HMMOCS 2022
International Workshop "Hybrid methods of modeling and optimization in complex systems"

ALGORITHM DEVELOPMENT FOR IMPLEMENTING THE INFORMATION RESOURCE LIFE CYCLE DIGITAL TWIN

A. M. Popov (a), A. A. Popov (b)*, N. A. Romanov (c), A. K. Ovsyankin (d),
A. Batyrbekova (e)
*Corresponding author

(a) Reshetnev Siberian State University of Science and Technology, 31, prospect Krasnoyarsky rabochy, Krasnoyarsk, Russia, vm_popov@mail.sibsau.ru
(b) Reshetnev Siberian State University of Science and Technology, 31, prospect Krasnoyarsky rabochy, Krasnoyarsk, Russia, tolynbms@yandex.ru
(c) Reshetnev Siberian State University of Science and Technology, 31, prospect Krasnoyarsky rabochy, Krasnoyarsk, Russia
(d) Reshetnev Siberian State University of Science and Technology, 31, prospect Krasnoyarsky rabochy, Krasnoyarsk, Russia, dikoti1994@yandex.ru
(e) Reshetnev Siberian State University of Science and Technology, 31, prospect Krasnoyarsky rabochy, Krasnoyarsk, Russia

Abstract

Currently, in many organizations that are engaged in information processing, there is a problem of operational processing of information resources. The daily volume of information processed in the structure of organizations increases every day. Therefore, the problem of processing and analysing an information resource in large volumes in a short time with an approximation to a real information system is very relevant. The problem of increasing the efficiency of information resources processing processes in IT departments is solved by creating a digital twin algorithm using simulation modeling. In this paper, the methods of simulation modeling, digital twin and the life cycle of an information resource are investigated. A digital twin of the information resource life cycle was designed, based on which a model was built in the AnyLogic simulation environment. The developed model will allow the selection of data ordering for the provision of processed data for other departments and the prediction of hardware capacities.

Keywords: Digital twin, simulation modeling, anylogic, information resource, life cycle

1. Introduction

Today, at many enterprises that are engaged in information processing, there is a problem of processing information resources. Classification, conducting analytical research, providing the necessary software and hardware solutions for processing information resources is an important task for the quality work of both individual departments and the organization as a whole.

According to Forrester (1961) every manager is faced with the need to expand information networks and increase data warehouses for electronic information resources. The management of an information resource in the modern world is becoming increasingly difficult, which required legislative support for this area of human activity. In particular, a number of laws regulating this activity have been adopted. The object of information processes is an information resource. Information resources of a country, region, and organization are considered as strategic resources, similar in importance to reserves of raw materials, energy, etc.

Information resources Bekey and Karplus (1968) notes are concentrated and used by IT organizations, which are understood as either organizations specializing in information technology services, or IT departments of organizations working in other fields.

In the Federal Law of the Russian Federation of 2006 "On Information, Information Technologies and Information Protection", information resources are defined as separate documents and separate arrays of documents, documents and arrays of documents in information systems intended and independently designed for distribution to an unlimited number of persons, or serving as the basis for the provision of information services. However, to date, no methods have been developed for creating a digital double of an information resource, taking into account their classification.

2. Problem Statement

Zhukova (2008) claims that currently, the developed life cycle models of information resource are based on the Markov chain formalism. Mathematical methods of Markov chains allow us to evaluate many characteristics of information processes of systems, such as the probable completion time of certain stages of work, average uptime, average productivity, and others, but they are not very flexible and visual tools for processing large amounts of data.

Therefore, the problem of processing and analyzing information resources in large volumes in a short time with an approximation to a real information system is very relevant. Dorrer (2001) consider this problem can be solved by creating an algorithm for the digital twin of information resources, which will help to simulate what will happen to real data in information systems.

Gritsenko (2005) notes this work is devoted to solving a number of problems related to the implementation of the digital twin algorithm of information resources, taking into account the developed methods and models of the life cycle of information resources.

3. Research Questions

The following questions were raised during the study:

▪ What is an information resource and what is its role?

▪ What is the role of simulation in business processes?

▪ What is the experience of using a digital double?

4. Purpose of the Study

The purpose of this work is to increase the efficiency of information resource processing processes through the development of a digital twin. Based on the classifications of information resources of the Dublin Metadata Core (2022) and the life cycle model of information resources, an algorithm has been developed and implemented as a simulation model of the digital twin of information resources.

The developed model of the digital twin of information resources will be used to select organize data to provide processed data for other departments and predict hardware capacities.

5. Research Methods

The authors used simulation-modeling methods in the Anylogic development environment.

5.1. The Concept of an information resource

An information resource (IR) is a collection of data in information systems. For example, a collection of books in a library, customer data in a service center.

IR is used in many different companies engaged in different activities. Someone uses an IR to store a list of products, and someone to store a list of employees.

Being objects of knowledge preservation, IRS serve as data for the construction of new knowledge, which, in turn, can be presented in the form of IR. The use of AI presupposes the possibility of their interpretation either by a person or by computer programs. Information resources exist in electronic form (for example, electronic document files) and other forms.

Each IR is accompanied by a description (specification) containing the data necessary to find it, determine the scope of application.

Since the storage and use of an information resource is expensive, there was a problem of simplifying it. This task has already been achieved by classifying the information resource using the Dublin Metadata Core (DCMI). Due to this, the time and money spent on using the information resource have been reduced.

Dorrer (2009) notes the Dublin Core is a set of metadata elements, the meaning of which is described verbally and fixed in the specifications of the standards defining it. The combination of the values of these elements can be used as a structured description of the content of various kinds of text documents or documents presented in other environments, as well as user requests.

Haytov (2021) consider the product life cycle (product life cycle) is the combination of several processes and phenomena that are repeated at intervals determined by the time of the existence of a

particular product, the cycle begins at the moment of conception before its disposal. Such a cycle is a frequent case that is applied to industrial products.

The information resource during its existence goes through the following stages of the life cycle:

- collecting information;
- keeping;
- processing;
- archiving;
- destruction.

5.2. Mathematical model of the information resource life cycle

During the life cycle, the relevance of the information contained in the information resource changes. Popov (2011) consider from this point of view, information can be classified as critical, important and unimportant.

Based on the above, the life cycle model of an information resource can be simplified in the form of a Markov chain.

Figure 1 shows a life cycle model of an information resource in the form of a Markov chain.
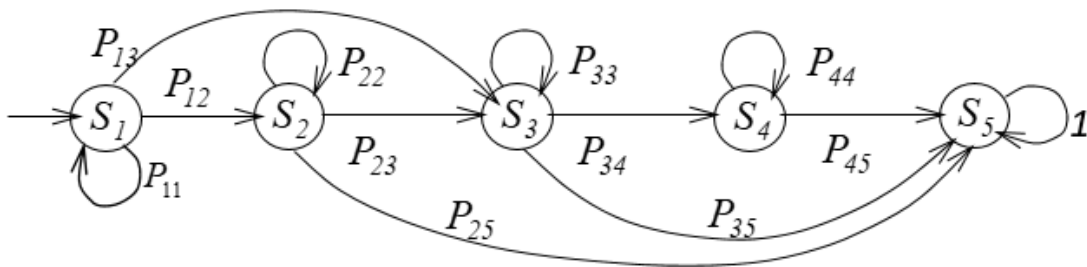

Figure 1. The life cycle model of an information resource

There are 5 states in this circuit, which mean:

- S_1 - creating an IR;
- S_2 - storage and processing of critical importance IR;
- S_3 - storage and processing of IR with important information;
- S_4 - archiving, storage and processing of IR with unimportant information;
- S_5 - removing an IR.

The transition from one state S_i to another S_j in one time step is a random event, the probability of which p_{ij} is determined by the stages of the life cycle of the IR, their "lifetime" and other factors.

The probabilities of the transition of the i −th type of resource also determine the average duration of the information resource's stay in each of the selected states.

The R_{ij} resource is characterized by the following parameters:

- the time of receipt of the IR in the information system τ, $0 \leq \tau \leq T$, where T – is the time of the study of the system;

- the duration of the "life" of the IR T_{Si} - the number of moments of time from the receipt of the resource in the system to the moment of its removal with a probability close to one;

- the volume of the received resource of the i - th type at the time of $\tau - V_{i0}(\tau)$, MB;

- $v_{ij}(t, \tau)$ $-$ the volume of the resource R_{ij}, MB, at the time $t \geq \tau$. The values $v_{ij}(t, \tau)$ form the vector $V_i(t, \tau) = [v_{i1}(t, \tau), \ldots, v_{i5}(t, \tau)]$, while $\sum_{j=1}^{5} v_{ij}(t, \tau) = V_{i0}(\tau)$, $\forall t \geq \tau$.

- $x_{ij}(t - \tau)$ $-$ the probability of finding the resource R_{ij} in the j $-$degree of relevance at the time $t \geq \tau$. Probabilities $x_{ij}(t - \tau)$ form a vector $X_i(t, \tau) = [x_{i1}(t, \tau), \ldots, x_{i5}(t, \tau)]$, while $\sum_{j-1}^{5} x_{ij}(t, \tau) = 1$, the initial probability distribution $X_i(\tau, \tau) = [1,0,0,0,0]$.

The resource that has entered the system (in the state S_1), is further redistributed between states in proportion to the probabilities of the system being in this state. Based on the above, we will get the following calculation formulas.

The dynamics of the change in the state of the IR is determined by the equations:

$$X_i(t + 1, \tau) = \begin{cases} 0, t \leq \tau, \\ X_i(t, \tau) P_i, t \geq \tau. \end{cases}$$

$$v_{ij}(t, \tau) = \begin{cases} 0, t \leq \tau, \\ V_{i0}(\tau) x_{ij}(t, \tau), t \geq \tau. \end{cases}$$

$$i = 1, \ldots, n, j = 1, \ldots, 5.$$

The total amount of resources of the i $-$th type located in the system will be determined by the vector

$$V_i^{\Sigma}(t) = \sum_{\tau=0}^{T} V_i(t, \tau), t = 0, \ldots, T_1,$$

where $T_1 = T + T_{Si}$; T_{Si} $-$ is the lifetime of the resource of the i $-$th type.

The total amount of resources of the j $-$th degree of relevance will be determined by the sum

$$u_j^{\Sigma}(t) = \sum_{\tau \leq t} \sum_{i=1}^{n} v_{ij}(t, \tau),$$

$$j = 2, \ldots, 4, 0 \leq t \leq T$$

The calculated volume of the IR of the j $-$th degree of relevance U_j^{Σ} for determining the necessary parameters of the storage device is equal to the maximum value of the value $u_j^{\Sigma}(t)$ at all the time interval under study:

$$U_j^{\Sigma} = max \, u_j^{\Sigma}(t), j = 2, \ldots, 4, 0 \leq t \leq T.$$

Taking this model as a basis, it is possible to build a digital twin of the life cycle of an information resource.

A digital twin is a digital model of any objects, systems, processes or people. It accurately reproduces the form and actions of the original and is synchronized with it.

A digital double is needed to simulate what will happen to the original in certain conditions. This helps, firstly, to save time and money, and secondly, to avoid harm to people and the environment.

The digital double helps to get answers to the "what if" questions, understand the behavior of the system and verify the results of the model. At the same time, different indicators of the real system can be used to verify the model.

Taking the Markov chain model as a basis, the next thing to do is to transfer the model to the AnyLogic simulation environment using the "Process Modeling Library" located in it. First of all, you need to create an imitation of each type of resource according to the Dublin core; a block called Source is suitable for this purpose. This block will generate agents, which, in this case, will represent the type of resources specified by the block. A total of nine resource types will be taken, which are shown in Table 1.

Table 1. Elements of the basic dictionary of DCMI resource types.

Dictionary element	Meaning
Collection	A collection is a collection of blocks, parts, and elements. The term collection means that the resource is described as a group.
Dataset	Data structure. Denotes information presented in strict formats (lists, tables, databases).
Event	Event. A time-dependent event does not have a long duration.
Image	Image. Such a resource is visual data in a symbolic format other than text.
Interactive Resource	An interactive resource is a resource that implies a reaction from the user that must be understood and processed. For example, these are forms in web pages, applets, multimedia learning objects, chats, virtual reality.
Service	A service is a system that implements one or more functions that are useful to the end user.
Software	A computer program, in the form of source code, which is available for installation on other computers. For programs that provide interactive functions only, it is recommended to use Interactive Resource
Sound	Sound is a resource whose content is designed for audio playback.
Text	A resource whose content is primarily a sequence of words to read.

Since different types of resources can come to the IR, with different intensity and at different times, there may be a situation that a certain type of resource may not come to the IR at all, for this, the model will need a side that will temporarily hold the generated agents and a block that will not allow them to exit until this not required. The Queue and Hold blocks will be used for these tasks. The first block will accumulate an unlimited number of agents, while the second block will keep agents in the first block, preventing them from exiting until the Hold block is unlocked, in which case the agents will start exiting the block Queue.

6. Findings

To simulate the process of storing a resource, the Delay, Hold and Select Output blocks will be used. The Select Output block differs from the Select Output 5 block in that it has only two outputs instead of 5 and the average chance of an agent exiting one or the other output is set. The Delay block delays agents in itself for a user-defined time or by calling a function. Like the Queue block, this block has a capacity.

After passing one of the processes, the resource is removed from the IR, respectively, since the generated agents act as resources, they must also be removed from the model, for this, the Sink block will be used. The block destroys the agents, after which they leave the model's memory.

The model should also have a dashboard, which will show information about the generated agents of each type of resources, the number of resources on the server and the amount of processed data in megabytes. Also have text fields with which the user can dynamically set the server volume and the time during which the resource will be at the processing stage.

When the model is designed and the necessary blocks are defined, then it is necessary to develop the model in the AnyLogic simulation environment.

The result of the work done indicates that the implemented simulation model solves the problem of increasing the efficiency of information resource processing processes

When starting the model, the user will see an experiment screen in which he will be able to set the number of agents of each resource type generated at a time.

Figure 2 shows the initial screen of the experiment.

Figure 2. The initial screen of the experiment

In the text fields, opposite the names of resource types, the receipt of agents of the corresponding resource is configured.

When starting the model, the user can independently set the number of agents generated at a time for each type of resource separately, if the value is 0, then this resource will not be generated. After starting the model, the user will be able to change these values only by restarting the model.

The life cycle process of an information resource was modeled on the provided model. This model shows the Source blocks, which represent all types of resources for the Dublin core. Now that the model has been designed and built, it is time to launch the model and monitor its operation. Figure 3 shows a simulation of the model.
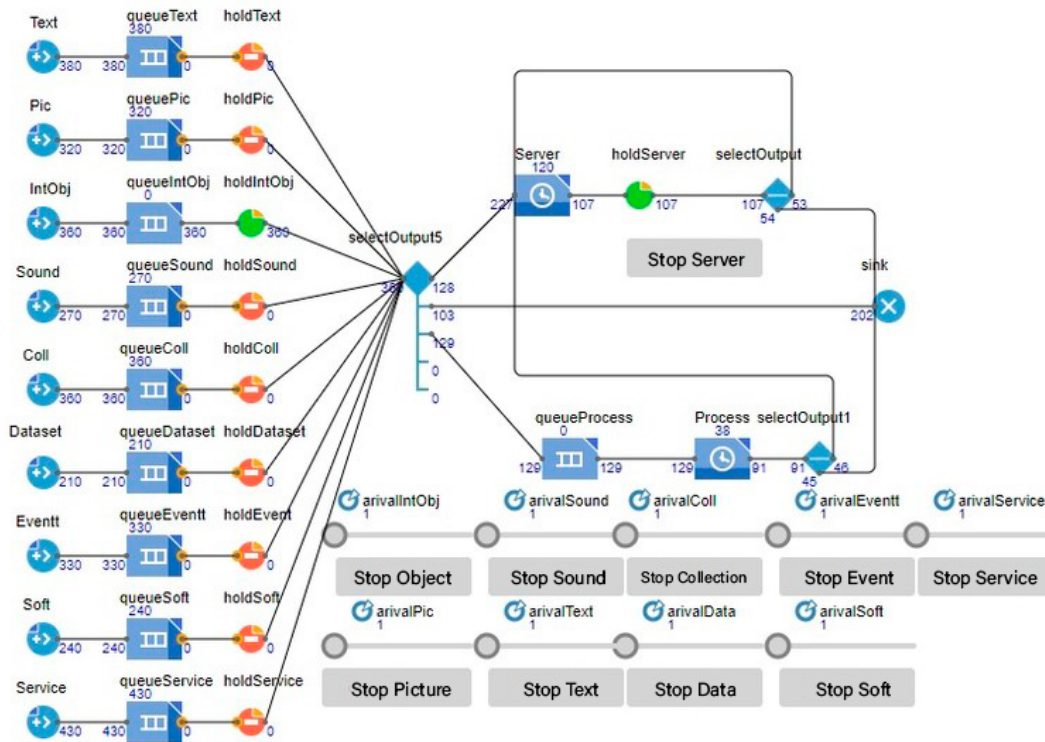
Figure 3. The initial screen of the experiment

In this experiment, the model worked for 40 seconds of real time or 35 hours of model time. In this experiment, only the agents of the interactive object resource entered the process and the holdServer block was unblocked, which allowed the agents to exit the server block. During the operation of the model, 360 agents of the IntObj block passed, 128 of which were sent to the server for storage, 103 agents were immediately deleted and 129 agents were sent for processing.

In the data storage subprocess, 107 agents left the server, 53 of which were returned back to the server and 54 agents were deleted.

In the queue processing subprocess, at the time the model was suspended, no agents were queued for processing, 38 resources were in active processing and 91 agents left the processing block, 46 of which were sent to the server for further storage, and the remaining 45 were deleted.

In Figure 4, you can clearly see how much of each resource the model generated during its operation; you need to take into account that the model only processed the agents of the interactive object resource. When the server is suspended, you can see that there are 120 agents on the server and 4068 megabytes of data were processed during operation. The resource processing time in the Process block was 5 hours, and the server capacity was 1000 agents.

Number of incoming IR		Number of resources on the server
Collection	360	120
Dataset	210	
Event	330	
Picture	320	Number of processed data, MB
Interactive object	360	
Service	430	4,068.774
Programm	240	
Sound	270	
Text	380	

Server capacity	1000	Resource processing time, hour	5

Figure 4. Dashboard model operation screen

Based on this work, it can be concluded that with these capacities, the server will not be overloaded for 35 hours of real time.

7. Conclusion

In the course of the work, a digital twin of the information resource life cycle was modeled and built, aimed at selecting the necessary hardware capacities for processing, storing and exchanging information. The development was carried out using the AnyLogic simulation tool using the process-modeling library.

Determining the necessary hardware capacity to maintain the operation of the AI is one of the important issues for IT companies; companies lose a large amount of money for permanent equipment replacements and improvements due to unforeseen circumstances of lack of storage space.

This model will allow IT companies to visually see the work of their AI and, based on the data obtained from the experiment, select the necessary hardware capacities. In the future, it is possible to test this model and put it into operation. In addition, to avoid unnecessary costs and clearly see the work of the information resource lifecycle, which will allow you to accurately assess the necessary hardware capacities that the company needs.

References

Bekey, G. A., & Karplus, W. J. (1968). *Hybrid Computation.* Wiley, Inc.

Dorrer, G. A. (2001). Probabilistic model of the interactive learning process. *Open education, 2*.

Dorrer, G. A. (2009). *Theory of Information processes and systems*. SibSTU.

Dublin Core. (2022). Metadata Initiative: DCMI is supported by its members and is a project of ASIS&T: DCMI: http://dublincore.org

Forrester, J. W. (1961). Industrial Dynamics. Massachusetts Institute of Technology Press. http://www.laprospective.fr/dyn/francais/memoire/autres_textes_de_la_prospective/autres_ouvrag es_numerises/industrial-dynamics-forrester-1961.pdf

Gritsenko, E. M. (2005). *Life cycle management of educational information resources*. SibGTU.

Haytov, X. O. (2021). Product life cycle and industry life cycle. *Economy and Society, 4*.

Popov, A. A. (2011). *Modeling of the life cycle of information resources and information exchange processes*. SibSTU.

Zhukova, S. A. (2008). *Modeling and optimization of the structure of an information resource*. IzhGTU.