Analysis of the life cycle of a library information resource using Process Mining technology

M G Dorrer, A A Popov and A N Bartuzanova
Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarsky Rabochiy Ave., Krasnoyarsk, 660037, Russia
E-mail: mdorrer@mail.ru

Abstract. The purpose of this work is to evaluate the applicability of the Process Mining methodology for the analysis of life cycle models of information resources. ProM software was used with the Inductive visual miner plugin. The source data was a database of the Seattle Public library hosted on the site Kaggle.com. Even from a fragment of the transaction history of library information resources, it was possible to recover the lifecycle model. The result confirms that it is possible to build and analyze models of the life cycle of information resources based on the event history logs.

1. Introduction
In the modern world, the flow of information and its exchange is increasing by the minute. Based on the existing data, new information resources are being used more often and have their distribution in more than one area. Thus, over time, existing long-term and rarely used information loses its relevance and use, which leads this data to the next stage of its life cycle.

Information resources are a collection of data organized to obtain reliable information in a variety of fields of knowledge and practice.

This topic has attracted the attention of researchers since the late 80's of the 20th century. Thus, in this work [1], the authors explore formalisms for designing and implementing information resource systems.

In this work [2], the authors suggest using methods of mathematical formalization of structuring information resources, describe the processes at the stages of structuring information, and consider the use of fuzzy logic at the stage of modeling information resources.

In work [3], the author suggests to use differential models of information processes of governing bodies that take into account the spatial distribution and transformation of data.

Based on the classification of business processes, an information resource during its existence goes through the following stages of the life cycle (LC):

• collection of information, the creation of an information resource, storage,
• processing (streamlining, searching, changing and updating),
• archiving,
• destruction.

In work [4], the authors propose a technique focused on the study of business artifacts and key conceptual objects, and the analysis of their behavior in the framework of life cycle models.
The Process Mining methodology, an approach to analyzing business processes based on modern tools in the field of collecting and processing data obtained directly from information systems, has shown its effectiveness as a tool for studying systems described by flows of events. The approach is proposed and developed by a group of researchers from Eindhoven University, led by Wil van der Aaalst. The main points of the approach are described, for example, in works [5, 6].

At the same time, the analysis of publications did not reveal the application of the Process Mining approach to the analysis of the life cycles of organizational and information objects. The authors previously attempted such an analysis in work [7], however, it was necessary to verify the operability of the approach on a verified database of a larger volume.

To solve this problem, an open database of the Seattle Public Library presented on the Kaggle.com Internet portal [8] was made an object of study. The Seattle Public Library provided an open access dataset that includes a log of all physical item checks from the library. Data for 2005 contain 1048576 records.

2. Methods

Dataset description

The data set presented in [8] contains the following columns:
- BibNumber – Bibliographic Number Identifier (integer type)
- ItemBarcode – Book barcode (integer type)
- ItemType – Book type (type string)
- Collection – Which collection the book belongs to (type string)
- CallNumber – Unique value of the book (type string, unique)
- CheckoutDateTime – Date and time of registration (type date)

Consider each stage of the life cycle in detail on the example of library information resources:

1. Creation, collection and storage

After a new book arrives in the library's warehouse (as in stock), the first stage of its life cycle begins. A library information resource now has a processing property that allows you to track detailed information about its changes and movements.

2. Processing (streamlining, searching, changing and updating)

At this stage, we can see at what time and in what quantity (for a given period of time) the library information resource was used. Thus, information appears that can be used to track all resource movements and draw conclusions about its relevance at the current time.

3. Archiving

After the library information resource loses its relevance and distribution in use, it is in an untouched state, which says that rarely anyone else needs the information stored in it. Thus, a simple resource occurs, and ultimately it goes to the destruction stage.

4. Destruction

At this stage, the cycle of the information resource completes. It is no longer used or used somewhere else, the resource has completely lost its distribution and needs to be finalized.

The Pro Tools framework was used to work with Process Mining algorithms [9].

In the course of the work, the task was set to study in detail the sequence of change of states of information resources in such a way as to reveal the relationship between groups of states and their participation in the stages of the life cycle.

3. Results

To consider a phased change in the state of a resource throughout the year, we need to identify a group of the most frequent events. Figure 1 shows the frequency histogram of the occurrence of events in the life cycle of an information resource.
The figure above shows that in 2005 there were 54 types of events in the sample. Among them may contain both frequently used types of events that occur many times during the year, and types of events that occur once. Using the event log, we identify rare types of events – Figure 2 and Figure 3.

![Figure 1. Event statistics.](image)

![Figure 2. The most rare events (part 1).](image)
Further, we will not consider these types of events in the analysis, since their influence will not affect the final result, that is, we can already conclude that 20 types presented in Table 1 can be excluded during the procedure for identifying patterns in the process of constructing a life model cycle of a library information resource.

**Table 1.** Event types excluded from life cycle analysis.

| No. | Name   | Event Type Occurrence in the Event Log |
|-----|--------|---------------------------------------|
| 1   | acart  | 11                                    |
| 2   | acdisk | 3                                     |
| 3   | acpam  | 2                                     |
| 4   | acper  | 4                                     |
| 5   | acpost | 2                                     |
| 6   | acside | 1                                     |
| 7   | arcd   | 16                                    |
| 8   | acrdrom| 2                                     |
| 9   | acdvc  | 2                                     |
| 10  | acmfc  | 3                                     |
| 11  | armfm  | 1                                     |
| 12  | armus  | 3                                     |
| 13  | arper  | 14                                    |
| 14  | arunkn | 2                                     |
| 15  | arvhs  | 1                                     |
| 16  | dcilll | 10                                    |
| 17  | jcrec  | 1                                     |
| 18  | jrkit  | 1                                     |
| 19  | jrvhs  | 2                                     |
| 20  | ucflpdr| 1                                     |
To illustrate the confirmation of our choice to exclude certain types of events, we apply the scatter diagram in Prom. The result is shown in Figure 4.

![Scatter plot of event types in the event log.](image)

**Figure 4.** Scatter plot of event types in the event log.

Before proceeding with the construction of a life cycle model based on cases of working with library resources in the considered time period, we will isolate and handle such cases using the ProM software tool.

Despite the fact that some library resources can have only one state transition (event) in the year under review, due to a large number of cases, we were able to identify and build the most likely routes for changing the state of library information resources. As a result of processing, the events and conditions shown in Figure 5 were included in the most likely route of the life cycle of an information resource.
Filtering the types of events of an information resource (book) life cycle during the analysis of the scatter diagram (see Figure 4) and constructing a graphic model of the life cycle (see Figure 5a, 5b) using the ProM software made it possible to compile a list of types that are most actively involved in the life cycle: acbk, acdvd, acvhs, jcvhs, jcdvd. These types of states belong to the main actions in relation to the book - this is mainly its archiving or saving on various devices. Thus, the life cycle model constructed on the basis of statistics analysis includes both events of a simple movement of a book from one reader to another, as well as a large number of events related to the collection and storage of information necessary for the reader from the book. These statistics allow assessing the demand for an information resource.
by users both in statics, at the current moment, and in dynamics, which will allow to predict the need for this resource in future time periods.

5. Conclusion
As a result of the analysis, the sets of state types that were most often encountered during the entire life cycle of the library resource were selected. A graphic model of the life cycle of a library information resource was constructed.

It is planned to continue research and analysis of the presented data set, expanding the set of considered parameters of events column "collection".

In addition, an interesting development of the work will be the identification of the correspondence of the life cycle of each of the information resources to the built generalized life cycle model.

Analysis of the data obtained during the Process Mining procedure for the library event log allows probabilistic forecasting for all library information resources.

In addition, this work can be generalized from the study of library information resources to the study of life cycles of a wider range of information resources, for which it will be possible to obtain event logs.

References
[1] Epstein R G and Aiken R M 1989 The Information Resource Model Computer Assisted Learning 2nd International Conference, ICCAL 89, May 9-11, 1989 (Dallas, Texas, USA) pp 77-101
[2] Shebanin V, Atamanyuk I, Kondratenko Y and Volosyuk Y 2016 Application of fuzzy predicates and quantifiers by matrix presentation in informational resources modeling XII International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH) (Lviv, Ukraine) pp 146-149
[3] Ivanov A K 2019 Modeling spatial transformation of information resources of management bodies Automated control systems 4 10-21
[4] Hompes B and Van der Aalst W 2018 Lifecycle-Based Process Performance Analysis Confederated International Conferences: CoopIS, C&TC and ODBASE, Proceedings, Part I (Valletta, Malta) pp 336-353
[5] Van der Aalst W 2011 Process Mining Manifesto International Conference on Business Process Management (Berlin) pp 169-194
[6] Van der Aalst W 2011 Process Mining: Discovery, Conformance and Enhancement of Business Processes (Berlin: Springer-Verlag) p 370
[7] Dorrer M G, Popov A A and Bartuzanova A N 2019 Analysis of information resource life cycle Journal of Physics: Conference Series 1399 033083
[8] Seattle Library Checkout Records 2015 Available at: https://www.kaggle.com/seattle-public-library/seattle-library-checkout-records.
[9] ProM 6 tutorial 2010 Available at: http://www.promtools.org/prom6/downloads/prom-6.0-tutorial.pdf