

Modeling and Optimization of the Life Cycle of the Organization's Information Resources

G A Dorrer¹, A A Popov¹

¹Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarsky Rabochy Av., Krasnoyarsk, 660037, Russian Federation

E-mail: g_a_dorrer@mail.ru

Abstract. Electronic information resources (EIR) play an important role in the activities of modern organizations and enterprises. The paper presented discusses the problem of increasing the efficiency of EIR processing and storage. To solve this actual problem, based on the classification of information resources within the framework of the Dublin core of metadata and business processes, a probabilistic dynamic model of the EIR life cycle was created using the Markov chains formalism. The model allows to predict the process of storing and updating EIR, as well as analyze the work of information systems. The authors have developed a software system "The life cycle of electronic information resources". Using this system, it's possible to visually see the dynamics of the EIR life cycle and analyze the loading of disk equipment by several types of resources over time. The data obtained allow to choose the equipment that is optimal in price, volume and access speed. The system was tested when analyzing the dynamics of information resources of the Scientific Library of the Reshetnev Siberian State University of Science and Technology (Krasnoyarsk).

1. Introduction

In modern information systems there are a big variety of different types of information resources that need to be stored and processed, as well as provide convenient and timely access to the final user, which requires the creation of systems for their classification. Initial attempts to use for this purpose the traditional systems of bibliographic description of information resources showed their insufficiency. In order to increase the efficiency of using information resources (IR) of organizations, it is necessary to classify them, estimate the intensity of input flow, conduct a comparative analysis based on the data obtained, which will best combine the information resources into a single information system [7].

A major contribution to the development of information theory and information management belongs to C. Shannon, J. Pierce, R. Fisher, N. Wiener, L. Brillouin, A. Kolmogorov. In recent years, work has been done that solved the issues of effective use of IR in optimizing the activities of organizations. These are works by V.A. Old, A.I. Bashmakov, V.A. Staryh, T.A. Gavrilov, A.M. Vender, A.M. Karminsky, S.A. Zhukova, K.G. Skripkina, S.V. Cheremnykh, G.A. Titorenko et al. [2,3,5,9,11,13,14]. In the works of G.A. Dorrer, G.M. Rudakova, E.M. Gritsenko, A.A. Popov considered the life cycle of educational EIR [1,3,4,5,7].

In the work reviewed, urgent problems, connected with the increase of efficiency of search, transfer and analysis of information was discussed. But the problem of the IR structure optimizing allowed us to reduce the current costs associated with servicing information resources.

A review of well-known works showed that the EIR theory, their classification, metadata in existing information systems are not sufficiently developed. All EIRs change over time, the value of the information contained in a particular information resource may be different. Information resource is born, lives and dies, like everything else. There is a need to introduce a new concept - the life cycle of an information resource. Currently, this issue is not well understood, unlike, for example, the life cycle of software and information systems. In addition, all software engineering is based on the software life cycle model.

Accordingly, in order to make EIR more accessible, as well as to reduce the time and cost of their use, it is necessary to optimize it using. This can be achieved using the EIR classification based on the Dublin Core Metadata (DCMI), considering their life cycle [12].

The most important purpose of information systems is to work with information resources. Due to the large volumes and numerous types of operations performed with information resources, areas of activity for working with EIR were identified - the business processes of managing an EIR that were identified in the overall Process Classification Framework developed by the American Center of Productivity and Quality [10]. Based on the analysis of EIR-related problems, the Markov chains formalism was chosen for a convenient description of the EIR life cycle and model development. The Markov chains mathematical apparatus makes it possible to evaluate many characteristics of the information processes of the systems: probable completion time of certain work phases, average uptime, average performance, and others.

2. Methods

Based on the classification of EIR on basis of DCMI, business processes for managing IR, the model of the life cycle of an information resource may be represented as a Markov chain; the model is shown in Figure 1.

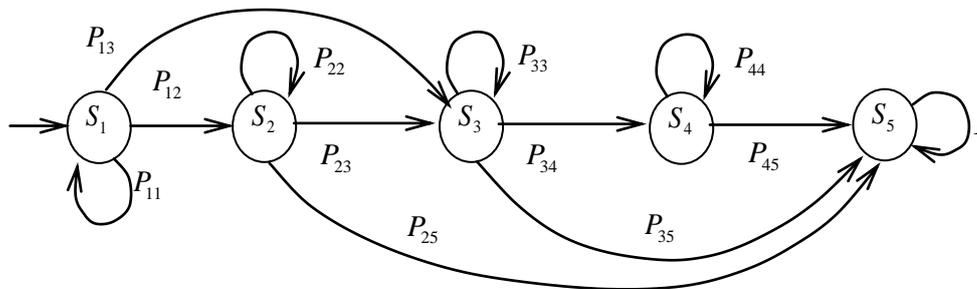


Figure 1. The model of the life cycle of an information resource.

Based on the J. Niessen’s [8] classification of business processes, the information resources pass five stages of the life cycle $E = \{E_1, E_2, E_3, E_4, E_5\}$, where:

- E_1 – the stage of creation (receipt) of IR,
- E_2 – storage phase of IR,
- E_3 – the stage of IR processing,
- E_4 – the stage of IR archiving,
- E_5 – the stage of destruction of IR;

During the life cycle, the relevance of the information contained in the IR changes. From this point of view, information can be classified as follows:

- $A = \{A_0, A_1, A_2, A_3\}$ is the set of levels of the IR importance, where:
- A_0 – is neutral,

A_1 – is of critical importance,
 A_2 – is important,
 A_3 – unimportant.

Model of the IR life cycle contains set of the considered states $S = \{S_1, \dots, S_5\}$, where:

$S_1 = \{E_1, A_0\}$ – creation of IR,

$S_2 = \{E_2, A_1\}$ – storage and processing of critical importance IR

$S_3 = \{E_2, A_2\}$ – storage and processing of information resources with important information,

$S_4 = \{E_4, A_3\}$ – archiving, processing and storage of information resources with unimportant information,

$S_5 = \{E_5, A_0\}$ – removal of IR.

Information resources are characterized by the set of parameters $R = \{T, RT, V, P, X, U\}$, where:

T – is the set of discrete points in time: $T = \{0 \leq t, \tau \leq T_{max}\}$.

$RT = \{RT_1, \dots, RT_{12}\}$ – the set of resources types according to the classification of the Dublin Core Metadata [15] listed in Table 1.

$V = \{V_i(t, \tau_i), i = 1, \dots, 12\}$ – set of volumes of IR of all types in dynamics, t - current time,

τ_i – time when the i -th resource arrives at the information system, $t \geq \tau_i \geq 0$.

$V_{i0}(\tau_i) = V_i(\tau_i, \tau_i)$ – volume of IR of the i -th type, which entered the system at the time τ_i .

$V_i(t, \tau_i) = [v_{ij}(t, \tau_i)]$ – vector of the i -th resources volume in its various states $j = 1, \dots, 5$.

Table 1. Type of IR according to DCMI classification.

RT_i	Type IR	Content
RT_1	Collection	Collection – a collection of blocks, parts, elements
RT_2	Dataset	Data Structure – lists, tables, databases
RT_3	Event	Time-dependent short-term event,
RT_4	Image	Visual Data
RT_5	Interactive Resource	Interactive resource, implying user reaction
RT_6	Service	Service – a system that implements functions that are useful for the end user
RT_7	Software	Computer program as source code, or compiled
RT_8	Sound	Sound – a resource designed for playback
RT_9	Text	Text – a sequence of words to read
RT_{10}	Physical Object	An inanimate object or substance in three dimensions
RT_{11}	Still Image	Static image – picture, picture, picture
RT_{12}	Moving Image	A series of images that evoke the impression of movement

In this case, the total amount of the resource saved:

$$\sum_{j=1}^5 v_{ij}(t, \tau_i) = V_{i0}(\tau_i), T_{i_{max}}; \tag{1}$$

$P = \{P_i = \|p_{kl}^i\|, i = 1, \dots, 12\}$ – set of matrixes of transition probabilities between the IR states of the i -th type for one time step; p_{kl}^i – the probability of a resource of i -th type from the state S_k to the state S_l , transition, $k, l = 1, \dots, 5$,

$$\sum_{k=1}^5 p_{kl}^i = 1, i = 1, \dots, 12; k = 1, \dots, 5. \tag{2}$$

$X = \{X_i(t, \tau_i) = [x_{ij}(t, \tau_i)], j = 1, \dots, 5\}$ - set of vectors that determine the probabilities of finding information resources in different states. The value $x_{ij}(t, \tau_i)$ represents the probability of the current time t being the resource of the i -th species in the state S_j , which entered the system in the moment τ_i .

$$\sum_{j=1}^5 x_{ij}(t, \tau_i) = 1, \forall i, t \geq \tau_i. \tag{3}$$

Subsequently, the volume of the incoming resource is distributed proportionally between the states, the probabilities of being in these states are determined. On the basis of the above and from the theory

of Markov chains, we obtain the following formulas for calculating the parameters of information resources.

The total amount of information resources of the i -th type that are in the system, is determined by the vector:

$$V_i^\Sigma(t) = \sum_{\tau_i=0}^{T_{\max}^+} V_i(t, \tau_i), \quad t=0, \dots, T_{\max}^+, \quad (4)$$

$T_{\max}^+ = T_{\max} + T_{Li}$ - lifetime of the resource i -th type.

$U = \{U_j(t), j=1, \dots, 5\}$ – set of the resources of one degree of relevance. The total amount of resources of the j -th degree of relevance is determined by the sum:

$$U_j(t) = \sum_{t \geq \tau} \sum_{i=1}^{12} v_{ij}(t, \tau), \quad j=1, \dots, 5. \quad (5)$$

The estimated amount of resources of the j -th degree of relevance, necessary for determining the parameters of the disk system $U_j^\Sigma(t)$, is equal to the maximum value of the value $U_j(t)$ for the entire calculated time interval: $U_j^\Sigma(t) = \max U_j(t), t \in [0, T_{\max}^+]$.

For the convenience of carrying out practical calculations using the formulas obtained, the program “Life cycle of information resources” has been developed.

3. Model of IR dynamics of Reshetnev Siberian State University of Science and Technology (SibSU) scientific library

In March - April 2017, research and consideration of the life cycle of information resources used in the scientific library of SibSU were conducted.

Within the framework of the Dublin core, the types of information resources of library were studied and classified, and the intensity of the flow of these resources to libraries was analyzed. An expert analysis of the life cycle of information resources in the library made it possible to estimate the matrix of transition probabilities p_{kl}^i for each type of IR.

Consider the dynamics of information resources used in libraries of SibSU. The calculation was carried out using the developed program system "Life cycle of IR". During the program operation, after filling in the matrices, files are automatically generated that define the scenario for the receipt of information on each type of EIR within the DCMI classification during the year. Each cycle of the simulation system corresponds to one day of the organization.

Conventionally, we assume that at the beginning of the modeling period the data warehouses were empty, $V_{i0}(0)=0, i=1, \dots, 5$. When performing calculations for each data stream, a graph was plotted that reflects the probability that the resource stays in each of the states on each clock cycle. Below is a graph for the information resource "Interactive Objects" (Figure 2).

Using the above graphs, the work of the information system is clearly visible, it is shown in what state the EIR is at a certain point in time, how much disk space is involved, so you can estimate the required capacity of the equipment necessary for storing and processing the EIR. For example, it is clear that the EIR "Interactive objects" does not involve archival storage (state S_4), and practically does not use rare access (state S_3).

The EIR of this type is in the state of storage and processing of critical importance, which means it mainly uses equipment with a high access speed (state S_2), after which it is immediately removed.

The following databases are replenished annually: “Novelties”, “Articles”, “Dissertations”, “Works”, “Electronic Edition”, “Readers”, “Completion”, “Zheleznogorsky Branch”, “State standards”, “Events” and others. The type of EIR in the framework of DCMI is a collection, these are journals: “Bulletin”, “Reshetnev’s readings”, “Forest inventory and forest management”, “Coniferous boreal zone”.

Therefore, in order to increase processing efficiency and ensure timely access for readers, it is necessary to use modern data storage systems that allow you to reliably process and store large amounts of information using modern automatic backup methods.

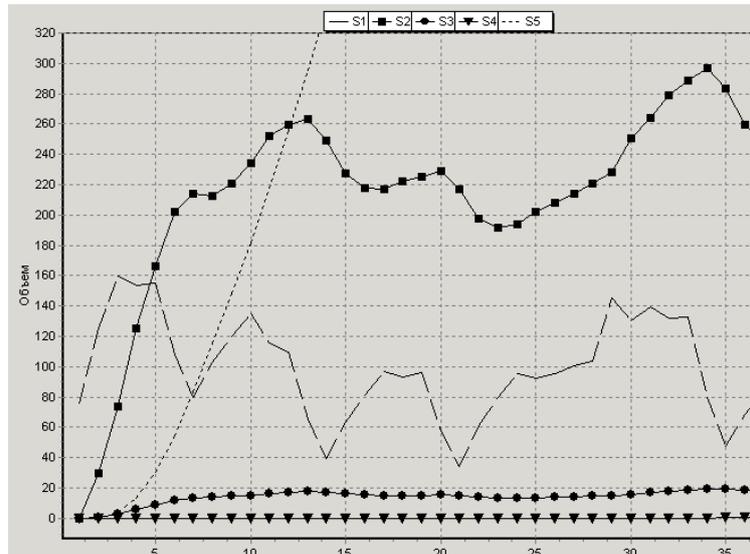


Figure 2. Dynamics of IR "Interactive objects".

After performing the calculation, all the original and received data can be exported to MS Excel for further processing and analysis. The program allows you to process from 1 to 5 types of resources to a depth of 1,000,000 cycles. The program is implemented input control and provides the ability to save the original data.

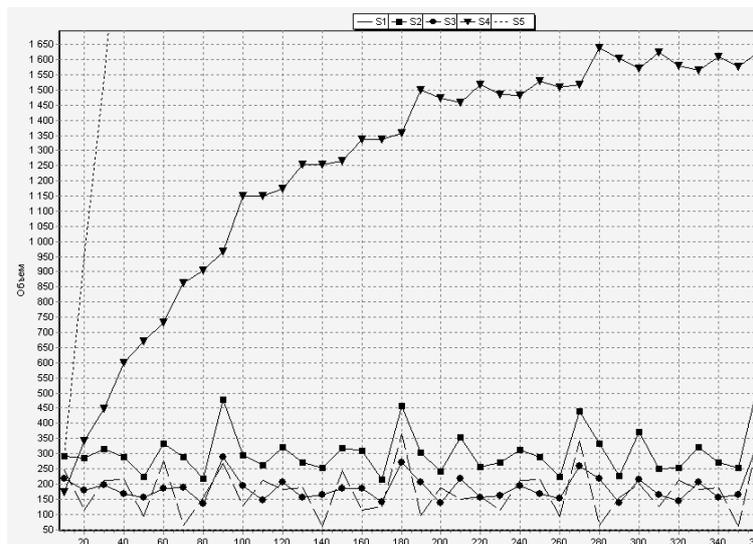


Figure 3. Dynamics of the total volume of all EIR categories.

In Figure 3 you can clearly observe the dynamics of the life cycle of an information resource. In this network, the flow and life cycle of the IR is carried out by transferring chips to the direction that is necessary and conditionally directed, which makes it possible to obtain a qualitatively adequate picture of the dynamics of IR for both a single resource and the aggregate.

Mathematical methods of Markov chains allow to estimate many characteristics of information processes of systems, such as the probable completion time of certain stages of work, the average uptime, average productivity, and others.

4. Conclusions

Using the system "Life cycle of IR", which implements the proposed methodology, it is possible to analyze the loading of disk equipment with several types of resources for any period of time and, based on the data obtained, select the equipment that is optimal in price, volume and access speed.

A certain difficulty in using the proposed methodology is to obtain estimates of the transition probabilities between states. For this purpose, in addition to the considered expert estimation method, you can use the methodology for estimating the parameters of Markov chains using experimental data contained in the book [8].

In addition, you can solve the problem of the optimal strategy for power management, considering the timing of replacement and equipment characteristics.

In the future, continue work planned on the study and improvement the life cycle model of information resources and methods for estimating model parameters. The developed algorithm is planned to use the example of cloud data storage.

References

- [1] Gritsenko E M 2005 Management of the life cycle processes of educational information resources *Abstract of the dissertation of the candidate* 22 (Krasnoyarsk: SibGTU)
- [2] Dorrer G A and Rudakova G M 2001 Probabilistic model of the process of interactive learning *Open education journal* 2 57-65
- [3] Dorrer G A 2009 Theory of information processes and systems *A manual for students of the specialty 230201.65 "Information systems and technologies"* 202 (Krasnoyarsk: SibGTU)
- [4] Dorrer G A and Rudakova G M 2006 Technology of modeling and development of educational electronic publications *Study guide for students* 272 (Novosibirsk)
- [5] Zhukova S A 2008 Modeling and optimization of the structure of the information resource *Abstract of the dissertation of the candidate* 18 (Izhevsk: IzhSTU)
- [6] 2017 Official site of Reshetnev Siberian State University of Science and Technology [Online] Available: <https://www.sibsau.ru/>
- [7] Popov A A 2011 Modeling the life cycle of information resources and information exchange processes *Abstract of the dissertation of the candidate* p 23 (Krasnoyarsk: SibGTU)
- [8] Niessen J 2008 Introduction to IT Service Management <http://allitil.ru/book/42-vvedenie-v-itsm.-glava-1..html/>
- [9] Riordan D 1966 *Probabilistic service systems* (Moscow) 184
- [10] 2005 *American Productivity & Quality Center* <http://www.apqc.org/>
- [11] Bekey G A and Karplus W J 1968 *Hybrid Computation* (New York) 197
- [12] 2019 Dublin Core Metadata Initiative *DCMI is supported by its members and is a project of ASIS&T* <http://dublincore.org>
- [13] Fishman G S and Kiviat P J 1967 The Analysis of Simulation-Generated Time Series *Management Science* 13(7) 67-77
- [14] Forrester J W 1961 *Industrial Dynamics* Massachusetts Institute of Technology Press Cambridge Mass 231
- [15] IEEE 1484.12.1-2002 *Learning Object Metadata standart* (New York: IEEE)