

Method of similarity business processes by genetic tagging for warehouses of processes and process mining

JAROSŁAW KOSZELA

Faculty of Cybernetics, Institute of Computer and Information Systems

Military University of Technology

Sylwestra Kaliskiego 2, 00-908 Warsaw,

POLAND

jaroslaw.koszela@wat.edu.pl, <http://www.isi.wat.edu.pl/pl/pracownicy/jaroslaw-koszela>

Abstract: - The article presents method of business processes similarity evaluation utilizing genetic tag sequences as a linearity definition of business processes. This method is an element of BI systems type for business processes analysis and process warehouse. This method describe transformation of business process to “genetic” tags sequences and evaluation of similarity measures for genetic sequences as a part of process mining and warehouse of processes.

Key-Words: business process, warehouse of processes, process mining, similarity of processes method, genetic tagging

1 Introduction

The presently applied methods of modeling, designing and creating IT systems are based on three basic approaches: structural, object and process. In the structural and object approach, the system structure and resources come to the fore, whereas the issue describing the processing dynamics stays in the background, even though this aspect is equally important in terms of a complete and consistent description of the modeled system. The situation is different in case of the process approach, where the priority is the description of the processing dynamics in the form of business process models, whereas the system models as well as the descriptions of the system structures and resources recede into the background.

The most popular currently applied notations of the business process is the BPMN v2.0 (Business Process Model and Notation) [1-3] standard, developed by the BPMI (Business Process Management Initiative) and OMG (Object Management Group) in January 2011. The standard defines a set of symbols and their semantics, allowing to model a diagram of the business process definition in a graphical form. The graphical notation of the process, presented in the BPMN standard, is in human-readable form. On the other hand, the following formats are used for the purpose of a notation allowing clear interpretation: XPD (XML Process Definition Language) [4] developed by WfMC (Workflow Management Coalition) and adapted for the purpose of exchanging process

definitions noted in this format between various systems, and the BPDM format (Business Process Definition Metamodel) developed by OMG, which described meta-models defining concepts, correlations and their semantics to exchange the business process models between different modeling processes. Both these formats are based on data notation in XML (Extensible Markup Language), developed by XSD (XML Schema Definition) and XMI (XML Metadata Interchange).

The modeling of business processes is a defining stage of the model that describes operations of a given organization or merged or collaborating elements of the modeled reality fragments. To perform the modeling activities in a manner allowing on one hand, their actual and physical embodiments, and on the other - necessary or desired efficiency, it is crucial to apply the automation tools. The group shall include process automation systems - the workflow systems, in particular the workflow management systems - WFMS (Workflow Management System) [5]. The main purpose of such IT systems is to process the process instances, according to the modeled business process definition. It should be mentioned that the main task realized by the WFMS systems is the processing of processes in accordance with their definitions, not activities described thereby. The implementing systems, independent of WFMS, or specialized services are responsible for performing activities and works described in the processes.

The basic scope and tasks related to the business processing IT systems have been presented by the WfMC organization, which defines workflow in the following way:

“The automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules” [6].

On the other hand, the workflow management system - WFMS - is defined by WfMC as:

“A system that defines, creates and manages the execution of workflows through the use of software, running on one or more workflow engines, which is able to interpret the process definition, interact with workflow participants and, where required, invoke the use of IT tools and applications” [6].

2 BI systems for process analysis

The system for the business process analysis (OLAP-BP) is categorized as Business Intelligence (BI), in case of solutions based or significantly supported by business processes [7-10]. The process modeling is an activity that requires constant modification, improvement and optimization of such processes. The variability of the environment, legislation, organization, economic situation makes it necessary to continuously introduce changes and upgrades of the process to increase its efficiency in broad context. The PDCA model [11] is used for the business process management. The model describes the process life cycle and consists of four main steps: Plan, Do, Check, Act. The model presents a closed cycle of the process management, where the processes should be all the time modified and upgraded. The developed processes may be automated, analyzed and explored through the process automation systems (workflow systems) [1,3],[5-10],[17-26],[28,33]. The use of the process warehouse, including process exploration, makes it possible to perform an in-depth analysis of the processes gathered in the warehouse [7,8], by applying special tools for process exploration, which may use various methods, e.g. analytical, forecast, simulation [12], structural, semantic, etc. The process warehouse as well as process exploration tools should be included in the group of the Business Intelligence systems.

2.1 Process warehouse – architecture

The functioning of the process warehouse in the organization is aimed at streamlining the management and decision-making processes by

providing current data and information regarding the implementation of business processes. The main tasks of the process warehouse are, among other things, the following [18]:

- integration with the process runtime environment, allowing to transfer data in time cycles and via an online model;
- possible extraction and upload of the processes and process instances;
- processing of data with the process definitions and instances derived from the process runtime environments, in cycle and event modes (online);
- storage of data on business process, which include, among other things, the following:
- collection of unified business process definitions and instances in terms of their structure and semantics,
- collection of metadata and semantic data concerning business processes,
- collection of indirect data and results of analytical processing and process exploration,
- in forms allowing to analytically process the data, while ensuring expected capacity, in particular
- analysis of deviations of the process instances from sample paths,
- identification of deadlocks and bottlenecks in the process instances,
- analysis of execution times of the process instances and particular components,
- analysis of the decisions made during the process implementation,
- identification of the so-called dead process or process fragments and resources, which are not implemented or are implemented very rarely,
- definition and analysis of the process similarity measures as well as event sequences,
- identification of new processes on the basis of the process execution logs,
- identification of patterns as well as rare and frequent event sequences,
- provision of apparatus and tools allowing the process analytical processing and exploration.

The process warehouse is a system which, through its software, analysis methods and process exploration, allows efficient analytical processing and exploration of the collected business processes and process instances. Fig. 1 shows the general architecture model of the process warehouse, in

which the following basic subsystems may be distinguished:

- extract, transformation and load from different sources of various formats of the definition notation regarding processes and executed instances of business process, together with their structural and semantic unification,
- process warehouses as thematic composition of the unified process and process instances,
- analyses and exploration of processes and process instances,
- access to the process warehouse allowing interactive questioning and answering in the form of reports, tables, charts, etc.

process warehouse, in which the following basic subsystems may be distinguished:

- extract, transformation and load from different sources of various formats of the definition notation regarding processes and executed instances of business process, together with their structural and semantic unification,
- process warehouses as thematic composition of the unified process and process instances,
- analyses and exploration of processes and process instances,
- access to the process warehouse allowing interactive questioning and answering in the form of reports, tables, charts, etc.

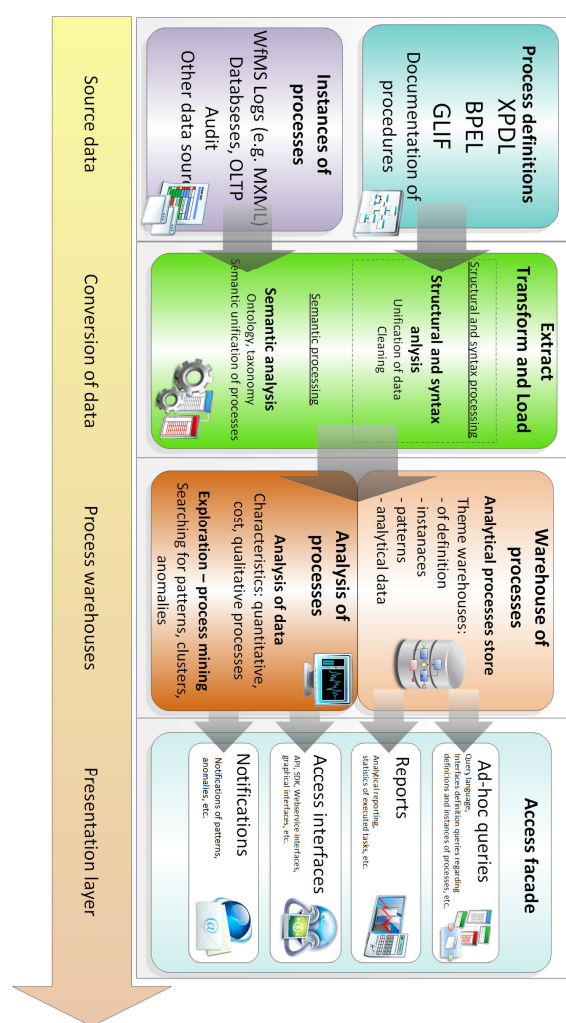


Fig. 1. Architecture of process warehouse

The process warehouse is a system which, through its software, analysis methods and process exploration, allows efficient analytical processing and exploration of the collected business processes and process instances. Fig. 3 shows the general architecture model of the

2.2 Process warehouse – analytical data store of the process definitions and instances

The process warehouse is an IT tool used for collecting, accessing and efficient processing of the business processes and process instances for the purpose of their analysis and exploration. The physical composition of the process warehouse includes the components of the process definitions and instances in the form of unified data, metadata, semantic data and analytical data (Fig. 2).

Warehouse of the unified processes and instances - a component responsible for the storage of data on the process definitions and instances, pursuant to the effective standards, e.g. XPDL [4,14], MXML [16], XES [15], which constitutes grounds for generating data on the processes stored in the analytical data warehouse. As part of the preliminary data, two variants of the warehouse structures were developed and tested - the original variant based on the physical warehouse of the matrix process structures and the warehouse using Neo4j graph database [31].

Warehouse of metadata and semantic data - a component responsible for the storage of semantic models expressed in ontological languages, which ensures, among other things, a possibility of expressing complex links between the model elements and a possibility of expressing logical equivalence of the terms and instances as well as open world assumptions, by introducing a specific manner of interpreting the lack of knowledge in the model. Correlations between various elements of the process warehouse - a component that implements metamodel of the processes warehouse and that is responsible for storing data on the links binding individual elements of the process warehouse, ensuring its coherence.

Analytical data and process warehouses - a component responsible for storing data on the

processes in the form of graphs and networks in the form of matrices, using the network databases - this method for representing data on the process definitions and instances allows to apply a mathematical apparatus for the purpose of the process analyses and streamlines analytical processing in terms of its capacity.

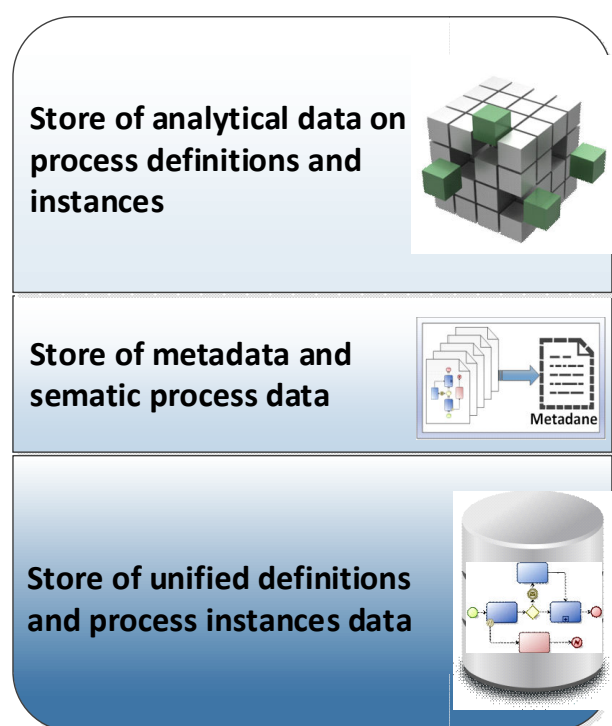


Fig. 2. Process warehouse model - storing processes and instances

Fig.3 shows a sample business process noted in graphical notation of the BPMN v2.0 standard, implemented with the use of the process modeling tools – Bizagi Modeler v3.1 [13] and the same process noted in the formal data structure, i.e. XPD. The developed transformation tool is used for the business process noted in the XPD format to be translated into Neo4j's Cypher Query Language [31,32], applied in the network environment. When comparing a visual embodiment of the process noted in BPMN v2.0 (Fig.3) and another visualization of the result after the transformation of such process saved in Neo4j database, it may be observed that the basic information included in the definition of such process was included in the network database.

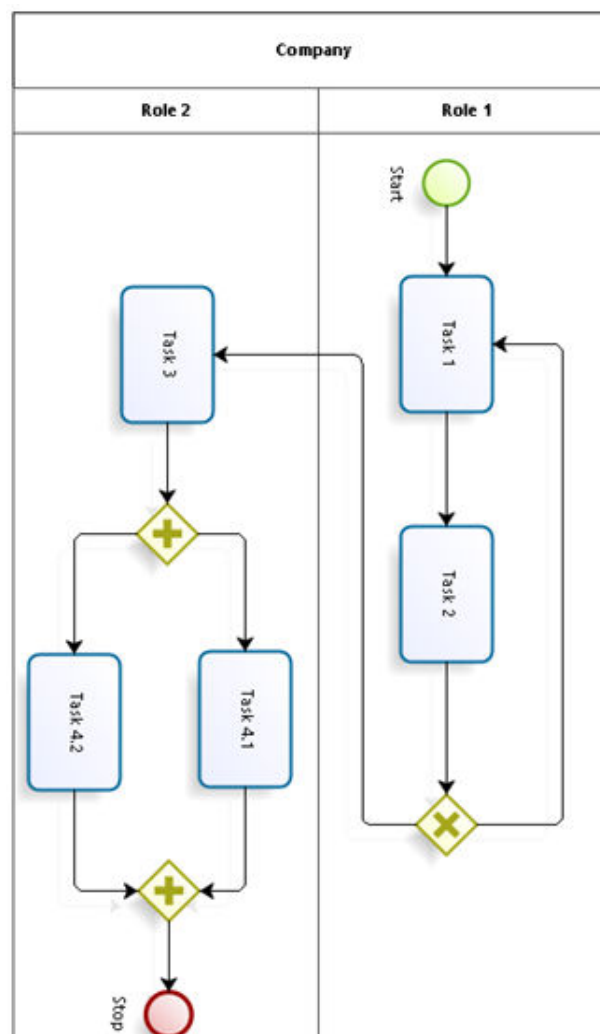


Fig. 3. Sample process in BPMN v2.0

3 Method of similarity processes by genetic tagging

One of the main groups of the process analysis methods are methods used for examining similarities between the business process definitions and instances. The methods directly based on the graph and network theories [9, 27, 29, 30] as well as mixed methods may be included in the above-mentioned group. The methods are usually quite complex in terms of calculations, which means that their application for the analysis of a large number of the process definitions and instances may be rather costly as far as calculations are concerned. Therefore, the process similarity method by genetic tagging is aimed at increasing the efficiency of the processes similarity analysis in case of a significant number of the processes. The method is based on the process transformation from the form based on graphs to the form of element sequences (genes), which correspond to the model form. Generally, the

similarity analysis is aimed at analysing two "genetic" sequences of the processes.

The first step is to transform the notation of the process definition from the graph (matrix) form to the linear tag sequence. Such transformation should include the relationship of a certain order, which shall transform two different notations of the same process definition into a uniform and identical process tag sequence.

$$\begin{aligned}
 G &= (V, E) \\
 E &\subseteq \{ \langle v_s, h, v_e \rangle : v_s, v_e \in V, h \in H \} \\
 v &= \{ \langle a, s \rangle : v \in V, a \in A, s \in S \}
 \end{aligned} \tag{1}$$

The process defined as graph G, in which a set of vertices V denotes a set of the process active elements (i.a. process, activities, gates, events), vs,ve represents start and end vertices. Set E contains edges - the flowing of controls and messages between the process active elements, set H means type of flow. An element of the set of vertices V is defined as a well-ordered pair of elements, where s means semantically unified feature of the process, and a means an element of set A defined as type of the process active element.

The process genetic tag sequence T and its well-ordered elements sequence t_j are described in the following form:

$$\begin{aligned}
 T &= (t_1, t_2, \dots, t_j, \dots, t_J) \\
 t_j &= (v_s, h, v_e)
 \end{aligned} \tag{2}$$

where J is size of set E.

To define order for genetic sequence of the business process, we introduce a mapping f(v) which includes graph layers of the process, and assigns types of active elements to natural numbers such that:

$$f(v) : V \rightarrow \{1, 2, \dots, |A|\} \tag{3}$$

Additionally the function satisfies the following:

$$\bigwedge_{\substack{v_1 = \langle a_1, s_1 \rangle, \\ v_2 = \langle a_2, s_2 \rangle \in V}} (a_1 \neq a_2) \rightarrow (\sim(f(v_1) = f(v_2))) \tag{4}$$

Analogously, a mapping g(h) must be applied to the types of transitions between flow active elements, such that:

$$g(h) : H \rightarrow \{1, 2, \dots, |H|\} \tag{5}$$

The mapping fulfils various conditions such that the mapping g(h) does not take the same value for any two different elements from the set of types of passages between the flow elements, i.e.:

$$\bigwedge_{h_1, h_2 \in H} (h_1 \neq h_2) \rightarrow (\sim(g(h_1) = g(h_2))) \tag{6}$$

To arrange the T chain, we use the order relationship between the elements $t_1 = \langle v_{s_1}, h_1, v_{e_1} \rangle$ and $t_2 = \langle v_{s_2}, h_2, v_{e_2} \rangle$:

$$\begin{aligned}
 t_1 &< t_2 : \\
 &f(v_{s_1}) < f(v_{s_2}) \\
 &\vee \left(\left(f(v_{p_1}) = f(v_{p_2}) \right) \right. \\
 &\quad \left. \wedge (g(h_1) < g(h_2)) \right) \\
 &\vee \left(\left(f(v_{p_1}) = f(v_{p_2}) \right) \right. \\
 &\quad \wedge (g(h_1) = g(h_2)) \\
 &\quad \left. \wedge (f(v_{k_1}) < f(v_{k_2})) \right)
 \end{aligned} \tag{7}$$

Transformation of Fig. 3 process noted in XPDL to the structure of the process genetic tag sequence is as follows:

<St1.Start; Tr1; Ta1.Task 1>, <Ta1.Task 1; Tr1; Ta1.Task 2>, <Ta1.Task 2; Tr1; Ga2.>, <Ta1.Task 3; Tr1; Ga3.>, <Ta1.Task 4.2; Tr1; Ga3.>, <Ta1.Task 4.1; Tr1; Ga3.>, <Ga2.; Tr1; Ta1.Task 1>, <Ga2.; Tr1; Ta1.Task 3>, <Ga3.; Tr1; Ta1.Task 4.1>, <Ga3.; Tr1; Ta1.Task 4.2>, <Ga3.; Tr1; En1.Stop>

For example, a slight modification of the process presented in Fig. 3 may be introduced by adding one activity, and we obtain the processes shown in Fig. 4. More modification of the process represented in Fig. 3 are presented in Fig. 5 and Fig. 6.

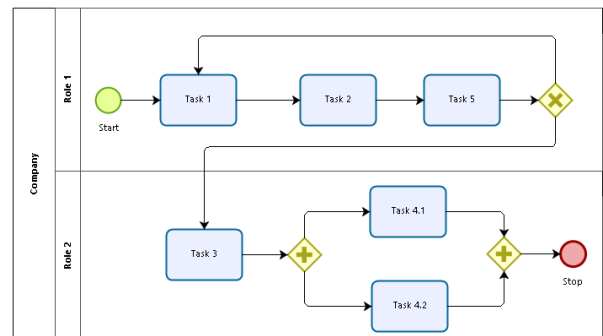


Fig. 4 Modified process Fig.3 (Task 5)

Process genetic tag sequence Fig. 4 is as follows:
 <St1.Start; Tr1; Ta1.Task 1>, <Ta1.Task 1; Tr1; Ta1.Task 2>, <Ta1.Task 2; Tr1; Ta1.Task 5>, <Ta1.Task 5; Tr1; Ga2.>, <Ta1.Task 3; Tr1; Ga3.>, <Ta1.Task 4.1; Tr1; Ga3.>, <Ta1.Task 4.2; Tr1; Ga3.>, <Ga2.; Tr1; Ta1.Task 1>, <Ga2.; Tr1; Ta1.Task 3>, <Ga3.; Tr1; Ta1.Task 4.1>, <Ga3.; Tr1; Ta1.Task 4.2>, <Ga3.; Tr1; En1.Stop>

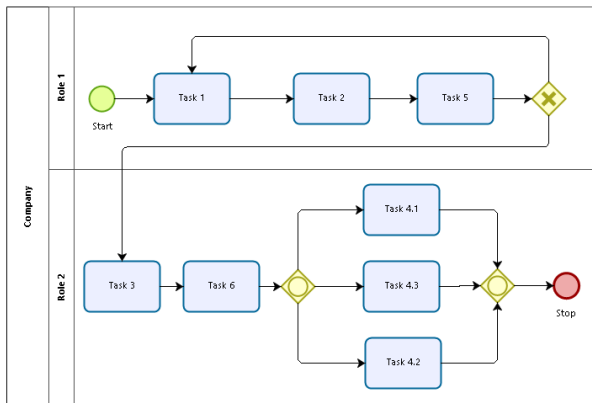


Fig. 5 Modified process Fig.3 (Task 5, Task 6, Task 4.3, gateways)

Process genetic tag sequence Fig. 5 is as follows:
 <St1.Start; Tr1; Ta1.Task 1>, <Ta1.Task 2; Tr1; Ta1.Task 5>, <Ta1.Task 1; Tr1; Ta1.Task 2>, <Ta1.Task 3; Tr1; Ta1.Task 6>, <Ta1.Task 5; Tr1; Ga2.>, <Ta1.Task 4.2; Tr1; Ga4.>, <Ta1.Task 4.1; Tr1; Ga4.>, <Ta1.Task 4.3; Tr1; Ga4.>, <Ta1.Task 6; Tr1; Ga4.>, <Ga2.; Tr1; Ta1.Task 1>, <Ga2.; Tr1; Ta1.Task 3>, <Ga4.; Tr1; Ta1.Task 4.2>, <Ga4.; Tr1; Ta1.Task 4.1>, <Ga4.; Tr1; Ta1.Task 4.3>, <Ga4.; Tr1; En1.Stop>

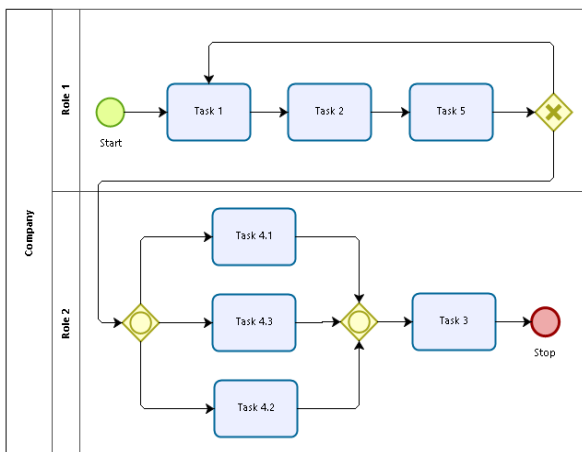


Fig. 6 Modified process Fig.3 (Task 5, Task 4.3, Task 3, gateways)

Process genetic tag sequence Fig. 6 is as follows:
 <St1.Start; Tr1; Ta1.Task 1>, <Ta1.Task 1; Tr1; Ta1.Task 2>, <Ta1.Task 2; Tr1; Ta1.Task 5>, <Ta1.Task 5; Tr1; Ga2.>, <Ta1.Task 4.1; Tr1; Ga4.>, <Ta1.Task 4.3; Tr1; Ga4.>, <Ta1.Task 4.2; Tr1; Ga4.>, <Ta1.Task 3; Tr1; En1.Stop>, <Ga2.; Tr1; Ta1.Task 1>, <Ga2.; Tr1; Ga4.>, <Ga4.; Tr1; Ta1.Task 4.2>, <Ga4.; Tr1; Ta1.Task 3>, <Ga4.; Tr1; Ta1.Task 4.3>, <Ga4.; Tr1; Ta1.Task 4.1>

The so prepared process genetic sequence is in the linear form of the input data to establish the similarity measures. Basic process similarity measures have been developed on the basis of the similarities established for the corresponding process genetic tag sequence:

- General numerical similarity measure (GNM) - the measure which defines the process similarity in terms of a quotient of occurrences of the types of tags in the particular process genetic tag sequences.

$$GNM = \frac{\sum^i F_L(i)}{\sum^i F_M(i)} \tag{8}$$

$$F_L(i) = \begin{cases} 0, & \text{when } w_i = p_i = 0 \\ p_i, & \text{when } p_i \leq w(i) \\ z, & \text{when } p(i) > w_i \wedge z \geq 0 \end{cases} \tag{9}$$

$$F_M(i) = \begin{cases} 0, & \text{when } w_i = p_i = 0 \\ w_i, & \text{when } p_i \leq w_i \vee (p_i > w_i \wedge z \geq 0) \\ |z|, & \text{when } p_i > w_i \wedge z < 0 \end{cases}$$

where:

- w_i – number of tag instances (triples) of type i in base genetic tag sequences,
- p_i – number of tag instances (triples) of type i in pattern genetic tag sequences.

- Detailed numerical similarity measure (DNM) - the measure which is similar to the general numerical similarity measure, except that the tags of the same type are those which have the same type of the initial active elements, including their properties, the same type of transition between the active elements and the same type of the final active element, including their properties.

- General chain similarity measure (GCM) - the measure which is aimed at defining the level of similarity between two business processes noted as the genetic tag sequence, based on the number of similar subchains. The ratio of the number of positive comparisons of the subchains with the general number of the compared subchains is the results of the above-mentioned comparison. The subchains are properly compared if all of their corresponding three are of the same type.

$$GCM = \frac{\sum_{d=1}^l (\sum_{L_1-i}^{i_1} (\sum_{L_2-i}^{i_2} F(i_1, i_2, l)))}{\sum_{d=1}^l (\sum_{L_1}^{i_1} (\sum_{L_2}^{i_2} 1))} \tag{10}$$

where:

$d = \min(L_1, L_2)$ - size of shorter chain,

$F(i_1, i_2, l)$

$$= \begin{cases} 0, & \text{when } L_1(i_1 + l) \neq L_2(i_2 + l) \\ 1 * F(i_1, i_2, l - 1), & \text{when } L_1(i_1 + l) = L_2(i_2 + l) \end{cases}$$

- Detailed chain similarity measure (DCM) - the measure is similar to the general chain similarity measurement, except that - when comparing particular tags in subchains - the tags of the same type are those which have the same type of the initial active elements, including their properties, the same type of transition between the active elements and the same type of the final active element, including their properties.

Values of such similarity measurements are scalar values - number from the range [0,1], where 1 means that the compared processes are identical in terms of a given similarity measure. When applying the aforementioned process similarity measures, based on the similarities of genetic tags for the processes shown in Fig. 3, Fig. 4, Fig. 5 and Fig. 6, we obtain the following values of the similarity measures:

- general numerical similarity measure:

	Process Fig.5	Process Fig.6	Process Fig.7	Process Fig.8
Process Fig.5	1,0000	0,9091	0,6667	0,6154
Process Fig.6	0,9167	1,0000	0,7500	0,7143
Process Fig.7	0,7333	0,8000	1,0000	0,7059
Process Fig.8	0,6667	0,7333	0,6667	1,0000

Table 1. Values of GNM

- detailed numerical similarity measure:

	Process Fig.5	Process Fig.6	Process Fig.7	Process Fig.8
Process Fig.5	1,0000	0,9091	0,2000	0,1500
Process Fig.6	0,9167	1,0000	0,2500	0,2381
Process Fig.7	0,2381	0,2857	1,0000	0,5882
Process Fig.8	0,1500	0,2000	0,5333	1,0000

Table 2. Values of DNM

- general chain similarity measure:

	Process Fig.5	Process Fig.6	Process Fig.7	Process Fig.8
Process Fig.5	1,0000	0,7532	0,5273	0,2929
Process Fig.6	0,8636	1,0000	0,5702	0,3824
Process Fig.7	0,7273	0,7051	1,0000	0,3333
Process Fig.8	0,4091	0,4872	0,3697	1,0000

Table 3. Values of GCM

- detailed chain similarity measure:

	Process Fig.5	Process Fig.6	Process Fig.7	Process Fig.8
Process Fig.5	1,0000	0,7532	0,0909	0,0707
Process Fig.6	0,8636	1,0000	0,1228	0,1078
Process Fig.7	0,1212	0,1538	1,0000	0,3238
Process Fig.8	0,1061	0,1538	0,3109	1,0000

Table 4. Values of DCM

4 Conclusion

The above material shows the results of the studies carried out in the field of modeling, implementation and analysis of the systems based on the process approach as well as systems of analysis and automation of business processes. The presented subject related to the business processes, in particular their modeling, analysis and automation is a relatively new area of research in both IT and management. However, the implementation of the described processes in both these fields stimulates their development dynamics, as the increased demand for more efficient activities may be a powerful engine for economic growth.

The presented aspects of the systems, tools and methods of analytical business processing should be considered as an introduction to further development of this subject matter in various directions. The directions include issues related to the tools and methods for collecting data on the process definitions and instances as well as tools and methods for analytical processing in different dimensions and at different levels.

References:

- [1] M. Weske, *Business Process Management. Concepts, Languages, Architectures*, Berlin, Springer-Verlag, 2007.
- [2] OMG, BPMN - standard documents, [Online] <http://www.omg.org/spec/BPMN/2.0/>, 2011.
- [3] A. Arsanjani, N. Bharade, M. Borgenstrand, P. Schume, J. K. Wood i V. Zheltonogov, *Business Process Management Design Guide. Using IBM Business Process Manager*, USA, IBM Corporation, 2015.
- [4] WfMC, XML Process Definition Language – standard XPD L documents, [Online] <http://www.xpdl.org/> (2012)
- [5] W. van der Aalst, K. van Hee, *Workflow Management. Models, Methods, and Systems*, The MIT Press Cambridge, Massachusetts London, 2002.
- [6] WfMC, *Terminology & Glossary - WfMC-TC-1011*, [Online], http://wfmc.org/docs/TC-1011_term_glossary_v3.pdf, 1999.
- [7] G. Bliźniuk, M. Chmielewski, T. Gzik, J. Koszela, Warehouse of processes, *Studia Informatica*, Vol. 33-2A, 2012, pp. 111-127.
- [8] P. Zadora, *Hurtownie danych w organizacji inteligentnej* (Akademia Ekonomiczna, Katowice, 2004)
- [9] W. van der Aalst, *Process Mining*, Springer-Verlag, Berlin, 2016.
- [10] B. Gawin, *Systemy informatyczne w zarządzaniu procesami Workflow*, PWN, Warszawa, 2015.
- [11] R. Moen, C. Norman, *Evolution of the PDCA Cycle*, 2009
- [12] A. Sidnev, J. Tuominen i B. Krassi, *Business Process Modeling and Simulation*, Otamedia Oy, Espoo, 2005.
- [13] Bizagi, [Online] <http://www.bizagi.com>, 2017.
- [14] WfMC, [Online] <http://www.wfmc.org>, 2017.
- [15] XES, [Online] <http://www.xes-standard.org>, 2017.
- [16] MXML-Mining eXtensible Markup Language, [Online] <http://www.processmining.org/logs/mxml>, 2003.
- [17] M. Kiedrowicz, *Zarządzanie informacjami wrażliwymi - wybrane aspekty organizacyjne, prawne i techniczne ochrony informacji niejawnych*, WAT, Warszawa, 2015.
- [18] J. Koszela, Business process modeling for processing classified documents using RFID technology in *20th International Conference on Circuits, Systems, Communications and Computers*, Vol. 76, 2016.
- [19] R. Waszkowski, M. Kiedrowicz, T. Nowicki, Z. Wesołowski, and K. Worwa, Business processes in the RFID-equipped restricted access administrative office, in *20th International Conference on Circuits, Systems, Communications and Computers*, Vol. 76, 2016.
- [20] M. Kiedrowicz, Uogólniony model danych w rozproszonych rejestrach ewidencyjnych, *Roczniki Kolegium Analiz Ekonomicznych*, Vol. 33, 2014, pp. 209-234.
- [21] M. Kiedrowicz, The importance of an integration platform within the organisation, *Zeszyty Naukowe*, Vol. 46, 2014, pp.83-94.
- [22] M. Kiedrowicz, Dostęp do publicznych zasobów danych - Big data czy Big brother, (in:) INTERNET. Publiczne bazy danych i Big data, (ed.) G. Szpor, 2015, pp. 15-39.
- [23] M. Kiedrowicz, R. Waszkowski, Business rules automation standards in business process management systems, (in:) Information Management in Practice, (eds) B.F. Kubiak and J. Maślankowski, 2015, pp. 187-200.
- [24] M. Kiedrowicz, T. Nowicki, R. Waszkowski, Z. Wesołowski, and K. Worwa, Software simulator for property investigation of document management system with RFID tags, *20th International Conference on Circuits, Systems, Communications and Computers, MATEC Web of Conferences*, Cedex A: E D P Sciences, 2016, Vol. 76.
- [25] M. Kiedrowicz, J. Koszela, Business processes modelling for the processing of classified documents using RFID technology, *Collegium of Economic Analysis Annals*, Vol. 42, 2016, pp: 53-66.
- [26] M. Kiedrowicz, T. Nowicki, R. Waszkowski, Business process data flow between automated and human tasks, *3rd International Conference on Social Science (ICSS 2016)*, 2016, pp. 471-477.
- [27] Dijkmana R., *Similarity of business process models: Metrics and evaluation*, Information Systems, Elsevier, 36, 2011.
- [28] M. Kiedrowicz, J. Koszela, *Secret office model for the processing of classified documents*

using *RFID technology*, Collegium of Economic Analysis Annals, Vol. 42, 2016, pp. 67-81.

- [29] M. Ehrig, A. Koschmider, A. Oberweis, *Measuring Similarity between Semantic Business Process Models*, Conferences in Research and Practice in Information, 67, Australian Computer Society, 2007.
- [30] R. Dijkman, M. Dumas, L. García-Bañuelos, *Business Process Graphs: Similarity Search and Matching*, Graph data management: techniques and applications, IGI Global, 2012.
- [31] Neo4j, [Online] <https://neo4j.com>, 2017.
- [32] I. Robinson, J. Webber, E. Eifrem, *Graph Databases*, O'Reilly Media, Sebastopol, 2015.
- [33] W. van der Aalst & others, *Process Mining Manifesto*, IEEE Task Force on Process Mining, [Online] <http://www.win.tue.nl/ieeetfpm/downloads/Process%20Mining%20Manifesto.pdf>, 2017.