УДК 004.91
DOI: 10.20535/SRIT.2308-8893.2018.4.04

ESTIMATION AND ANALYSIS OF BUSINESS PROCESS MODELS SIMILARITY IN ENTERPRISE CONTINUUM REPOSITORY

A.M. KOPP, D.L. ORLOVSKYI

Abstract. This paper considers the problem of the store, share, and reuse of organizational knowledge represented using business process models. Various studies related to managing large collections of business process models are reviewed. The core concept of Business Process Model Repository was outlined as well as the reference architecture provided in related works. This research is focused on considering the Business Process Model Repository as part of the whole Architecture Repository defined in the field of Enterprise Architecture. The knowledge-based model used to store process models, as well as the similarity measure used to identify process models in the repository that are similar to a given process model or a fragment thereof are proposed. Besides that, the elaborated approach proposes the decision tree model for business process models classification according to the Enterprise Continuum concept of Enterprise Architecture, as well as the conceptual model of the Business Process Model Repository. The software prototype developed to implement the proposed approach was used to upload sample process models and estimate their similarity according to the Enterprise Continuum categories. The accuracy of the proposed similarity measure is analyzed for the different Enterprise Continuum categories of artifacts.

Keywords: business process model, similarity measure, organizational knowledge, repository, enterprise continuum.

INTRODUCTION

At higher levels of BPM (Business Process Management) maturity, a lot of organizations tend to accumulate considerable amounts of business process models [1]. Thus, business process model repositories might contain hundreds or even thousands models represented using various modeling notations [2].

A Business Process Model Repository offers organizations a space for storing, maintaining, and changing process knowledge (business rules, relationships, process elements, etc.) for future reuse. Also it enables business users to retrieve process models for various purposes like understanding, updating, simulating, and analyzing business process models [3]. The Business Process Model Repository also might be considered as the software for storing, managing and sharing of process models for future reuse [4].

Another area where the repository concept appears is EA (Enterprise Architecture). It is an important concept of an extremely popular architectural framework TOGAF (The Open Group Architecture Framework). The Architecture Repository can be used to store diverse types of architectural outputs, each at varying levels of abstraction [5]. Whereas TOGAF supports four architectural domains, business process models belong to the Business Architecture domain.

Thus, to support an iterative cycle of business process models transformation from reference models to organization-specific models and their further reuse as building blocks, the Business Process Model Repository concept should be considered as part of the whole Architecture Repository.

Since business process modeling technique is used to describe knowledge about organizational activities, the problem of store, share, and reuse of organizational knowledge, represented using business process models, becomes relevant. Hence in this paper the similarity measure used to retrieve process models from the repository in order to their further reuse in a business process continuous improvement cycle according to BPM concept is proposed.

RELATED WORK

In the study [6] authors noted that collections that contain hundreds or even thousands of business process models become more common for organizations that describe their operations in terms of business processes. They analyzed existing business process model repositories, which provide specific functions for managing collections of process models, such as managing the consistency and extracting knowledge from existing processes to better design new processes. As a result, they have proposed a framework for repositories that assists in managing large collections of business process models.

Quite similar ideas of the Business Process Model Repository are shown in papers [2, 3, 4]. Elias in [4] proposes the open and language-independent process model repository, which allows any potential users to capture, share, and reuse of process models. As the central function of the repository, author of [4] called supporting reuse of process models among different stakeholders, across organizations and industries. Studies [3, 4] also provide requirements for the business process models repository with considering its place and role in the BPM lifecycle.

Authors of paper [2] proposed the reference architecture for the Business Process Model Repository, which is based on analysis of existing solutions in this field. This reference architecture includes four layers:

1. “Presentation Layer” provides user interface.
2. “Process Repository Management Layer” provides access management, version control etc.
3. “Database Management Layer” provides basic functions of database management system.
4. “Storage Layer” provides storage of business process models.

As for EA, according to TOGAF the Architecture Repository concept is tightly related to another architectural concept called Enterprise Continuum. This concept explains how certain generic solutions can be customized and used as per specific requirements of an organization. The Enterprise Continuum provides a view of Architecture Repository that provides ways and techniques for classifying architecture and other related artifacts as they transform from generic architecture to specific architectures that are suitable for specific needs of an organization. This interaction allows stakeholders to use all architectural resources and assets that are available in an organization-specific architecture. The Enterprise

Continuum provides a very good context for understanding various architectural models, the building blocks, and relationships between building blocks [5]. Scheme that represents the structure and relationships between the Architecture Repository and Enterprise Continuum according to the TOGAF framework is shown in Fig. 1 [7].

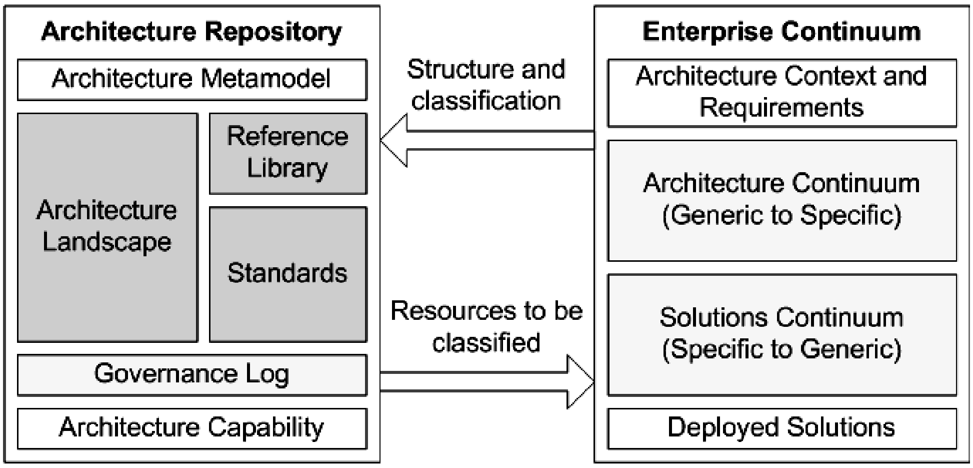
Architecture Repository
Architecture Metamodel
Architecture Landscape
Reference Library
Standards
Governance Log
Architecture Capability
Structure and classification
Resources to be classified
Enterprise Continuum
Architecture Context and Requirements
Architecture Continuum (Generic to Specific)
Solutions Continuum (Specific to Generic)
Deployed Solutions

Fig. 1. Structure and relationships between the Architecture Repository and Enterprise Continuum

The problem of retrieving similar business process models from the repository has been earlier considered in studies [1, 8, 9], which propose label similarity, structural similarity, and behavioral similarity measures based on labels comparison of business process models nodes. Authors of these papers discussed the foundations of detecting and measuring similarity between business process models described in BPMN (Business Process Modeling and Notation) and EPC (Event-driven Process Chain) notations. In the survey on business process similarity measures [10], Becker and Laue concluded that there is not a single “perfect” similarity measure. They also gave some recommendations for the selection of an appropriate similarity measure for different use cases.

Another interesting paper [11] considers not only similarity measures of business process models, but also provides an approach to similarity search in large business process model repositories. Proposed indexing approach is based on metric trees, a hierarchical search structure that saves comparison operations during search with nothing but a distance function at hand. Dijkman et al. have also mentioned the ideas of similarity search of process models, which are based on computationally inexpensive metrics, comparison of models’ fragments, and clustering techniques [12].

PROPOSED APPROACH

Earlier we have proposed using of the knowledge representation model called RDF (Resource Description Framework) to describe business process models that are used to represent organizational activities [13]. The RDF model is based on “subject-predicate-object” statements, which are convenient for machine processing [14]. A set of such statements might be represented as a marked directed graph.

Proposed RDF Schema used to describe a business process model as the RDF graph includes classes and properties (Fig. 2), such as:

1. “FlowObject” is the class that describes process flow objects, such as functions (“Function” class), processes (“Process” class), events (“Event” class), and gateways (“Gateway” class) related to each other using “isPredecessorOf” property. Classes “DataStore” and “ExternalEntity” derived from the class “Process” are used to provide description for business processes in DFD (Data Flow Diagram) notation.

2. “OrganizationalUnit” is the class that describes organizational units, such as departments (“Department” class) and positions (“Position” class) related to functions and processes using “isPerformedBy” property.

3. “ApplicationSystem” is the class that describes supporting IT-systems related to functions and processes using “isSupportedBy” property.

4. “BusinessObject” is the class that describes domain objects that might be considered as inputs (“requires” property), outputs (“produces” property), and regulations (“isRegulatedBy” property) of functions and processes.

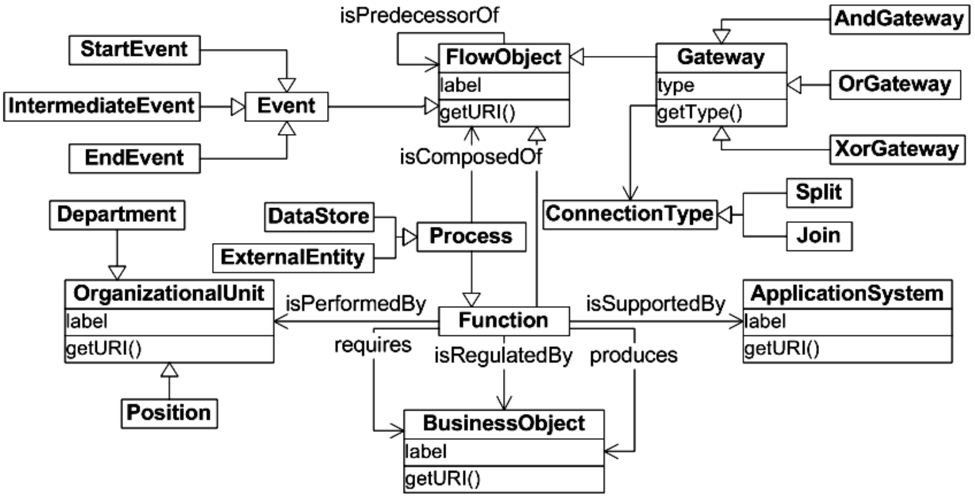

Fig. 2. Classes and properties of the proposed RDF Schema

Maintenance of the repository of business process models, represented as RDF graphs, might allow various possibilities, such as store and retrieve knowledge about organizational activities, and its further reuse to design new or improve existing business processes. Improvement of an existing organizational business process, according to BPM concept, assumes selection of its design variants and further transformation using obtained recommendations.

Therefore, business process models that are similar to an existing business process model should be retrieved from the Business Process Model Repository. Hence, the similarity measure of two business process models $BPModel_1$ and $BPModel_2$ represented using RDF graphs is proposed:

$$BPModelSim(BPModel_1, BPModel_2) = \alpha_1 \frac{1}{1+\left\| N_1 \right| - \left| N_2 \right\|} +$$

$$+\alpha_2 \frac{2}{|N_1| + |N_2|} \sum_{x \in Flow_1 \wedge x \in Flow_2} \frac{1}{1+\left| m_{Flow_1}(x) - m_{Flow_2}(x) \right|} +$$

$$+\alpha_3\frac{2}{|F_1|+|F_2|}\sum_{x\in Org_1\wedge x\in Org_2}\frac{1}{1+\left|m_{Org_1}(x)-m_{Org_2}(x)\right|}+$$

$$+\alpha_4\frac{2}{|F_1|+|F_2|}\sum_{x\in App_1\wedge x\in App_2}\frac{1}{1+\left|m_{App_1}(x)-m_{App_2}(x)\right|}+$$

$$+\alpha_5\frac{1}{|F_1|+|F_2|}\left(\sum_{x\in In_1\wedge x\in In_2}\frac{1}{1+\left|m_{In_1}(x)-m_{In_2}(x)\right|}+\right.$$

$$\left.+\sum_{x\in Out_1\wedge x\in Out_2}\frac{1}{1+\left|m_{Out_1}(x)-m_{Out_2}(x)\right|}\right)+$$

$$+\alpha_6\frac{2}{|F_1|+|F_2|}\sum_{x\in Reg_1\wedge x\in Reg_2}\frac{1}{1+\left|m_{Reg_1}(x)-m_{Reg_2}(x)\right|},$$

where α_i, $i=\overline{1,6}$ are the weights of structure similarity by size, control flow, organizational units, supporting IT-systems, regulation objects, input objects, and output objects respectively, $\alpha_1+\alpha_2+\alpha_3+\alpha_4+\alpha_5+\alpha_6=1$; N_i is the set of flow objects; $Flow_i$ is the multiset of tuples that contain in-degree and out-degree values for each flow object (event, function, connector, etc.); Org_i, App_i, Reg_i, In_i, and Out_i are the multisets of degree values for each function (or process) with considering only organizational units, IT-systems, regulation objects, input objects, and output objects respectively; $m_A(x)$ is the number of occurrences of the element x in a certain multiset A.

For example, considered multisets for a given business process model $BPModel_{\text{sample}}$ shown in Fig. 3 will be the following:

$$Flow_{\text{sample}}=\{(0,1),(1,1),(1,1),(1,1),(1,1),(1,2),(1,1),(1,1),(2,1),(1,0)\},$$

$$Org_{\text{sample}}=\{1,1,1,1\},\ App_{\text{sample}}=\{1,1,1,0\},$$

$$Reg_{\text{sample}}=\varnothing,\ In_{\text{sample}}=\varnothing,\ Out_{\text{sample}}=\varnothing.$$

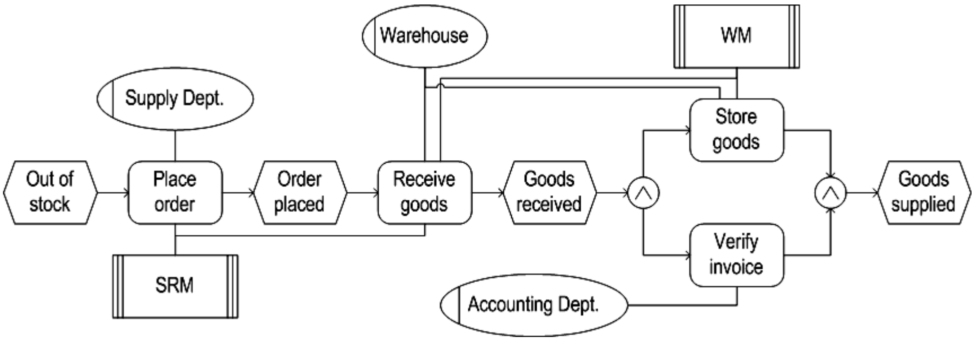

Fig. 3. Sample business process model

Whereas known similarity measures of business process models [1, 8, 9] contain difficulties of syntactic, semantic, and contextual label comparisons, proposed measure is based on process model graph structural characteristics. Moreover, proposed measure allows configuring similarity degree according to specific features of business process modeling notations. The corresponding values of weights α_i, $i = \overline{1,6}$ that depend on notations used to describe the compared business process models are shown in table 1. Those weights that are missing for a certain notation (e. g., column that corresponds to BPMN contains only α_1 and α_2 weights) should be considered as zeros.

Table 1. Weights of the similarity measure components according to various business process modeling notations

Notation	BPMN		DFD		eEPC					IDEF0				
	α_1	α_2	α_1	α_5	α_1	α_2	α_3	α_4	α_5	α_1	α_3	α_4	α_5	α_6
BPMN	0,5	0,5	1	0	0,5	0,5	0	0	0	1	0	0	0	0
DFD	1	0	0,5	0,5	0,5	0	0	0	0,5	0,5	0	0	0,5	0
eEPC	0,5	0,5	0,5	0,5	0,2	0,2	0,2	0,2	0,2	0,25	0,25	0,25	0,25	0
IDEF0	1	0	0,5	0,5	0,25	0	0,25	0,25	0,25	0,2	0,2	0,2	0,2	0,2

Proposed similarity measure of business process models is normalized $BPModelSim(x,y) \in [0,1]$, symmetric $BPModelSim(x,y) = BPModelSim(y,x)$, and reflexive $BPModelSim(x,x) = 1$, where x stands for an existing business process model and y is a model retrieved from the business process model repository.

Since the similarity measure of business process models is normalized, its values might be evaluated using the Harrington's desirability function and the corresponding scale [15]. Mapping categories of Harrington's scale to categories of architectural artifacts provided by the TOGAF Enterprise Continuum allows classifying business process models as they transform from generic models to organization-specific models in the following manner by applying the decision tree model shown in Fig. 4.

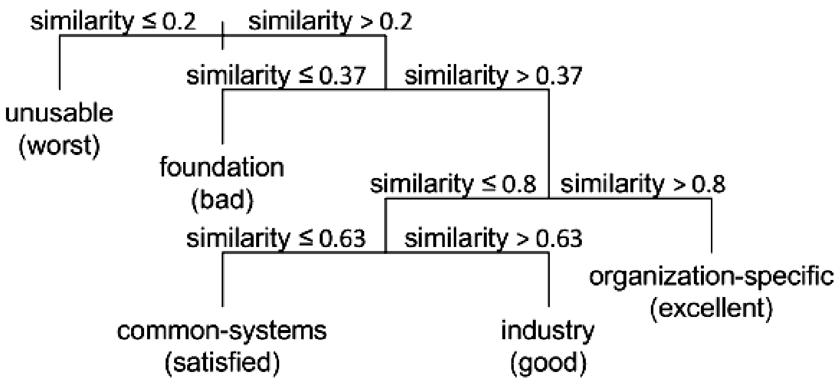

Fig. 4. Decision tree model for business process models classification

Proposed decision tree model was built using the machine learning algorithm CTree which is implemented in the R programming language and serves as the implementation of conditional inference trees method [16].

In the context of business process modeling, we assume artifacts as business process models provided using various notations and languages and then translated into the corresponding RDF graphs according to the proposed RDF Schema (Fig. 2). Therefore, proposed conceptual model of the Business Process Model Repository is shown in Fig. 5.

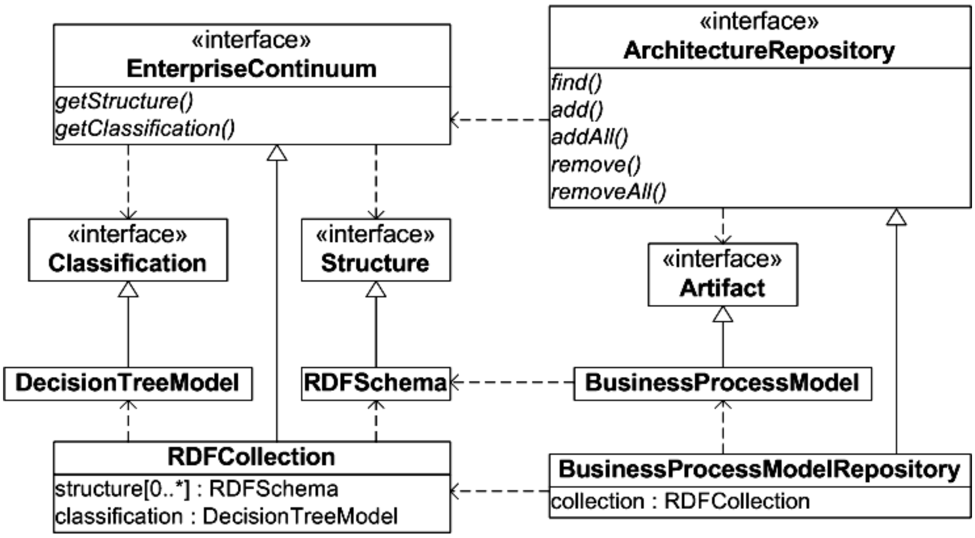

Fig. 5. Proposed conceptual model of the Business Process Model Repository

This model is based on the Repository Pattern that acts as a collection of artifacts [5]. As it is shown, the Business Process Model Repository might be considered as the concrete implementation of the Architecture Repository provided by TOGAF. According to this model, proposed similarity measure might be used to find process models in the repository that are similar to a given process model or a fragment thereof. Classification of a found business process model according to the Enterprise Continuum categories of artifacts (from foundation to organization-specific assets) is provided by the decision tree model (Fig. 4).

RESULTS

Proposed conceptual model of the Business Process Model Repository was used to implement the prototype of such tool using the Java-based open source library Apache Jena [17]. This library was used as the RDF triples storage and the framework to operate the set of business process models described using RDF graphs according to the RDF Schema shown in Fig. 2.

Since the version control of the Business Process Model Repository content is one of the basic requirements [3, 4], we used the distributed version control system Git in order to satisfy this requirement and provide maintenance of several versions of business process models [18]. The architecture of the Business Process Model Repository implementation is shown in Fig. 6.

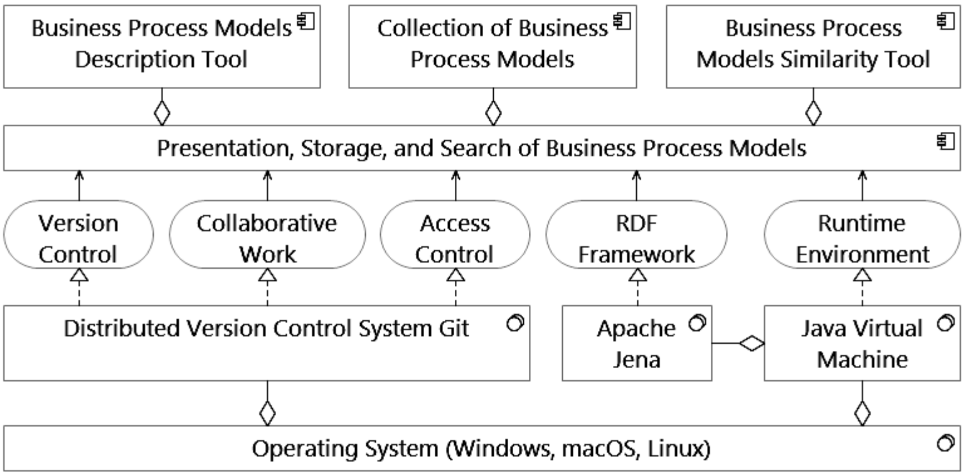
Business Process Models Description Tool
Collection of Business Process Models
Business Process Models Similarity Tool
Presentation, Storage, and Search of Business Process Models
Version Control
Collaborative Work
Access Control
RDF Framework
Runtime Environment
Distributed Version Control System Git
Apache Jena
Java Virtual Machine
Operating System (Windows, macOS, Linux)

Fig. 6. Architecture of the Business Process Model Repository implementation

The software prototype developed to implement the Business Process Model Repository was used to translate into RDF graphs and upload 73 business process models provided by the business diagramming software vendors Conceptdraw, EDraw, and MyDraw on their websites. Proposed similarity measure was used to find similar business process models among the considered 73 models (including 20 eEPC models, 25 BPMN models, and 28 DFD models). The Rand index [19] accuracy values were calculated in order to validate proposed similarity measure by comparing it with the similarities based on business process model metrics: size, density, and coefficient of network connectivity [20]. Moreover, the accuracy values were defined for each similarity threshold according to the proposed classification (see Fig. 5). These results are outlined in Table 2.

Table 2. Accuracy and estimation of similar business process models

Artifacts categories	Size	Density	Connectivity	Similar models
Foundation	0,51	0,5	0,48	55,49%
Common Systems	0,82	0,78	0,77	24,26%
Industry	0,98	0,91	0,92	7,19%
Organization-Specific	0,98	0,92	0,93	4,79%

Obtained results demonstrate decrease of the number of similar pairs of models and growth of accuracy during the transition from foundation to organization-specific categories according to the TOGAF Enterprise Continuum. The values shown in Table 2 demonstrate correctness of the proposed similarity measure since its accuracy grows according to the Enterprise Continuum categories – from generic to specific artifacts.

CONCLUSIONS

In this paper we have proposed the similarity measure between business process models. In contrast with already known measures based on labels comparison [1, 8, 9], it uses graph structural characteristics to define similarity of business process models described using various modeling notations and standards. Besides, proposed measure allows considering similarity not only by the process flow objects, but also by the organizational units, supporting IT-systems, and business objects. Proposed measure could be used to identify process models in the repository, which are similar to a given process model or a fragment thereof.

Proposed conceptual model of the Business Process Model Repository is based on the TOGAF framework in order to provide interoperability with the whole Architecture Repository and Enterprise Continuum. Being the concrete implementation of the Architecture Repository, proposed conceptual model of the Business Process Model Repository uses the RDF Schema (Fig. 2) and decision tree model (Fig. 4) to provide structure and classification for stored business process models according to the Enterprise Continuum concept.

Proposed similarity measure was used to estimate similarity of the sample business process models in order to analyze accuracy of this measure while going from the foundation to organization-specific categories of the stored artifacts according to the TOGAF Enterprise Continuum.

Future work includes additional consideration of the business process models similarity search technique that should be elaborated taking into account a large collection of business process models stored in the repository in which pairwise comparison of models is not feasible due to performance reasons.

REFERENCES

1. *Dumas M.* Similarity search of business process models / M. Dumas, L. Garcia-Banuelos, R.M. Dijkman // Bulletin of the IEEE Computer Society Technical Committee on Data Engineering. — 2009. — **32**. — P. 23–28.
2. *Yan Z.* Business process model repositories – Framework and survey / Z. Yan, R. Dijkman, P. Grefen // Information and software technology. — 2012. — **55**. — P. 380–395.
3. *Shahzad K.* Requirements for a business process model repository: A stakeholders' perspective / K. Shahzad, M. Elias, P. Johannesson // Business Information Systems. — 2010. — **47**. — P. 158–170.
4. *Elias M.* Design of business process model repositories: requirements, semantic annotation model and relationship meta-model / M. Elias. – Department of Computer and Systems Sciences, Stockholm University, 2015. — 252 p.
5. *Pethuru R.* Architectural Patterns / R. Pethuru, R. Anupama, H. Subramanian. — Packt Publishing, 2017. — 458 p.
6. *Yan Z.* A Framework for Business Process Model Repositories / Z. Yan, P. Grefen // International Conference on Business Process Management. — 2010. — **66**. — P. 559–570.
7. *Architecture* Repository. The TOGAF Standard, Version 9.2. — Available at: http://pubs.opengroup.org/architecture/togaf9-doc/arch/
8. *Dijkman R.* Similarity of business process models: Metrics and evaluation / R. Dijkman // Information Systems. — 2011. — **36**. — P. 496–516.
9. *Van Dongen B.* Measuring similarity between business process models / B. Van Dongen, R. Dijkman, J. Mendling // Seminal Contributions to Information Systems Engineering. — 2013. — P. 405–419.
10. *Becker M.* A comparative survey of business process similarity measures / M. Becker, R. Laue // Computers in Industry. — 2012. — **63**. — P. 148–167.

11. *Kunze M.* Metric Trees for Efficient Similarity Search in Large Process Model Repositories / M. Kunze, M. Weske // BPM 2010: Business Process Management Workshops. — 2010. — **66**. — P. 535–546.
12. *Dijkman R.* Managing large collections of business process models – Current techniques and challenges / R. Dijkman, M. La Rosa, H. A. Reijers // Computers in Industry. — 2012. — **63**. — P. 91–97.
13. *Kopp A.* An approach to business process models repository development / A. Kopp, D. Orlovskyi // Information Processing Systems. — 2018. — **153** (2). — P. 60–68.
14. *Resource* Description Framework (RDF). Semantic Web Standards. — Available at: https://www.w3.org/RDF/
15. *Kondruk N.* Clustering method based on fuzzy binary relation / N. Kondruk // Eastern-European Journal of Enterprise Technologies. — 2017. — **4** (2) — P. 10–16.
16. *Hothorn T.* ctree: Conditional Inference Trees / T. Hothorn, K. Hornik, A. Zeileis // The Comprehensive R Archive Network. — 2015. — Available at: https://rdrr.io/rforge/partykit/f/inst/doc/ctree.pdf
17. *Apache* Jena. Semantic Web Standards. — Available at: https://www.w3.org/2001/sw/wiki/Apache_Jena
18. *Chacon S.* Pro git / S. Chacon, B. Straub. — Apress, 2014. — 456 p.
19. *Sivogolovko E.* Validating cluster structures in Data Mining tasks / E. Sivogolovko, B. Novikov // Proceedings of the 2012 Joint EDBT/ICDT Workshops. — ACM, 2012. — P. 245–250.
20. *Sanchez-Conzalez L.* Quality assessment of business process models based on thresholds / Sánchez-González L. // OTM Confederated International Conferences “On the Move to Meaningful Internet Systems”. — 2010. — P. 78–95.

Received 31.10.2018

From the Editorial Board: the article corresponds completely to submitted manuscript.