Kopp A.M., **Orlovskyi D.L.**

National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine

An approach to measure similarity of business process models

Introduction. At higher levels of BPM (Business Process Management) maturity, organizations tend to accumulate considerable amounts of business process models [1]. Thus, business process model repositories may contain hundreds or even thousands models, represented using various modeling notations [2]. Since business process modeling technique is used by organizations to describe knowledge about their activities, the problem of store, share, and reuse of organizational knowledge, represented using business process models, becomes relevant.

In this paper a similarity measure, used to retrieve business process models from a repository in order to their further reuse in a business process continuous improvement cycle according to BPM concept, is proposed. The problem of similar business process models retrieving has been earlier considered in studies [1,3,4], which propose several measures, based on label similarity, structural similarity, behavioral similarity etc.

Details. Earlier we have proposed using of the knowledge representation model RDF (Resource Description Framework) to describe business process models, developed in organization. RDF is based on "subject-predicate-object" statements, which are convenient for machine processing [5]. A set of such statements may 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, such as:

1. $FlowObject$ – class, which includes process flow objects, such as functions ($Function$), processes ($Process$), events ($Event$), and gateways ($Gateway$), related using $triggers$ property.
2. $OrganizationalUnit$ – class, which includes organizational units, such as departments ($Department$) and roles ($Role$), related to functions and processes using $executes$ property.
3. $SupportingSystem$ – class, which includes supporting IT-systems, related to functions and processes using $usedBy$ property.

Therefore, maintenance of a repository of business process models, represented as RDF graphs, may 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.

Thus, business process models, 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:

$$
\begin{aligned}
BPModelSim&(BPModel_1, BPModel_2) = \\
= \alpha_1 \cdot &\left[\beta_{lab} \cdot Sim(U_1, U_2) + \beta_{str} \cdot Sim\left(\bigcup_{u \in U_1 \cap U_2} executes_1(u), \bigcup_{u \in U_1 \cap U_2} executes_2(u) \right) \right] + \\
+ \alpha_2 \cdot &\left[\beta_{lab} \cdot Sim(S_1, S_2) + \beta_{str} \cdot Sim\left(\bigcup_{s \in S_1 \cap S_2} usedBy_1(s), \bigcup_{s \in S_1 \cap S_2} usedBy_2(s) \right) \right] + \\
+ \alpha_3 \cdot &\left[\beta_{lab} \cdot Sim(F_1, F_2) + \beta_{str} \cdot Sim\left(\bigcup_{f \in F_1 \cap F_2} triggers_1(f), \bigcup_{f \in F_1 \cap F_2} triggers_2(f) \right) \right],
\end{aligned}
\tag{1}
$$

where $\alpha_i, i = \overline{1,3}$ – weights represent importance of organizational units, supporting IT-systems, and process flow objects similarity respectively, $\alpha_1 + \alpha_2 + \alpha_3 = 1$; β_{lab} and β_{str} – weights represent importance of label (unique resources used in both RDF graphs) and structural (unique statements used in both RDF graphs) similarity respectively, $\beta_{lab} + \beta_{str} = 1$; U – a set of organizational units;

S – a set of supporting IT-systems; F – a set of process flow objects; Sim – a similarity measure of two sets, e.g. Jaccard index J; $executes(u)$ – a set of processes, executed by the organizational unit $u \in U$; $usedBy(s)$ – a set of processes, supported by the IT-system $s \in S$; $triggers(f)$ – a set of process flow objects, triggered by the object $f \in F$.

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

Result. As an example, two models (fig. 1), which describe goods order business process in EPC (Event-driven Process Chain) notation, were translated to RDF graphs. Similarity of these models was calculated using the proposed measure (1).
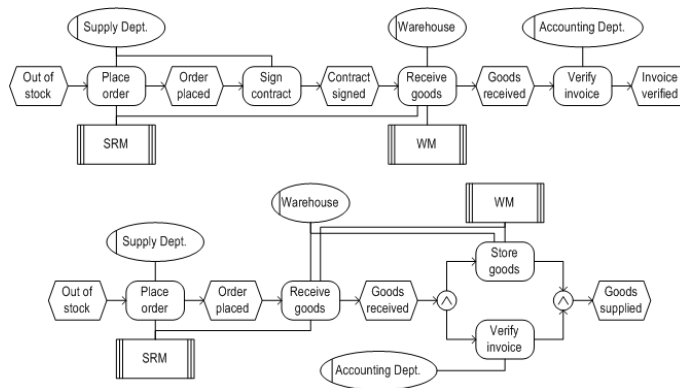

Figure 1. Sample pair of business process models (SRM – supplier relationship management system, WM – warehouse management system)

According to the obtained results, considered business process models are totally similar ($BPModelSim$ is 1) if only organizational units or supporting IT-systems are considered as important ($\alpha_1 = 1$ or $\alpha_2 = 1$) to compare only by the label closeness ($\beta_{lab} = 1$). By the other side, comparison of these models only by the structural closeness ($\beta_{str} = 1$) if only process flow objects are considered as important ($\alpha_3 = 1$), has shown that considered models are not that similar ($BPModelSim$ is 0,38). With considering that organizational units, supporting IT-systems, and process flow objects are equally important ($\alpha_1 = 0, 33$, $\alpha_2 = 0, 33$, and $\alpha_3 = 0, 34$) to compare both by the label and structural closeness ($\beta_{lab} = 0, 5$ and $\beta_{str} = 0, 5$), we have defined that these models are quite similar ($BPModelSim$ is 0,69).

Conclusion. In this paper we have proposed the similarity measure between business process models. In contrast with already known measures [1,3,4], it combines both label and structural approaches to define similarity of business process models, described using knowledge representation model RDF. Besides, proposed measure allows considering similarity not only by the process flow, but also by the whole process environment, including organizational units and supporting IT-systems.

References. **1.** M. Dumas, L. Garcia-Banuelos, and R.M. Dijkman, "Similarity search of business process models", Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, vol. 32, no. 3, pp. 23–28, 2009. **2.** Z. Yan, R. Dijkman, and P. Grefen, "Business process model repositories – Framework and survey", Information and Software Technology, vol. 55, pp. 380–395, 2012. **3.** R. Dijkman et al., "Similarity of business process models: Metrics and evaluation", Information Systems, vol. 36, no. 2, pp. 498–516, 2011. **4.** B. Van Dongen, R. Dijkman, and J. Mendling, "Measuring similarity between business process models" // Seminal Contributions to Information Systems Engineering, pp. 405–419, 2013. **5.** RDF. [Online]. Available: https://www.w3.org/RDF/.