type in contract register*, *Record state in contract register*, *Contract monitoring stage*, *Product classification*, *Item on budget*, *Register of needs*, *District*, *Month*, *Quinter*, *Year*, *Government agent*}. Using numbers and letters as abbreviations, we also can write $F = \{f1, f2, f3, f4, f5, f6, f7, f8, f9\}$ and $D = \{d1, d2, d3, d4, d5, d6, d7, d8, d9, d10, d11, d12, d13, d14, d15, d16, d17, d18, d19, d20, d21, d22, d23, d24, d25, d26, d27, d28, d29, d30, d31\}$. Relation R is represented by the crosses which can formally be expressed by $R = \{(f1, d1), (f1, d2), (f1, d3), (f1, d4), (f1, d5), (f1, d14), (f1, d14), (f1, d24), (f1, d25), (f1, d27), (f1, d28), (f1, d29), (f1, d30), (f1, d31), \ldots, (f9, d30)\}$.

Formal concepts are formed based on the established context. Either of the concepts is a complete analytical cube (OLAP-cube). In our example, the formal context contains 44 formal concepts. For example, the measures of set $A = \{f1, f2, f4, f7\}$ can be commensurable only with dimensions of set $B = \{d1, d3, d4, d5, d6, d7, d15, d24, d25, d28, d29, d30\}$ at the same time. Therefore, pair (A, B) is a formal concept and is a complete analytical cube for solving group of analytical problems of the municipal procurement. Fig. 2 shows a concept lattice as an integral analytical model of the municipal procurement. This cube lattice gives means for the adaptive manipulation with all the analyzed objects and it absorbs the maximum collection of analytical problems of the municipal procurement.

Conclusion and Future Work

Using FCA can increase the efficiency of on-line analytical processing. Expert knowledge about domain objects (measures and dimensions), their properties, relations, algorithms of the analytical measures calculation is represented as a lattice (a cube lattice). The knowledge representation is used for the manipulation with all the analyzed objects at the same time. The constructed integral analytical model of domain includes all possible combinations of analyzed objects. Therefore, the suggested approach improves the government decision making process.

The future research will be connected with solving the following problems:

- Developing the algorithm to form concepts presented as cubes and a concept lattice, presented as a cube lattice.
- Developing the storage model for cubes and a cube lattice.
- Developing the algorithm to determine the optimal concept (OLAP-cube) for an analyzed problem.
- Developing the logical inference rules for the adaptive manipulation support of the analyzed objects.
- Developing the tools for analytical decision making support based on FCA.

Acknowledgements

This paper was supported by a grant for leading scientific schools (No: NSh-3209.2010.9.), Federal Agency of Science and Innovation and Council grants, the President of Russia.

		f1 Procurement count	f2 Lot count	f3 Proposal count	f4 Contract count	f5 Threshold price	f6 Contract price	f7 Budget	f8 Proposal guarantee	f9 Contract guarantee
d 1	Procurement activity	×	×	×	×	×	×	×	×	×
d 2	Procurement state	×				×	×			
d 3	Product category	×	×	×	×	×	×	×		
d 4	Procurement privilege category	×	×		×	×	×	×		
d 5	Central purchasing body	×	×	×	×	×	×	×		
d 6	Contracting authority name	×	×	×	×	×	×	×		
d 7	Contracting authority type	×	×	×	×	×	×	×		
d 8	Lot state		×	×		×	×		×	
d 9	Proposal status			×						
d 10	Proposal state			×					×	
d 11	Place			×	×				×	×
d 12	Supplier name			×	×				×	×
d 13	Supplier category			×			×		×	×
d 14	Ownership type			×	×					
d 15	Sponsor type	×	×	×	×	×	×	×		
d 16	Payment method	×	×	×	×	×	×			
d 17	Procurement identifier		×	×		×	×		×	×
d 18	Procurement name		×	×		×	×		×	×
d 19	Lot name			×		×	×		×	×
d 20	Contract status		×		×		×			×
d 21	Record type in contract register				×					
d 22	Record state in contract register				×					
d 23	Contract monitoring stage				×					
d 24	Product classification	×	×	×	×			×		
d 25	Item on budget	×	×		×			×		
d 26	Register of needs					×	×	×		
d 27	District	×	×	×						
d 28	Month	×	×	×	×	×	×	×	×	×
d 29	Quinter	×	×	×	×	×	×	×	×	×
d 30	Year	×	×	×	×	×	×	×	×	×
d 31	Government agent	×								

Fig. 1. Context of municipal procurement

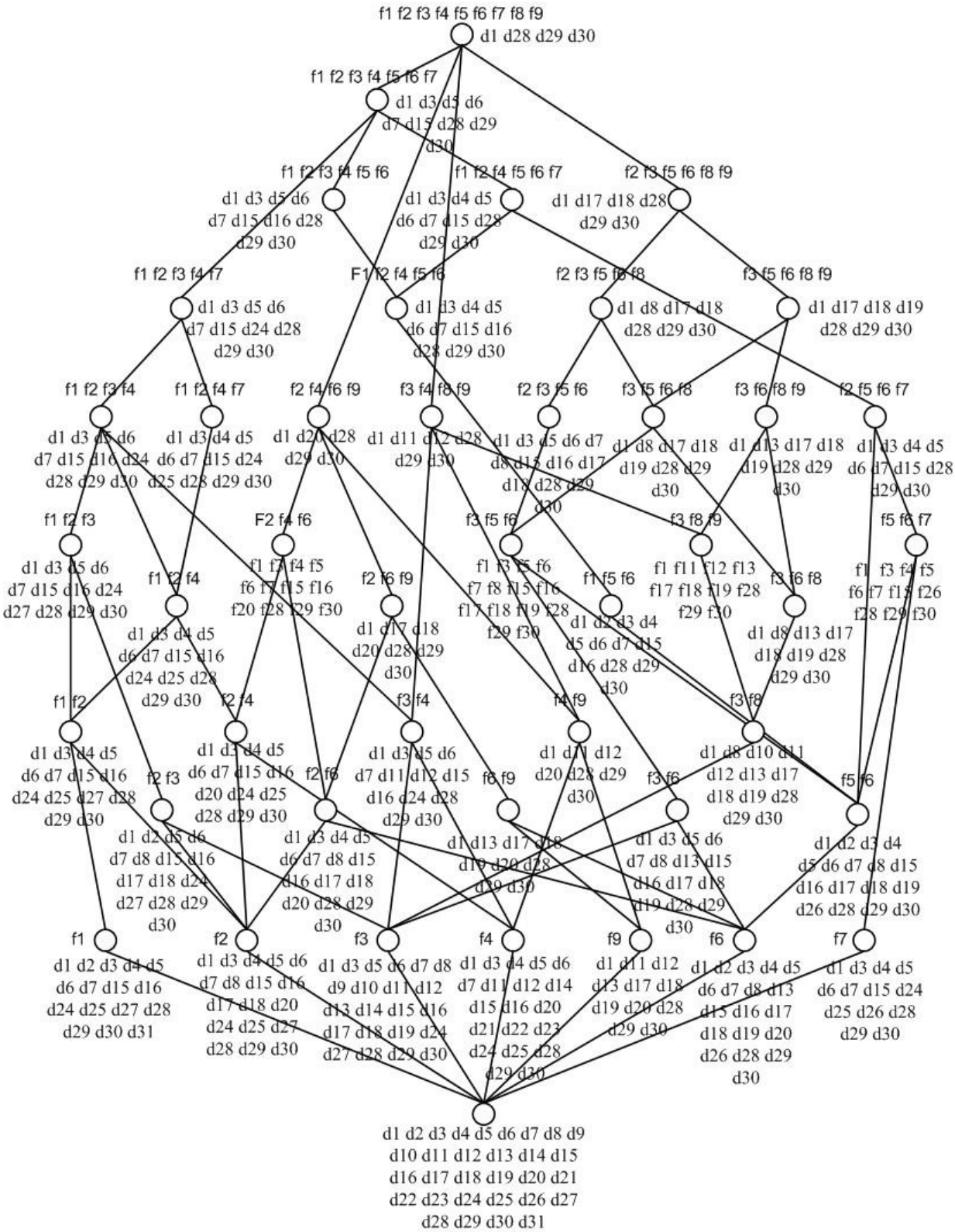

Fig. 2. Concept lattice of municipal procurement

References

1. L. Nozhenkova, A. Evsukov and A. Nozhenkov, Management techniques and GIS modeling of OLAP technology, Journal of Siberian Federal University. Engineering & Technologies №1, 2009. pp. – 49-58.
2. O. Vasilecas and A. Smaizys, Business rule based data analysis for decision support and automation, Proc. International Conference on Computer Systems and Technologies, Vol. II, 2006.
3. S. Palaniappan and C. Ling, Clinical Decision Support Using OLAP With Data Mining, International Journal of Computer Science and Network Security, VOL.8, No.9, 2008. – pp. 290-296.
4. A. Javed, M. Shaikh and B. Bhatti, Conceptual Model for Decision Support System Based Business Intelligence OLAP Tool for Universities in Context of E-Learning, Proc. World Congress on Engineering and Computer Science, 2008.
5. Yu. Dudina and P. Ishenin, Usage of OLAP-means of the System «Analytics» for the Problem of Health Protections, 9th International Scientific and Practical Conference of Students, Post-Graduates and Young Scientists «Modern Techniques and Technologies», 2003. – pp. 151-152.
6. A. Korobko and T. Penkova, OLAP-modeling of municipal procurement automation support problem, Proc. «Knowledge and Ontology *ELSEWHERE*» 17th International Conference on Conceptual Structures (ICCS'09), Moscow, 2009. – pp. 87-91.
7. O. Baryshnikova, A. Korobko and A. Nozhenkov, Technological and instrumental tools for consolidation and analytical processing automation in the social sphere, Proc. 10th Russian Scientific Conference "Informatization Problems of Region", Krasnoyarsk, 2007. – pp. 24-27.
8. V. Nicheporchuk, A. Lepikhin, V. Moskvichev and K. Simonov, Estimation method of risk parameters for atmosphere pollution of Krasnoyarsk region , International Conference on Computational information Technologies for Environmental Sciences «CITES-2009», Krasnoyarsk, 2009.
9. V. Nicheporchuk, Investigation of ecological danger for territories of Siberia, 18th International Symposium ECOLOGY & SAFETY, Burgas, Bulgaria, 2009.
10. E. Codd, S. Codd and C. Salley, Providing OLAP to User-Analysts: An IT Mandate, Arbor Software Corp. Papers, 1996.
11. J. Gray, A. Bosworth, A. Layman and H. Priahesh, Data Cube: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab and Sub-Totals, Proc. 12th International Conference on Data Engineering. IEEE, 1995. – pp. 152-159.
12. W. Inmon, Willey John & Sons, Building the Data Warehouse, New York, 1992.
13. E. Sperley, Enterprise data warehouse: planning, building and implementation. Vol.1 Prentice Hall PTR, Upper Saddle River, NJ 07458, 2001. – 400 p.
14. E. Thomsen. OLAP solutions: building multidimensional information systems. Second Edition. – New York, USA: Willey & Sons inc., 2002. – 608 p.
15. D. Mandy On OLAP / R. Kimball // Intelligent Enterprise. – 2003. – N18.
16. S. Kedrov, S. Kuznetsov, Study of Groups of Web Users with Methods Based on Formal Concept Analysis and Data Mining, Business-informatics, №1, 2007. – pp. 45-51.
17. R. Wille, Restructuring Lattice Theory: an approach based on hierarchies of concept. / Ordered sets / editor I. Rival. –Reidel, Dordrecht-Boston, 1982. – pp. 445-470.
18. B. Ganter, R. Wille, Formal Concept Analysis: mathematical Foundations, Springer-Verlag, Berlin Heidelberg New York, 1999.
19. G. Birkhoff, Lattice theory, Moscow: Nauka, 1984. – 568 p.
20. A. Hramkin, O. Vorobeva, B. Vdovin, A. Volosatova and B. Ermakov, Public procurement. – Moscow: Jurisprudence, 2006. – 312 p.