

- [11] Kalynychenko, O., Chalyi, S., Bodyanskiy, Y., Golian, V., Golian, N. (2013). Implementation of search mechanism for implicit dependences in process mining. 2013 IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS). doi: <https://doi.org/10.1109/idaacs.2013.6662657>
- [12] Sergii, C., Ihor, L., Aleksandr, P., Ievgen, B. (2018). Causality-based model checking in business process management tasks. 2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT). doi: <https://doi.org/10.1109/dessert.2018.8409176>
- [13] Levykin, V., Chala, O. (2018). Method of automated construction and expansion of the knowledge base of the business process management system. EUREKA: Physics and Engineering, 4, 29–35. doi: <http://dx.doi.org/10.21303/2461-4262.2018.00676>
- [14] Levykin, V., Chala, O. (2018). Method of determining weights of temporal rules in markov logic network for building knowledge base in information control systems. EUREKA: Physics and Engineering, 5, 3–10. doi: <http://dx.doi.org/10.21303/2461-4262.2018.00713>
- [15] Levykin, V., Chala, O. (2018). Development of a method for the probabilistic inference of sequences of a business process activities to support the business process management. Eastern-European Journal of Enterprise Technologies, 5 (3 (95)), 16–24. doi: <https://doi.org/10.15587/1729-4061.2018.142664>
- [16] Gupta, M., Gao, J., Aggarwal, C. C., Han, J. (2014). Outlier Detection for Temporal Data: A Survey. IEEE Transactions on Knowledge and Data Engineering, 26 (9), 2250–2267. doi: <https://doi.org/10.1109/tkde.2013.184>

## METHOD OF CONSTRUCTING AN ATTRIBUTE DESCRIPTION OF THE BUSINESS PROCESS "AS IS" IN THE PROCESS APPROACH TO ENTERPRISE MANAGEMENT

**Serhii Chalyi**

*Department of Information Control Systems  
Kharkiv National University of Radio Electronics  
16 Nauka ave., Kharkiv, Ukraine, 61166  
[serhii.chalyi@nure.ua](mailto:serhii.chalyi@nure.ua)*

**Ievgen Bogatov**

*Department of Information Control Systems  
Kharkiv National University of Radio Electronics  
16 Nauka ave., Kharkiv, Ukraine, 61166  
[bogatovevgeniy@gmail.com](mailto:bogatovevgeniy@gmail.com)*

---

### Abstract

The problem of constructing an attribute description of a business process with the automated formation of process models “as is” using logs of information systems in which the tracks of individual processes are not identified is considered. It is shown that to solve this problem, it is advisable to distinguish the distinctive properties of individual business processes represented by the attributes of log events. A method for constructing an attribute description of a business process is proposed. The method is based on the comparison of combinations of attributes for intervals of events of a fixed length and the subsequent selection of subsets of attributes with the same values. The method includes the steps of forming the intervals of events, constructing combinations of attributes for specified intervals, as well as calculating and subsequently averaging the weights of combinations of attributes on these intervals. The result of the method is a weight-ordered set of event attributes and their values, which takes into account the attribute and temporal aspects of the business process. The method creates conditions for a more efficient transition from functional to process management based on splitting the log into processes using the resulting attribute description and subsequent prototyping of business process models “as is” by means of process mining.

**Keywords:** business processes, process management, process trace, event log, event attributes.

## 1. Introduction

The process approach to enterprise management is based on the construction of business process models in the form of interrelated sequences of actions, allowing to achieve the set goals of the enterprise [1]. As part of this approach, a regular increase in the efficiency of the enterprise's activity is being implemented on the basis of continuous improvement of the running business processes [2].

Continuous improvement of business processes is performed based on a comparison of the current process model "as is" and the desired process model "as it should be". As a result of the comparison, a set of control actions is formed, allowing to modify the process, convert it to the desired form.

Obtaining a process model "as is" is carried out either on the basis of monitoring the activities of the executors, as well as their survey, or by means of process mining [3]. The first approach requires considerable time and can't always be automated [4]. The second approach is based on the analysis of the logs of existing information systems, which makes it possible to eliminate the human factor when building a business process model.

Logs of information systems that have a special structure act as input data when building process models. The key feature of the structure of such logs is that the events in them should be grouped by process [3, 5].

At the same time, the logs of real information systems do not always meet the specified requirements. On the one hand, traditional systems that support functional management do not fix the execution of processes. Therefore, their logs contain sequences of events related to various processes and functional tasks [6]. On the other hand, process-oriented system logs often contain incomplete or erroneous information that is not fully suitable for building a process model [3].

At the same time, the log contains additional information provided by the event attributes. This information is different for different processes, which indicates the possibility of dividing these events by process and further using process mining methods for building models "as is".

Thus, the problem of constructing an attribute description of a business process with the automated formation of process models "as is" using the logs of information systems is relevant.

## 2. Literature review

Existing methods for constructing business process models and their components "as is" as the source data use the logs of information systems, each of which contains a large amount of data on the business process, obtained as a result of its multiple execution [3, 7].

Structurally, the log consists of a sequence of traces. Each trace contains information about only one execution of a business process as a sequence of events reflecting the sequence of the process. Events contain information about the actions of the process, the performers and used resources [8]. The links between events reflect the knowledge embedded in the business process model, as well as the implicit knowledge of the performers. Such dependencies can be obtained by applying the approach proposed in [9]. Heuristic methods, methods of fuzzy logic and genetic programming are used to obtain process models in logs with errors [3, 10, 11].

However, in the case of functional control, the source log contains events from various processes or functional tasks in the form of a single sequence of events. It is not possible to build process models "as is" by traditional methods.

Approaches to the formation of process-oriented trails from the general sequence of events are proposed in [6, 12]. These approaches make it possible to isolate connections between log events in the event of significant differences in event attributes, but they require further development in order to effectively separate events from different processes with similar attributes, as well as similar attribute values.

To overcome this drawback, it is advisable to use an attribute description of a business process that contains characteristic sets of attributes and their values. This will allow to divide the overall sequence of events into ordered sets of events for different business processes by comparing the attributes of the next log event with attribute models of different business processes.

The purpose of this article is development of a method for constructing an attribute description of a business process in order to provide the possibility of forming process-oriented sequences of events from the information management system log, which implements a functional management approach.

### 3. The method of constructing an attribute description of a business process

The developed method uses as a source data a sequence of events  $e_1, e_2, \dots, e_j, \dots$ . These events relate to different business processes or to various functional tasks, but business processes are not specified in this sequence. Each event contains a time stamp and a set of attributes and their values, which characterizes the state of one of the business processes running in the enterprise:

$$e_j = \{t_j, A_j, B_j\}, \quad (1)$$

where  $t_j$  – the timestamp of the event  $e_j$ ;  $A_j$  – a set of attribute names;  $B_j$  – a set of attribute values. Attribute values are defined at specific points in time  $t_j$ :

$$\forall a \in A_j \exists \beta \in B_j | t_j. \quad (2)$$

As attributes  $a \in A_j$ , the action of a business process, the executor of this action, the product being processed, a division of an enterprise, etc are indicated. Therefore, the values  $\beta$  of different attributes at any time do not match:

$$\forall t_j \forall a, a^* \in A_j \beta \neq \beta^*. \quad (3)$$

The main idea of the method is searching for such combinations of event attributes that have associated sets of values. Such value sets characterize both process-specific data and typical processing sequences for this data. Let's consider an example where in the three attributes of several events the name of the same division is indicated, the same product being released, but the name of the operation changes. Such a sequence of attribute values indicates that they belong to the same business process.

The method uses a sliding interval  $Win_l = \langle e_1, \dots, e_{l+k} \rangle$ . This interval consists of a sequence of  $k$  events. This sequence searches for combinations of attributes that have associated values. The size of the interval (the number of events) depends on the particular enterprise. Obviously, the more processes run at the same time, the larger the window must be. In the window of a small number of events, the laws will not be visible, since it will contain data only from part of business processes or functional tasks. Therefore, the size of the event window is determined experimentally.

For each event interval, a combination of all possible attributes from  $H = \max_{Win} |A_j|$  by  $1 \leq h \leq H$  is formed and the weights of combinations of attributes with the same values are calculated. In the calculation, let's consider a set of events  $Win_{l,i}^h$  with the same set of values  $h$  for the attributes. It is possible to create several combinations  $h$  of attributes  $A_{l,i}^h$  from  $H$  event attributes. The index  $i$  is used to distinguish between these sets.

The number of events with the same values for  $h$  attributes for each combination of attributes  $A_{l,i}^h$  is defined as the power of the subset  $Win_{l,i}^h$ :

$$Win_{l,i}^h = \{e_j | \forall a_{j,m} \in A_{l,i}^h \beta_{j,m} = \beta_{j+1,m}\}, \quad 1 \leq j < |Win_{l,i}^h|, \quad (4)$$

where  $a_{j,m}$  –  $m$ -attribute of event  $e_j$ ;  $\beta_{j,m}$  – value of attribute  $a_{j,m}$ .

The calculation of the weight  $w_{l,i}^h$  of the set of events  $Win_{l,i}^h$  with attributes from the set  $A_{l,i}^h$  in the event window  $Win$  is calculated according to the expression:

$$w_{l,i}^h = \frac{|Win_{l,i}^h|}{|Win|}. \quad (5)$$

The results of calculations by intervals are averaged for the entire sequence of events:

$$W_i^h = \frac{\sum_l w_{l,i}^h}{|\{w_{l,i}^h\}|}. \quad (6)$$

Expression (6) is used only for a subset of intervals that have common values of the same attributes.

The input data of the method are: the initial sequence of events and the number of events in the interval  $k$ .

The developed method includes the following stages:

*Stage 1.* Select the event window  $\langle e_l, \dots, e_{l+k} \rangle$ , where  $l$  – the number of the current iteration.

*Stage 2.* Formation of a set of combinations of event attributes  $\{A_{l,i}^h\}$  for a given interval  $Win_l$ .

*Stage 3.* Formation of subsets of events  $Win_{l,i}^h$  with the same attribute values for each combination  $A_{l,i}^h$  according to expression (4).

*Stage 4.* Calculation of weights  $w_{l,i}^h$  for each subset  $Win_{l,i}^h$  according to expression (5).

*Stage 5.* Checking the end of the sequence of events – i. e.  $l+k$  must not exceed the length of the log.

If the sequence is not finished, the transition to the next iteration is performed:

5. 1. The current iteration index is incremented:  $l=l+1$ .

5. 2. Going to stage 1.

Otherwise, proceed to stage 6.

The overall result of steps 1–5 is the weighting table  $\{w_{l,i}^h\}$  for all event intervals and for all combinations of events at these intervals. Each weight  $w_{l,i}^h$  is associated with a list of attributes and their values for which this weight is calculated.

*Stage 6.* Averaging the weights  $w_{l,i}^h$  for several intervals according to (6). Weights are averaged only for those intervals on which events have common attribute values for the combination  $A_{l,i}^h$ .

The result of this stage is a set of weights  $W_i^h$  for  $i$  – combination of  $h$  attributes, as well as a set of event attributes sets from intervals for which the weight is calculated.

*Stage 7.* Ordering weights  $W_i^h$  by value.

The result of this stage allows to use only combinations of attributes with maximum weights when building business process models “as is”.

#### 4. Experiments

The purpose of the experiment is verification of the effectiveness of the method proposed in the article for source data with similar attributes. Experimental verification of the method is carried out on a synthetic sequence of events generated from the logs of several business processes. Such a log is generated in order to test the functionality of the algorithm for events with similar sets of events and the values of their attributes.

Each log event is described by the following set of attributes: “org:group”, “org:role”, “product”, “org:resource”, “organization involved”.

In the experiment, the method is performed three times for sequences of 3, 4, and 5 events. Attributes are compared in a sliding interval of a specified number of events. For example, the 1st–5th, then the 2nd–6th, etc. events of the original log are analyzed sequentially.

The weighted sets of attributes and their values resulting from the operation of the method are ordered in descending order of weights. The sets of attributes with the maximum value of weight are presented in **Table 1**.

Selected as a result of applying the method weighted sets of combinations of attributes and their values characterize subsequences of events with different properties that correspond to different business processes.

**Table 1**  
Results of the method operation

Event attribute sets	Weight of log event attribute combinations			Attribute values
	3 events	4 events	5 events	
“product”, “org:resource”	0.1511117	0.04333363	0.2762764	{PROD453, Denny} {PROD453, Paul}
“product”, “org:group”, “org:role”	0.10222296	0.0422224	0.2288988	{PROD453, N52 2nd, n} {PROD453, O3 3rd, n}
“org:resource” “org:group”	0	0	0.2120002	{Denny, N52 2nd} {Denny, O3 3rd}
“product”, “org:resource”, “org:group”	0	0	0.2080002	{PROD453, Denny, V30} {PROD453, Paul, V30} {PROD453, Denny, V30} {PROD453, Paul, V30}
“organization involved”, “org:group”, “org:resource”	0	0	0.190987	{Orgline A2, N52 2nd, Sofia} {Orgline A2, N52 2nd, Steve}
“organization involved”, “org:resource”, “org:group” “org:role”	0	0	0.190873	{Orgline C, Denny, N38, n} {Orgline C, Denny, N38, n}
“product”, “org:resource”, “org:group”, “org:role”	0	0	0.1802576	{PROD204, John, N52 2nd, n} {PROD204, Irwin, N52 2nd, n}

## 5. Results and discussion

The research result is a method for constructing an attribute description of a business process, focused on using an automated way to build business process models “as is” based on the analysis of logs of the information management system.

The weighted attribute sequences obtained as a result of the method’s operation determine the invariant of a separate process, its distinctive features as part of a common log for several business processes and functional tasks. This allows to separate the log events for individual processes and then build models of these processes using process mining tools. The resulting models are compared with “as it should be” models, built on the basis of a priori knowledge of the activities of the respective enterprise. Based on the results of the comparison, decisions are made on improving business processes.

The difference of the proposed method lies in the generalization of the characteristics of a business process based on the construction of a sequence of the most frequently repeated combinations of attributes of events of the original log. Such a generalization allows to obtain a description of the business process in the attribute and temporal aspects. In the attribute aspect, the groups of the most essential features are distinguished, which are represented by the values of the log event attributes. The time aspect is reflected by the sequence of values for the selected combination of attributes.

The advantage of the method is the ability to isolate the most significant attribute characteristics of individual business processes with regard to changes in their values, which makes it possible to separate the sequence of events by potential business processes even if the models of these processes have not yet been built. As a result, conditions are provided for building prototypes of business processes and transition to process management.

The disadvantage of the method is that the resulting attribute description of the business process largely depends on the specifics of the enterprise in which the log is generated. Therefore, the number of events in the sliding interval must be chosen experimentally for each enterprise. It should also be noted that the computational costs are nonlinearly dependent on the number of event attributes, since all possible combinations of these attributes are used.

The method can also be used to separate the data stream from several sources of information if this data is multidimensional – have a set of attributes.

## 6. Conclusions

The problem of constructing an attribute description of a business process with the automated formation of process models “as is” by analyzing logs is considered.

The feasibility of building such an attribute description is justified by the weak suitability of source logs from function-oriented control systems for the automated construction of business process models “as is”.

A method for constructing an attribute description of a business process based on an analysis of the information system log is proposed.

The proposed method has the following differences from traditional approaches to the analysis of logs.

First, the attribute and temporal aspects of the business process are taken into account. The attribute aspect defines the properties of the subject area, which are characteristic of a particular business process. The time dimension sets the order in which these properties change as the business process runs. The combination of these two characteristics allows to distinguish prototypes of business processes, even in the absence of process management in the enterprise.

Secondly, the method allows to build several alternative attribute descriptions of business processes and thereby increase the efficiency of designing the enterprise business process system based on the selection and improvement of one of the proposed options, taking into account the semantics of the attributes specified in the description.

In the practical aspect, the developed method creates conditions for improving the efficiency of process management based on comparing business process models “as is” and “as it should be” and improving the latter. The method also provides the possibility of more efficient transition from functional to process management based on the selection of event sequences and the construction of prototypes of process models by means of process mining.

---

#### References

- [1] Dumas, M., La Rosa, M., Mendling, J., Reijers, H. A. (2013). *Fundamentals of Business Process Management*. Springer, 400. doi: <https://doi.org/10.1007/978-3-642-33143-5>
- [2] Van der Aalst, W. M. P. (2013). *Business Process Management: A Comprehensive Survey*. ISRN Software Engineering, 2013, 1–37. doi: <https://doi.org/10.1155/2013/507984>
- [3] Van der Aalst, W. (2016). *Process Mining: Data Science in Action*. Springer-Verlag, 467. doi: <https://doi.org/10.1007/978-3-662-49851-4>
- [4] Vom Brocke, J., Rosemann, M. (Eds.) (2015). *Handbook on Business Process Management 1. Introduction, Methods, and Information Systems*. Berlin: Springer-Verlag. doi: <https://doi.org/10.1007/978-3-642-45100-3>
- [5] Mannhardt, F., de Leoni, M., Reijers, H. A., van der Aalst, W. M. P., Toussaint, P. J. (2016). From Low-Level Events to Activities – A Pattern-Based Approach. *Business Process Management*, 125–141. doi: [https://doi.org/10.1007/978-3-319-45348-4\\_8](https://doi.org/10.1007/978-3-319-45348-4_8)
- [6] Chalyi, S., Levykin, I., Petrychenko, A., Bogatov, I. (2018). Causality-based model checking in business process management tasks. 2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT). doi: <https://doi.org/10.1109/dessert.2018.8409176>
- [7] Kalenkova, A. A., van der Aalst, W. M. P., Lomazova, I. A., Rubin, V. A. (2015). Process mining using BPMN: relating event logs and process models. *Software & Systems Modeling*, 16 (4), 1019–1048. doi: <https://doi.org/10.1007/s10270-015-0502-0>
- [8] Tax, N., Sidorova, N., Haakma, R., van der Aalst, W. M. P. (2016). Mining local process models. *Journal of Innovation in Digital Ecosystems*, 3 (2), 183–196. doi: <https://doi.org/10.1016/j.jides.2016.11.001>
- [9] Kalynychenko, O., Chalyi, S., Bodyanskiy, Y., Golian, V., Golian, N. (2013). Implementation of search mechanism for implicit dependences in process mining. 2013 IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS). doi: <https://doi.org/10.1109/idaacs.2013.6662657>
- [10] Van der Aalst W. M. P. (2013). A General Divide and Conquer Approach for Process Mining. *Federated Conference on Computer Science and Information Systems (FedCSIS 2013)*. Available at: [http://www.processmining.org/\\_media/blogs/pub2013/2-fedcis-wvdaalst-keynote.pdf](http://www.processmining.org/_media/blogs/pub2013/2-fedcis-wvdaalst-keynote.pdf)
- [11] Van der Aalst, W. M. P. (2018). Process discovery from event data: Relating models and logs through abstractions. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8 (3), e1244. doi: <https://doi.org/10.1002/widm.1244>
- [12] Chalyi, S. F., Bogatov, E. O., Pribylnova, I. B. (2018). Techniques of reordering traces in the event logs in business process management tasks. *Bulletin of National Technical University “KhPI”. Series: System Analysis, Control and Information Technologies*, 21, 43–47. doi: <https://doi.org/10.20998/2079-0023.2018.21.08>