Detection of Semantic Relations between Business Process Activity Labels

Lobna Makni, Nahla Zaaboub Haddar and Hanene Ben Abdallah
Mir@cl Laboratory, University of Sfax, Airport Road Km 4, Sfax, Tunisia

Keywords: Semantic Relation, Subsumed Activity, Part-of Relation Activity, Semantic Relation, Decision Rule.

Abstract: Determining similarities between business process models needs to be resolved in several cases, *e.g.,* when business processes must be merged, configured or integrated; when reusing fragments of process models, etc. Similarity can be tackled at the semantic, structural and/or behavioral levels. In this paper, we are interested in the semantic level similarity between business process models. More specifically, we present a method for detecting semantic relations between activity labels in process models belonging to the same business domain.

1 INTRODUCTION

Detecting the similarity between BPMs has recently gained interest in the business process management community (Awad et al., 2008). Similarity was addressed separately either at the semantic, structural or behavioral level of process models. The semantic level of a process model is the description of the meaning of its elements including its activity and event labels.

Comparing business processes at this level is a fundamental step for several model management activities. For example, in a merger scenario, it is necessary when integrating similar business process models into one process (La Rosa, 2010). In addition, this comparison is needed for managing process model repositories and ensuring the uniqueness of process models (Dijkman, 2011). Furthermore, multi-national enterprises need a similarity measurement to identify specialized processes of a national branch that no longer comply with the procedures defined in the company reference model (Van Dongen, 2008).

The aim of this paper is to use a linguistic comparison between activity labels to derive additional semantic relation types such as hyponyms and meronyms. A hyponym is an activity label whose semantics is included within another activity label, its hypernym. A hyponym shares a *type-of* semantic relation with its hypernym. On the other hand, a meronym activity label denotes a constituent part of another activity label. The detection of such relations between activity labels helps to identify *relations* between process *fragments*. Indeed, a process fragment is made up of activities that have semantic relations and control dependency among them. Identifying common fragments between business processes avoids redundancies when merging them. It is also necessary for establishing relations between semantically close process fragments.

In addition to the new semantic relations, the second contribution of this paper is a relation detection method that finds semantic relations among activity labels of process models belonging to the same business domain. This method can be used to extract process fragments that are semantically close and frequently present in the analyzed process models.

In the remainder of this paper, we first give an overall view of our relation detection method. Then, we present its set of decision rules to detect semantic relations among activity labels of process models belonging to the same business domain. Finally, we place the presented work in the context of already proposed approaches.

2 SEMANTIC RELATION DETECTION METHOD

A *semantic relation* between activities appears when

Makni L., Zaaboub Haddar N. and Ben Abdallah H..
Detection of Semantic Relations between Business Process Activity Labels.
DOI: 10.5220/0004003102730277
In *Proceedings of the 14th International Conference on Enterprise Information Systems* (ICEIS-2012), pages 273-277
ISBN: 978-989-8565-12-9

activities use different labels among which there is a semantic relation of equivalence, synonymy, subsumption or part-of. An activity *A* in one process *subsumes* an activity *B* in another process if *A* represents the same unit of work as *B*, but includes other units of work as well (Dijkman et al., 2007). On the other hand, an activity *A* in one process *partly corresponds* to an activity *B* in another process, if *A* and *B* partly represent equivalent units of work, but both represent also different units of work.

The semantic relation detection method we propose is composed of two main steps (Figure 1): *label refactoring,* and *definition of semantic links between labels*.

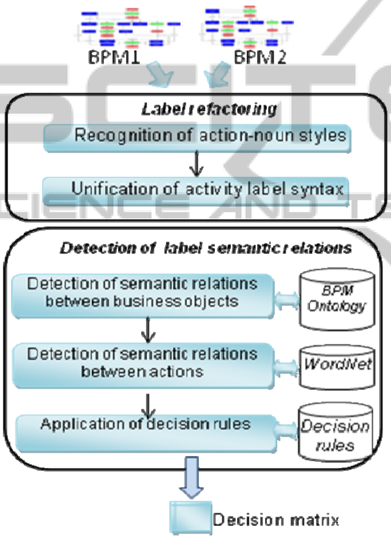

Figure 1: Semantic relation detection method.

2.1 Label Refactoring

An activity label captures an action and a business object on which the action is performed. To compare two activity labels, it is preferable that they have the same syntactic pattern, *e.g.,* an action followed by a business object. Unfortunatelly, designers use different styles when labeling activities. Thus, before comparing activity labels, a preliminary step of label *refactoring* is needed: This step aims at harmonizing the writing styles of the activity labels into the following form:

$$\underbrace{\text{action}}_{\text{Action field}} + \underbrace{\text{Business-Object}}_{\text{BO field}}$$

According to (Leopold and al., 2010), in practice, label writing styles can be classified into three syntactic patterns: verb-object labels, action-noun labels, and rarely other irregular forms. (Leopold and al., 2010) developed an algorithm to detect the syntactic pattern of an activity label in a BPM. In addition, this work proposes a derivation algorithm to rewrite labels from the action-noun pattern to the verb-object pattern. To refactor the actvity labels into the above action-business object style, we use this derivation algorithm to extract from each label the action and the associated business object. For example, the label 'Creation of invoice' is an action-noun pattern; after the derivation to the verb-object label, we get the action 'create' and the object 'invoice'. Thus, after the label refactoring step, each activity label can be represented as a couple *(a, O)*, where *a* is an action and *O* is a business object. The third irregular syntactic form is rarely found in business process models (1.1%). It concerns frozen forms which can be added as they are to the ontology used for the detection of semantic relations.

2.2 Label Semantic Relations Definition

To identify semantic relations between activity labels, we proceed in three steps (Figure 1). First, we compare the business objects specified in the labels' object fields; we believe that activities that do not act on the same object cannot be compared. For this *business* objects comparison, we need a business domain ontology. In our study, we used the *eTVSM* ontology (Awad, 2008) which offers a reference terminology that defines semantic correspondences between business objects. For example, using *eTVSM,* we can detect that the business object 'customer order' is *a kind of* business object 'order'. If two activity labels use the same (or synonym) business objects, we continue with the detection of the semantic relation between their actions. Since we are looking for relations among verbs in general, in this second step, a general-purpose dictionary is sufficient. In our study, we used WordNet to determine semantic relations between actions (*i.e.,* the verbs in the labels).

As illustrated in Figure 1, once the semantic relations between the business objects and the actions are identified, our semantic relation detection method applies a set of decision rules that we describe next.

3 DECISION RULES

To detect semantic relations, we need to construct,

for each *pair* of process models belonging to the same business domain, a matrix to compare the activity labels. In a comparison matrix, lines and columns correspond to activity labels of the compared process models. Each element in the comparison matrix represents the semantic relation (*synonym, subsumption, part-of, different*) according to the decision rules defined below in this section. To detect the semantic relations between activities, our decision rules are based first of all on the comparison of business objects and then continue with the comparison of their associated actions.
Given two activities, there are four types of binary semantic relations between them: identity, synonymy, subsumption and part-of. These relations are determined according to the following rules:

Rule 1 (synonym activities): Let A_1 = (*a1, O1*) and A_2 = (*a2, O2*) be two activities of two process models M_1 and M_2 respectively. A_1 and A_2 are *synonyms*, noted $A_1 \equiv A_2$, if and only if $O_1 \equiv O_2$ and $a_1 \equiv a_2$.

Rule 2 (subsumed activities): Let A_1 = (*a1, O1*) and A_2 = (*a2, O2*) be two activities of two process models M_1 and M_2 respectively. A_1 *subsumes* A_2, noted $A_1 \uparrow A_2$, if and only if $O_1 \equiv O_2$ and a_1 is an hypernym of a_2 ($a_1 \uparrow a_2$).

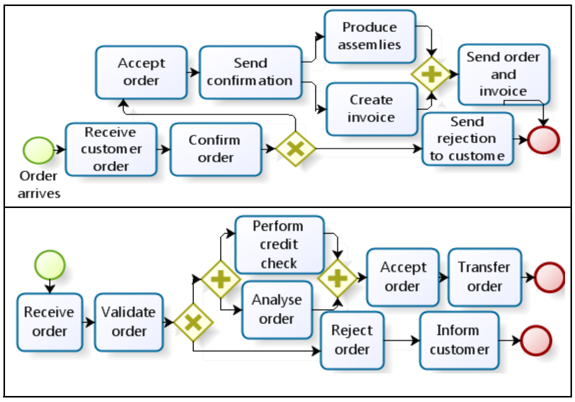

Figure 2: Two ordering processes.

Rule 3 (part-of activities): Let A_1 = (*a1, O1*) and A_2 = (*a2, O2*) be two activities of two process models M_1 and M_2 respectively. A_1 is *part-of* A_2, noted A_1 -♦ A_2, if O_1 is a part-of O_2 and $a_1 \equiv a_2$.

Rule 4 (different activities): Let A_1 = (*a1, O1*) and A_2 = (*a2, O2*) be two activities of two process models M_1 and M_2 respectively. A_1 and A_2 are *different*, noted A_1 Ð A_2, if and only if O_1 Ð O_2.

To illustrate our approach, we consider the two models in Figure 2 describing an ordering process. By applying the above defined decision rules, we obtain the comparison/decision matrix shown in Table 1. For example, the semantic relation between 'confirm order' and 'validate order' is subsumption. Indeed, after decomposing these activity labels into the verb-object grammatical structure, we obtain 'confirm' and 'validate' as verbs in the action field and 'order' as noun in the business object field. Using the *eTVSM* ontology, we can detect that these activity labels have the same object 'order'. Then, using WordNet, we find that the action 'validate' subsumes 'confirm'. Therefore, the activity label 'validate order' subsumes the activity label 'confirm order'.

4 RELATED WORKS

In the business process management domain, there are many issues and difficulties related to terminology mismatches and the unstructured and isolated knowledge representation in process models. To tackle these weaknesses, several approaches were proposed in the literature. They are based on solutions using domain *ontology* construction, *aggregation* techniques or *semantic* similarity metrics.

Two kinds of ontologies are used to enable the semantic support of modeling BP activities. The *sBPMN* ontology (semantic Business Process Modeling Notation) represents BPMN process models, featuring basic concepts and attributes for

Table 1: Decision matrix for BPM 1 and BPM 2.

A_1 / A_2	(Receive, order)	(Validate, order)	(perform, credit check)	(Analyze, order)	(Accept, order)	(Transfer, order)	(reject, order)	(Inform, customer)
(Receive, customer order)	$A_1 \equiv A_2$	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2
(Confirm, order)	A_1 Ð A_2	$A_1 \uparrow A_2$	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2
(Accept, order)	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	$A_1 \equiv A_2$	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2
(Send, confirmation)	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2
(Produce, assemblies)	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2
(Create, invoice)	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2
(Send, order and invoice)	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 -♦ A_2	A_1 Ð A_2	A_1 Ð A_2
(Send, rejection)	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2	A_1 Ð A_2

standard BPMN elements (Abramowicz et al., 2007). In (Born et al., 2007), the authors extended sBPMN to provide for the definition of the states of a business object before and after the execution of corresponding activities and to link objects, states and activities to elements of domain ontologies describing them. With these extensions, the *sBPMN* ontology can be used as an internal and external format for semantically augmented BPMN process models. The domain ontology covers information concerning domain objects and states which help to model business processes more precisely.

On the other hand, it is common for large organizations to maintain repositories of BPMs in order to document and to improve their operations. To retrieve process models from such BPM repository, a comparison means is required (Dijkman, 2011) (Ehrig, 2007) (Van Dongen, 2008). Based on label similarity, (Dijkman et al., 2011) propose a label matching similarity metric. The metric definition depends on the syntactic or semantic similarity notions or a weighted average of them (Dijkman et al., 2011). In addition, (Ehrig et al., 2007) also proposed a combined metric that computes similarity degrees between a pair of process element names based on syntactic, linguistic and structural measures. In (Dijkman et al., 2011) and (Ehrig et al., 2007), the authors use the WordNet dictionary to detect synonymous words.

Furthermore, in the context of company mergers, teams of analysts need to compare similar process models to identify commonalities and differences, and to create a configurable process model that captures a family of process models in an integrated manner (La Rosa et al., 2010). (La Rosa et al., 2010) used a matching score of a mapping between two functions or events based on the similarity between their labels. The matching score depends on syntactic and linguistic similarity measures. In (Dijkman et al., 2011) and (Makni et al., 2011), the authors use the same mapping functions to calculate the similarity between activity labels based on synonym words.

A significant point in the design of the aggregation operation is activity aggregation. Existing BPM abstraction techniques from the semantics of activities in business process models. In (Smirnov et al., 2010), the authors developed an aggregation technique clustering activities according to their domain semantics. The technique can guide the user during a process model abstraction providing recommendations on related activities. Aggragation of actions requires them to be related by a part-of or meronym relation. This work proposed a metric for comparing activity aggregations and the algorithm for aggregation mining (Smirnov et al., 2010). The metric is applied on a meronymy forest represented by the MIT Process Handbook (Malone et al., 2003). This latter describes business processes obtained by researchers with the help of business process experts. It represents several business domains such as sales, distribution, and production. The handbook illustrates about 5000 activities with their semantic relations like hyponymy and meronymy.

5 CONCLUSIONS

The main contributions of this paper are to propose decision rules to detect semantic relations between activity labels and a semantic relations detection method. The proposed method determines semantic relations between activity labels such as subsumption and part-of relation.

We are currently automating the presented method in order to evaluate its advantages and limits. In addition, we will validate our relations detection method by an empirical study on process models to determine its precision rate.

REFERENCES

Abramowicz W., Filipowska, A., Kaczmarek, M., and Kaczmarek, T., 2007. Semantically enhanced business process modelling notation: *Workshop on Semantic Business Process and Product Lifecycle Management* (SBPM 2007).

Awad A., Polyvyanyy, A., Weske M., 2008. Semantic Querying of BPM: EDOC 2008, - pp. 85-94.

Born M., Dörr, F., Weber, I., 2007. User-friendly Semantic Annotation in Business Process Modeling: *WISE Workshops*, pp-260-271

Dijkman R., Dumas, M., Van Dongen, B., Käärik,R., Mendling, J., 2011. Similarity of BPMs: *Metrics and Evaluation: Information System*, - Vol. 36(2): pp. 498-516 .

Ehrig M., Koschmider, A., Oberweis, A., 2007. Measuring Similarity between Semantic BPMs: *APCCM 2007* - pp. 71-80.

La Rosa M., Dumas, M., Uba, R., Dijkman, R. M., 2010. Merging BPMs: *In Proc. of OTM* - Vol. 6426: pp. 96–113.

Leopold H., Smirnov, S., Mendling, J., 2010. Refactoring of Process Model Activity Labels: *NLDB'10*- pp. 268-276.

Makni L., Zaaboub, N., Ben Abdallah, H. Semantic Design Patterns for Business Processes: *the 6th International Conference on Software and Data Technologies* (ICSOFT'11), 2011.

Malone T. W., Crowston, K., Herman, G. A., 2003. Organizing Business Knowledge: The MIT Process Handbook: *The MIT Press*, Cambridge.

Smirnov S., Dijkman, R., Mendling, J., Weske M. , 2010. Meronymy-Based Aggregation of Activities in BPMs: ER - pp. 1-14.
Van Dongen, B., Dijkman, R. M., Mendling, J, 2008. Measuring Similarity between Business Process Models: *In CAiSE*, Vol. 5074 Springer, pp. 450-464.