Formulation of the Problem of Structural and Parametric Synthesis of Electronic Document Management System of Research and Education Institution

Mikhail Krasnyanskiy
D.Sc., Professor, Acting Rector, Tambov State Technical University, Tambov Oblast, Russia.

Andrey Ostroukh
D.Sc., Professor, Professor of Chair ‘Automated control systems’ State Technical University – MADI, Russia.

Sergey Karpushkin
D.Sc., Professor, Professor of Chair ‘Computer-integrated systems in mechanical engineering’ Tambov State Technical University, Tambov Oblast, Russia.

Artyom Obukhov
Undergraduate, Tambov State Technical University, Tambov Oblast, Russia.
E-mail: obuhov_art@mail.ru.

Abstract

The paper gives the formulation of the problem of structural and parametric synthesis of electronic document management system (EDMS). We propose a mathematical model of the EDMS that accounts for key elements and permits a set-theoretic analysis of document management processes. We provide a detailed description of documentation, operations and conditions of their implementation; we consider hardware and software characteristics of the EDMS. The solution of the problem of structural-parametric synthesis will increase the economic efficiency of the developed EDMS, meet the requirements of productivity, quality, and profitability and satisfy end-user needs.

Keywords: electronic document management, DMS, structural and parametric synthesis, mathematical model of electronic document management system.
INTRODUCTION

Ever-increasing volumes of information, requirements to reliability of storage, processing and transmission speed, the need to shift from paper-based to electronic document management necessitates the creation and implementation of EDMS. Implementation of a large-scale, highly organized EDMS is a laborious and lengthy process requiring a significant amount of resources. Its creation and adaptation for specific tasks, development of universal design solutions using advanced methods becomes an urgent task of information technology, which cannot be solved without the use of methods of system analysis and mathematical modeling [1-6].

As this problem belongs to the class of extreme combinatorial problems, we will use structural and parametric synthesis to determine the structure and values of object parameters, ensuring the optimality of the developed system to meet certain criteria and required constraints.

Through the analysis of approaches to modeling document management processes, we identified the main types of models: set-theoretic, graph-theoretic, automatic, functional, descriptor and multi-agent models, and their strengths and weaknesses [1, 7-14]. As a result, the decision was made to construct a mathematical model of EDMS on the basis of a graph-theoretic model of document management. This model will account for the structure of documents, describe the operations that lead to changes in the state of documents, conditions of their implementation subject to the rules of restricting access to information, and allow formalizing the processes of document circulation and interaction, as well as basic software, hardware and structural parameters of the information system.

In this article, we describe the EDMS model, which meets the above criteria and allows solving the problem of structural and parametric synthesis to create an optimal information system in terms of its structure and parameters, with minimum costs of EDMS development, delivery, handling and storage of documents, and availability of documents in due time [4, 9, 12]. The main parameters include the number of documents, deadlines for their delivery, the structure of document management in the organization, software and hardware. Input and output variables as well as conditions of document management process are shown in Fig. 1.

Fig. 1 Document management processes of research and education institution.

To solve this problem it was necessary to find a range of permissible values of the
studied characteristics, select an optimization criterion, formalize processes occurring in the object, and present them in the form of a mathematical model.

**EDMS Mathematical Model**

Having analyzed document management objects, document interaction and movement processes, and the existing document automation systems, we determined the necessary components of the EDMS mathematical model, which can solve the problem of structural and parametric synthesis. We presented the EDMS mathematical model as a tuple \([1, 7-14]\)

\[
EDMS = (S, HW, SW, V_s, E_s, Q_s)
\]  

(1)

where \(EDMS\) is EDMS model; 
\(S\) is the structure of EDM of the organization; 
\(HW = (ST, VST, QST, CST)\) is a set of EDMS hardware characteristics, including information storage data \(ST\), costs \(VST\), volume \(QST\) and capacity \(CST\); 
\(SW = (SD, SM, SR)\) is a set of EDMS software characteristics, including possible combinations of design tools, \(SD\), EDMS module structure \(SM\) and program implementation \(SR\); 
\(V_s\) is annual costs of EDMS; 
\(E_s\) is document management performance measurement; 
\(Q_s\) is EDMS quality evaluation.

Under the structure of the electronic document organization, we mean a formalized representation of a tuple set model of document management objects that change their status as a result of user operations at some point in time under a number of external and internal influences [9]:

\[
S = (U, P, E, O, T, G)
\]  

(2)

where \(U = \{u_i | i = 1..I\}\) is a set of document management objects, \(I\) is total number of objects;

\(P = \{p_q | q = 1..Q\}\) is a set of users, \(Q\) is total number of users;

\(T\) is a set of discrete points in time represented by natural numbers;

\(E\) is a set of influences, both external and internal;

\(O = \{o_l | l = 1..L\}\) is a set of operations on the objects, \(L\) is total number of operations;

\(G = \{G_i(C_i, O) | i = 1..I\}\) is a set of structures of information flows, each of which describes the object life cycle \(u_i\), sequential change of states \(C_i\) through a set of operations \(O\).
Object $u_i$ has a set of states, each of which defines the contents of the object in a certain period of its life cycle:

$$u_i \rightarrow C_i,$$  \hspace{1cm} (3)

where $C_i = \{c_{ij} | j = 1..J_i\}$ is a set of states of object $u_i$, $J_i$ is the number of such states. Each state is defined as a tuple from the set of object attributes and their values in a specified period of time:

$$c_{ij} = ((a_{ijn}, d_{ijn}) | a_{ijn} \in A_i, d_{ijn} \in D_{ij}, n = 1..N_i, T_{c_{ij}})$$  \hspace{1cm} (4)

where $A_i = \{a_{ijn} | n = 1..N_i\}$ is a set of attributes of object $u_i$ with the corresponding elements of the set of values of attributes $D_i = \{D_{ij} | j = 1..J_i\}$ for each state $c_{ij}$:

$$D_{ij} = \{d_{ijn} | n = 1..N_i\},$$  \hspace{1cm} (5)

where $N_i$ is the number of object attributes; $T_{c_{ij}}$ is a set of discrete moments of time $t$, for the state $c_{ij}$.

The influences are divided into external $EE$ and internal $IE$. External influences $EE = \{ee_w | w = 1..W_e\}$ include ministry decrees, new laws and standards, third-party orders and other external influences. Internal influences $IE = \{ie_w | w = 1..W_i\}$ are caused by external influences or generated within the organization internally.

The influence is directed to achieving a particular result, namely, a document or a set of documents, which satisfy the conditions restricted by the influence [10]. The influence can be formalized as a tuple:

$$ee_{we} = (U^*, p_q, TU), ie_{we} \rightarrow \{ie_w | w = 1..W_i, ie_w \in IE\},$$

$$ie_{we} = (U^*, P^*, O^*, TU),$$

$$ee_{we} \in EE, U^* \in U, p_q \in P, P^* \in P, O^* \in O, TU \in T,$$  \hspace{1cm} (6)

where $ee_{we}$ is external influence directed at producing the result by a person in charge $p_q$ as a set of objects $U^*$ within a predetermined time frame $T^*$;

$ie_{we}$ is internal influences generated by influence $ee_{we}$ and directed at producing objects $U^*$ by a subset of users $P^*$ through a set of operations $O^*$ in a time $T^*$.

In order to achieve results in the form of final documents, users perform different operations with documents. Operations performed on object $u_i$ and executed by user
Formulation of the Problem of Structural and Parametric Synthesis of

$p_q$ in the moment of time $t$, in a general form can be represented by the function [9]:

$$o_t(u_i,p_q,t): u_i \rightarrow u_m \rightarrow c_{m_k} \rightarrow c_{m_k}(\{a_{i,j}, d_{j,m}\}, T_{c_{m_k}}) \rightarrow ((a_{m_n}, d_{m,k}), T_{c_{m_k}}),$$

$$t \rightarrow t + t \in T_{c_{m_k}}, t + t \in T_{c_{m_k}}, c_{i,j} = \psi(u_i,t), c_{m_k} = \psi(u_m,t),$$

$$7) \quad i \epsilon \omega = (U_*, P_*, O^*, T*), o_t \in O^*, u_m \in U_*, p_q \in P^*, t \in T,$$

$$u_i, u_m \in U, i = \overline{1..I}, l = \overline{1..L}, f = \overline{1..J}, k = \overline{1..J_m}, n = \overline{1..N_i}.$$

Typical document operations include: viewing, editing, deleting, copying, consolidation and compression of states [1.9]. Each operation has to satisfy conditions for its implementation. These conditions can be arranged in three different ways: discrete, role-based and attribute representations, each of which is used depending on the situation, object structure, performed operations and information security requirements [1.15].

Another component of the EDSM mathematical model is evaluation of annual economic costs [16]:

$$V_s = V_{hw} + V_{st} + V_p + V_t + V_d, \quad (8)$$

where $V_s$ is annual economic costs; $V_{hw}$ is equipment costs; $V_{st}$ is data storage costs; $V_p$ is data processing costs; $V_t$ is data delivery costs; $V_d$ is system development costs.

Equipment costs include the costs of all repositories $ST_k$:

$$V_{hw} = \sum_{k=1}^{K} VST_k, \quad (9)$$

$$VST_k = V_{k1} + V_{k2} + V_{k3} + V_{k4} + V_{k5}, \quad (10)$$

$$V_{k1} = 0.01 \cdot (V_h + 0.1 \cdot V_h), \quad (11)$$

$$V_{k2} = m_e \cdot T_e \cdot V_m \cdot k_e, \quad (12)$$

$$V_{k3} = \frac{V_h \cdot N_1 + (0.1 \cdot V_h) \cdot N_2}{100}, \quad (13)$$

$$V_{k4} = N_3 \cdot (V_h + 0.1 \cdot V_h), \quad (14)$$

$$V_{k5} = 0.01 \cdot V_{hw}, \quad (15)$$

where $VST_k$ is the sum of annual operation costs of repository $ST_k$, rubles;
2400

Mikhail Krasnyanskiy et al

$V_{k1}$ is costs of basic and auxiliary materials, rubles;
$V_{k2}$ is power costs, rubles;
$V_{k3}$ is annual depreciation amount, rubles;
$V_{k4}$ is equipment repairs and maintenance costs, rubles;
$V_{k5}$ is other expenses in the amount of 1% of capital expenditures, rubles;
$V_h$ is the cost of basic equipment of repository $ST_k$, 0.1-$V_h$ is the cost of auxiliary equipment, rubles;
$m_c$ is capacity of equipment in $ST_k$, kw;
$T_e$ is operation time of equipment $ST_k$ per annum, hours.

Data storage costs are calculated by the formula:

$$V_{st} = \sum_{k=1}^{K} \frac{VST_k}{QST_k} \cdot \sum_{i=1}^{I} f_{st}(u_i, ST_k) \cdot \sum_{j=1}^{J_i} a_{ij4} \cdot k_{ij},$$

under the constraint on the amount of stored information:

$$\sum_{k=1}^{K} \sum_{i=1}^{I} f_{st}(u_i, ST_k) \cdot \sum_{j=1}^{J_i} a_{ij4} \leq \sum_{k=1}^{K} QST_k.$$

Data processing costs are calculated as follows:

$$V_p = \sum_{i=1}^{I} \sum_{q=1}^{Q} \sum_{l=1}^{L} k_{iql} \cdot \lambda_{iql} \cdot t_{iql} \cdot \frac{v_w}{T_w} \cdot \lambda_{iql},$$

where $k_{iql}$ is the number of operations $o_l$ performed by user $p_q$ on object $u_i$ per annum; $\lambda_{iql}$ is coefficient that indicates that user $p_q$ performed operation $o_l$ on object $u_i$ by the formula:

$$\lambda_{iql} = \begin{cases} 1, & \exists o_l(u_i, p_q, t), \\ 0, & \text{otherwise}; \end{cases}$$
\( t_{iq} \) is the time given user \( p_q \) to perform operation \( o_t \) on object \( u_i \); \( v^w_q \) is monthly salary of user \( p_q \); \( T_w^q \) is the number of working hours per month of user \( p_q \).

Data delivery costs are calculated by the formula:

\[
V_i = \frac{1}{\text{CST}_k} \sum_{j=1}^{J_i} \sum_{i=1}^{I_i} a_{ij} \cdot \eta_{kij} \cdot f_{sd}(u_i, \text{ST}_k),
\]

under constraint on data delivery time:

\[
\frac{1}{\text{CST}_k} \sum_{j=1}^{J_i} \sum_{i=1}^{I_i} a_{ij} \cdot \eta_{kij} \cdot f_{sd}(u_i, \text{ST}_k) \leq \sum_{i=1}^{I_i} \sum_{j=1}^{J_i} \eta_{kij} t_{ui}, \tag{21}
\]

where \( \eta_{kij} \) is the number of deliveries of state \( c_{ij} \) from repository \( \text{ST}_k \);
\( t_{ui} \) is maximum permissible time to work with \( i \) document, \( t_{ui} \in TU \) (\( TU \) is a set of given permissible times);

\( \text{CST}_k \) is capacity of repository \( \text{ST}_k \) calculated by the formula:

\[
\text{CST}_k = S_{Pk} \cdot c_{k1} \cdot c_{k2} \cdot c_{k3} \cdot c_{k4} = [0,1], i=1..3,
\]

where \( S_{Pk} \) is capacity of server network card, \( c_{k1}, c_{k2}, c_{k3} \) are factors determining the estimates of the CPU performance, speed of main memory and main storage device, respectively. Estimates are normalized and determined by comparing hardware specifications; where 1 corresponding to the solution, providing maximum performance, 0 corresponds to the worst performance, leading to failure of the entire system.

The EDMS development costs are calculated as follows:

\[
V_d = V_{dw} + V_{dt} + V_{dt}, \tag{23}
\]

where \( V_{dw} \) is EDMS developers’ wage costs; \( V_{dt} \) is the cost of purchasing licensed development tools; \( V_{dt} \) is costs for technical equipment for the EDMS development. Estimated productivity of document management system depends on the number of documents processed per unit of time taking into account the processing costs. We represent the productivity of document management system as follows:

\[
E_s = \frac{\sum_{j=1}^{J_i} \sum_{i=1}^{I_i} w_{ij}}{\sum_{j=1}^{J_i} \sum_{i=1}^{I_i} t_{ij} w_{ij}}, \tag{24}
\]

where \( w_{ij} \) is the number of documents in the \( i \)-th category; \( J_i \) is the number of
operations on the document in the \( i \)-th category; \( t_{ij} \) is time required to perform \( j \)-th operation to process document of the \( i \)-th category; \( K \) is the total amount of documents.

Document management productivity is directly influenced by the power of hardware information system, the performance of its software components, which is expressed by the following formula:

\[
t_{ij} = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} a_{ij} \cdot n_{ij}}{CST_k} + \tau(u_i, o_j),
\]

where \( \tau(u_i, o_j) \) is the time to perform the \( j \)-th operation to process document of the \( i \)-th category when using the specific software implementation \( SR \), which was produced using the development tools \( SD \) and the modules structure \( SM \).

By the EDMS quality, we understand the result of evaluation of the developed system by a group of experts using certain criteria. The evaluation criteria, which is a list of quality indicators approved by GOST 28195-89 "Evaluation of Software Quality (PC)", are as follows:

\[
Q_s=\{(Q_{ijk}, q_{ijk})\}; 0 \leq q_{ijk} \leq 1,
\]

where \( Q_{ijk} \) is a set of evaluation elements, index \( i \) is a number of factor, index \( j \) is a number of metrics, index \( k \) is a sequence number of the evaluation element, \( q_{ijk} \) is the value of the evaluation element.

We introduce the weighting factors for the evaluation elements at all levels:

\[
\sum_{k=1}^{K} V_{jk} = \text{Const} = 1, \quad \sum_{j=1}^{J} V_{ij} = \text{Const} = 1, \quad \sum_{i=1}^{I} V_{i} = \text{Const} = 1.
\]

The EDMS quality evaluation is calculated by the formula:

\[
Q_s = \sum_{i=1}^{I} (V_i \cdot Q_i), \quad Q' = \sum_{j=1}^{J} (V_j \cdot Q_j) \cdot Q_s = \sum_{k=1}^{K} q_{ijk} \cdot V_{jk}.
\]

The EDMS quality evaluation is influenced by stability and performance of information system, which is ensured by the choice of specific EDMS software implementation \( SR \), produced using the development tools \( SD \) and the modules structure \( SM \).
Formulation of the Problem of Structural and Parametric Synthesis of 

\[(SD, SM, SR) \rightarrow \{q_{uk}\}.\]  

(29)

It should be noted that the highest quality software is not always the best choice for the end user. When implementing new information systems, it is essential to take into account user preferences and choose development tools, interface forms and interaction tools that are convenient to use, intuitive and do not require a long period of adaptation and retraining. In a formalized form, it can be represented as follows:

\[\varphi(S_{D'}, S_{R'}) = [0;100],\]  

(30)

where \(\varphi(S_{D'}, S_{R'})\) is function returning a percentage of the number of users of the organization that can interact with EDMS under specific software implementation \(S_{R'}\), produced using the development tools \(S_{D'}\), of the total number of the EDMS users.

For the formulation of the problem of structural and parametric synthesis of the EDMS it is necessary to select optimization criteria [17,18]. The analysis of scientific literature devoted to document management optimization showed that it is not sufficient to consider only economic costs in the design of information systems, as this results in the biased evaluation of the system in terms of money expenditures [1]. In other words, such an approach does not take into account such important characteristics of the system as its speed, convenience, compliance with the user requirements; therefore, it is necessary to consider them as prerequisites for solving the problem.

Thus, for the evaluation of variables and optimization of the EDMS we use the criterion of economic efficiency, having the form:

\[H = \frac{V_0 - V_s}{V_0},\]  

(31)

\[V_0 = V_h^0 + V_{st}^0 + V_p^0 + V_{it}^0 + V_d,\]  

(32)

\[0 \leq H \leq 1,\]  

(33)

where \(V_0\) is maximum allowable economic costs including the costs of the equipment \(V_h^0\), storage costs \(V_{st}^0\), обработки \(V_p^0\), obtaining information \(V_{it}^0\) before the EDMS implementation and EDMS development costs \(V_d\).

If there were no information systems for document management before the EDMS implementation, we will use the following formulae:

- annual costs of storing information in paper-based form:

\[V_{st}^0 = \sum_{i=1}^{l} V_{pp} \cdot n_{it}^{pp} \cdot a_i,\]  

(34)
- annual costs of processing data in paper-based form:

\[ T_p^0 = \sum_{i=1}^{f} n^p_i \cdot t_{iq} \cdot \omega_i \cdot \frac{v^w_{iq}}{T^w_q} \]

(35)

- annual costs of data delivery in paper-based form:

\[ V^0_i = \sum_{i=1}^{f} n^p_i \cdot t_{iq} \cdot \omega_i \cdot \frac{v^w_{iq}}{T^w_q} \]

(36)

where \( V^p \) is the cost of 1 sheet of paper; \( n^p_i \) is the amount of sheets in the \( i \)-th document; \( \omega_i \) is the amount of \( i \)-th documents per annum; \( t_{iq} \) is the amount of handlings of the \( i \)-th document per annum; \( t_{iq} \) is the handling time (delivery) of the \( i \)-th document by user \( p_q \); \( v^w_{iq} \) is monthly salary of employee \( p_q \); \( T^w_q \) is the amount of working hours per month of the user \( p_q \); \( n^p_i \) is the number of delivery requests of the \( i \)-th document per annum.

PROBLEM STATEMENT

The problem of structural and parametric synthesis of the EDMS of research and education institution is formulated as follows:

it is necessary to find the vector of variables \( x^* = (S^*, HW^*, SW^*) \) from the set of permissible solutions \( X \), wherein the value of the objective function reaches its maximum:

\[ \{S^*, HW^*, SW^*\} = \arg\max_{S,HW,SW} (H), \]

(37)

to satisfy equations of constraints in the EDMS mathematical model (1-30) and constraints on:

- EDMS payoff:

\[ V_s \leq V_0 - V_d, \]

(38)

- the required document management productivity:

\[ E_s \geq \sum_{i=1}^{K} w_i \frac{t_{iq}}{t_{ui}}, \]

(39)

\[ 0 < \sum_{j=1}^{J} t_{ij} \leq t_{ui}, \]

(40)

- the required EDMS quality:

\[ Q_s \geq Q_s^*, \]

(41)
Formulation of the Problem of Structural and Parametric Synthesis of the EDMS:

\[ \varphi(SD^*, SR^*) \geq \frac{\sum_{r=1}^{n} \varphi(SD_r, SR_r)}{|SD| \cdot |SR|}, \quad (42) \]

where \( Q_s^* \) is basic values of information system quality evaluation determined by expert group or competitors’ software evaluation;

Thus, we formulated the problem of structural and parametric synthesis of the EDMS of research and education institution. The proposed model takes into account hardware and software characteristics of the system and allows for the system optimization by the economic efficiency criterion given a number of additional conditions.

**PRACTICAL APPLICATION**

The results of theoretical studies formed the basis for the practical implementation of the EDMS of Tambov State Technical University, the general scheme of which is shown in Fig. 2. The model was used in the design of information subsystems for the Department of patent and intellectual property protection, Department of basic and applied research and Department of training and certification of highly qualified personnel [1.19].

![INFORMATION SYSTEM OF DOCUMENT MANAGEMENT](image)

**Figure 2. The structure of EDMS of Tambov State Technical University**
Application of the proposed approach to the EDMS development exemplified by the information subsystem of Department of training and certification of highly qualified personnel improved the economic efficiency by 1.84 times compared to the cost of document management before the implementation of EDMS. For the system implementation we used Oracle APEX environment as a platform, the system was implemented as ‘Oracle Application’, optimal values of EDMS hardware specifications had the following values: the size of QST repository was 500 (GB), the server bandwidth CST was 3 (MB / sec), hardware costs VST amounted to 30000 (rubles), the annual economic cost s were 92605 (rubles).

The developed subsystem enabled to move from paper-based to electronic document management, automate the work with electronic individual plans postgraduates, reporting and statistics. Fig. 3 illustrates operation of information system.

![Figure 3](image)

**Figure 3:** Information subsystem for Department of training and certification of highly qualified personnel

**CONCLUSION**

Using the system analysis methods we formulated and solved the problem of structural and parametric synthesis of the EDMS of research and education institution; the problem is based on the theoretical and applied research, system links, movement patterns and document life cycle. The system allows increasing the economic efficiency of electronic document management by optimizing the structure, software and hardware features of the EDMS.
Based on the set-theoretic analysis of the subject area we developed a mathematical model of the EDMS of research and education institution; the model accounts for the description of software and hardware parameters of the information system, changes of states of document operations and conditions of their performance given the rules of access differentiation. We developed the EDMS optimization criterion, which accounts for the evaluation of economic efficiency of the information system, costs of paper work and system development costs, as well as additional constraints on productivity, quality of work and degree of user satisfaction. This approach made it possible to create the system, which satisfied the specified requirements, and evaluate the cost-effectiveness of the solution.

The above approaches found practical application in solving the problem of structural and parametric synthesis of the EDMS subsystem for the Department of training and certification of highly qualified personnel at Tambov State Technical University. The results confirmed the adequacy of the mathematical model and the effectiveness of the proposed method of the EDMS optimization. The results of the study can be used for creation, evaluation and optimization of document management systems of different scale and complexity.

This work was financially supported by the Ministry of Education and Science of the Russian Federation in the framework of the basic part of the state task.

REFERENCES


