Refactoring of Process Models Activity Labels

Henrik Leopold[1], Sergey Smirnov[2], and Jan Mendling[1]

[1] Humboldt-Universität zu Berlin, Germany
{henrik.leopold,jan.mendling}@wiwi.hu-berlin.de
[2] Hasso Plattner Institute, Potsdam, Germany
sergey.smirnov@hpi.uni-potsdam.de

Abstract. Recently many companies have expanded their business process modeling projects such that thousands of process models are designed and maintained. Activity labels of these models are related to different styles according to their grammatical structure. There are several guidelines that suggest using a verb-object labeling style. Meanwhile, real-world process models often include labels that do not follow this style. In this paper we investigate the potential to improve the label quality automatically. We define and implement an approach for automatic refactoring of labels following action-noun style into verb-object labels. We evaluate the proposed techniques using a collection of real-world process models—the SAP Reference Model.

1 Introduction

Business process modeling is an integral part of process management in large enterprises. Many of the companies design and maintain up to several thousands of models that capture their operations [17]. Nowadays, laymen and casual modelers design a great share of process models with detrimental consequences for model quality. In this context, there is an ever growing demand for automatic techniques to check and improve the quality of process models.

Process model quality has been approached from different angles, including verification, error probability, and comprehension [1, 13, 14]. Also the small pieces of text that capture the names of activities (activity labels) have been investigated from a usability perspective. Such activity labels represent actions, which take place during the execution of a business process. Typically, an activity label captures an action and a business object, on which the action is performed, like *Validate address* or *Creation of order*. In essence, three classes of activity labels have been found in practice: verb-object labels, action-noun labels, and a rest category [15]. An interesting point is that verb-object labels are superior to action-noun labels in terms of perceived ambiguity. Therefore, it is desirable that all labels follow the verb-object style.

In this paper we address the problem of automatic refactoring of action-noun labels to verb-object labels. A key challenge is the identification of actions and business objects in action-noun labels. As activity labels in process models are only sentence fragments, we use contextual information from neighboring

control flow elements instead of standard natural language parsing techniques. The approach has been implemented and evaluated using the SAP Reference Model, a large collection of real-world business process models [10]. The results emphasize the potential of our approach.

The remainder of the paper is structured as follows. Section 2 illustrates the research problem and presents the main styles we identified. Section 3 defines algorithms for label substyle recognition, action and business object derivation, and refactoring methods. In Section 4 we evaluate the presented algorithms. Related work is discussed in Section 5. Section 6 concludes the paper.

2 Background

In this section we outline the problem of label refactoring, starting with an example covering different labeling styles. Then, we present the results from a study of action-noun labels in the SAP Reference Model.

2.1 Motivation

The problem of activity label quality can be motivated by the business process fragment in Fig. 1. It captures a part of a profit center planning process. One can see, that it is easy to misinterpret activity label *Plan data transfer to EC-PCA from profitability analysis*. Ignoring the preceding and succeeding events, a reader might conclude that the label *Plan data transfer to EC-PCA from profitability analysis* instructs to *plan* a *data transfer*, and label *Plan integration of profit centers* advises to *plan* the *integration of profit centers*. However, event *Plan Data transferred from other Applications* reveals that the action in the activity on the left branch is given by noun *transfer*. Consequently, the activity label does not instruct to *plan* a *data transfer*, but to *transfer plan data*. This ambiguity partially stems from the style of labeling: the first word is a verb referring to an action, while in other cases the first word is a business object and the action is given as a noun.

Fig. 1. Poor labeling in the SAP Reference Model

A consistent application of the verb-object style increases understandability of labels [15]. Verb-object labels are verb phrases headed by an infinitive and succeeded by a noun phrase. The verb captures an action, while the noun phrase—a business object. An action-noun label states the action as a noun, which can often be confused with a business object. We propose to refactor labels of potential ambiguity by analyzing and transforming action-noun labels.
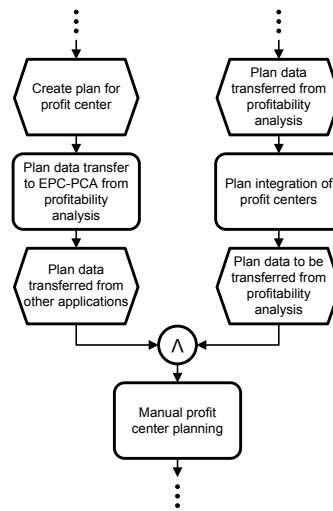
Table 1. Substyles of action-noun activity labeling style

Name	Structure	Example	Share, %
Noun phrase	NP → NN (bo), NN (a)	Invoice creation	78.8
Noun phrase with *of* prepositional phrase	NP → NP [NN (a)], PP [IN ('of'), NP [NN (bo)]]	Creation of invoice	15.0
Verb phrase (gerund)	VP → VBG (a), NP [NN (bo)]	Creating invoice	5.1
Irregular	-	LIFO: Invoice: Creation level	1.1

2.2 Label Classification

Development of effective algorithms deriving actions and business objects from activity labels requires a thorough understanding of current labeling practices. We approached this problem in a bottom-up way by investigating the different action-noun labels of the SAP Reference Model. This model collection includes 604 Event-driven Process Chains (EPCs), each containing several activities. Table 1 shows four identified substyles of action-noun labels, each described according to their structure and illustrated by an example.

The labels of *noun phrase* style are noun phrases where a business object comes first and an action follows. A business object may be absent. The labels of *Noun phrase with of prepositional phrase* style are also noun phrases. However, an action is represented by a leading noun and is succeeded by a prepositional phrase. The prepositional phrase is headed by a preposition *of* and refers to a business object. The labels of noun phrase with *of* prepositional phrase style can have an optional prepositional phrase. The labels of *verb phrase (gerund)* substyle are verb phrases headed by a gerund. The action is captured as a gerund, succeeded by the business object captured as a noun. This style may have an optional prepositional phrase.

The substyles described above cover almost 99% of action-noun labels in the model collection. The remaining labels cannot be assigned to any of these substyles and are related to *irregular* style. The specific property of these labels is the use of characters, linking together parts of the label in a special way. These characters do not allow labels to qualify into any of the above named substyles. Examples are *Transfer Posting FI-LC* and *LIFO: Valuation: Pool Level.* The

Table 2. Properties of action-noun label substyles

Label class	Property
Noun phrase	none
Noun phrase with *of* prepositional phrase	label contains a prepositional phrase with *of* as a leading preposition
Verb phrase (gerund)	the leading gerund signifies an action and is followed by a business object
Phrase with coordinating conjunction	the phrase contains a coordinating conjunction, e.g., *and* or *or*
Irregular	label contains characters ':' or ' - '

majority of irregular labels can be recognized by the use of the characters ':' and '–'.

Some labels refer to more than one business object or instruct to perform more than one action. Such labels contain a conjunction, coordinating the relations between homogeneous parts. Examples of conjunctions are *and*, *or*, comma symbol, and slash symbol. These labels can be decomposed to several labels, each capturing one action on one object.

3   Automatic Refactoring of Action-noun Activity Labels

This section presents a stepwise approach to automatic refactoring of activity labels from action-noun style into verb-object style. The refactoring process includes (1) label style recognition, (2) derivation of an action and a business object from the label, and (3) composing a verb-noun label.

3.1   Label Style Recognition

The recognition process is driven by a set of label properties. Each label is evaluated against the set of properties and, according to evaluation results, is categorized into a particular style. Table 2 enumerates the action-noun label substyles and points to their featured properties.

Algorithm 1 formalizes label style recognition. The input of the algorithm is an action-noun label *label*, the output is *prop*—an object storing the label properties with a substyle among them. We assume that all the boolean properties in *prop* object are initialized with *false*.

First the algorithm examines, if the label contains characters that allow to classify the label as irregular (see lines 3–5). If the label contains such characters, the style of the label is irregular and the algorithm terminates. Otherwise, the algorithm continues seeking for prepositions (lines 6–8) and conjunctions (lines 9–11). If conjunctions or prepositions are found, respective flags

---

Algorithm 1 Recognition of action-noun substyles

1: **recognizeSubstyle(Label** *label***)**
2: *prop* = new LabelProperties();
3: **if** *label* contains ':' OR ' - ' **then**
4:    *prop.style* = UNCLASSIFIED;
5:    **return** *properties*;
6: **if** *label* contains prepositions **then**
7:    *prop.hasPrepositions* = **true**;
8:    *prop.pIndex* = getFirstPrepositionIndex(*label*);
9: **if** *label* contains conjunctions **then**
10:    *prop.hasConjunctions* = **true**;
11:    *prop.cIndex* = getConjunctionIndex(*label*);
12: **if** first word in *label* has suffix 'ing' **then**
13:    *prop.hasSuffixING* = **true**;
14: *verbSize* = getVerbSize(*label*);
15: **if** *prop.hasSuffixING* **and** *label.size* > *verbSize* **and** (**!***prop.hasPrepositions*
     or *prop.pIndex* > *verbSize* + 1) **then**
16:    **if** verb == action derived from label context **then**
17:        *prop.style* = GERUND;
18:        **return** *prop*;
19: **if** *prop.hasPrepositions* **and** *label*.getWordAt(*prop.pIndex*) == 'of' **then**
20:    *prop.style* = PREPOSITION_OF;
21:    **return** *prop*;
22: *prop.style* = NOUN;
23: **return** *prop*;

---

hasConjunctions and *hasPrepositions* are set to *true* and the position of the first conjunction/preposition is stored in *pIndex/cIndex*.

The algorithm proceeds checking, if the label starts with a gerund (lines 12–18). It is verified, if the first word of the label has an *ing* suffix. Next, WordNet [16] is used to learn, if the first word is a verb and which infinitive it has. An assessment, whether the gerund represents an action, requires a deeper investigation: if the first word of a label is a gerund, it does not imply that this word also represents the action. Consider label *Planning scenario processing*. Although *planning* is a gerund, it might also be a part of a business object. In order to resolve this ambiguity, we consider event nodes preceding and succeeding the activity with the inspected label. Returning to the example, we notice that the activity is preceded by an event labeled with *Planning scenario is processed*. A part of speech analysis of this label identifies *planning* and *scenario* as nouns and *process* as a verb. Hence, we can infer that *processing* captures an action.

If the algorithm qualifies a label to be a gerund, it terminates. In the opposite case, the algorithm proceeds checking prepositions in the label (lines 19–21). A label containing prepositions, with the first *of* preposition, is qualified as a *noun phrase with of prepositional phrase*. If the label is categorized to none of the enumerated substyles, the algorithm relates it to a *noun phrase* substyle.

3.2   Derivation of Action and Business Object

Recognition of a label substyle enables the next step—derivation of an action and a business object. We provide an informal discussion of the algorithm enabling this step. The input of the algorithm is an action-noun label *label* and a corresponding *LabelProperties* object *prop* storing the properties of *label* obtained with Algorithm 1. The output of the algorithm is *prop* with *action* and *bObject* properties set.

The algorithm starts with an analysis of labels following noun phrase style. It checks for an optional prepositional phrase. If the label has a prepositional phrase, the phrase is omitted and not studied any more. If the label has only one word, e.g., *Deployment* or *Classification*, this word is recognized as an action. Otherwise, the algorithm checks, if the last two words of the label constitute a phrasal verb, for instance, *set up* and *carry forward*. If the first two words are recognized as a phrasal verb, this verb is perceived as an action. The rest of the label, if it exists, is recognized as a business object. If the phrasal verb is not revealed, the last word is recognized as an action, while the rest as a business object. It continues with analysis of *verb phrase (gerund)* labels. Analysis of these labels resembles the analysis of labels of noun phrase style. The key difference is that the action is expected to appear in the beginning of the label, while the business object—in the end. The algorithm concludes with an analysis of activity labels following *noun phrase with prepositional phrase* style. The label part preceding preposition *of* is recognized as an action. The label part between preposition *of* and the next preposition is treated as a business object.

3.3   Label Refactoring

Refactoring aims to transform an action-noun label into a verb-object label signifying the same action performed on the same business object. Derivation of actions and business objects from activity labels enables construction of labels in verb-object style. In fact, after the analysis of the previous steps the task becomes a trivial concatenation of a verb representing an action and a noun phrase representing a business object.

4   Empirical Evaluation

We have conducted an experiment to validate, how well the proposed algorithms approximate a human interpretation of activity labels. To evaluate the algorithms we have designed a test collection that includes the SAP Reference Model as a process model collection, and human interpretations of activity labels. Human interpretations are captured by two mappings: one mapping from an activity label to a set of corresponding actions and another mapping from an activity to a set of business objects. Within the evaluation we compared (1) recognition of label substyles by the algorithm and by humans and (2) derivation of actions and business objects by the algorithm and by humans.
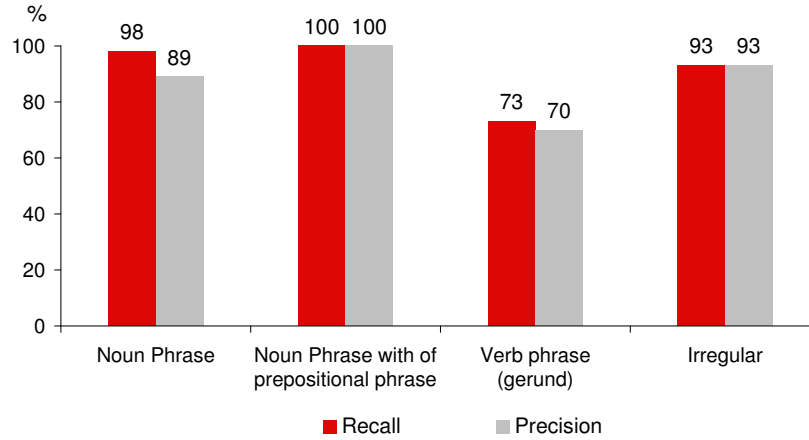
Fig. 2. Precision and recall of algorithms for label style recognition

To evaluate the substyle recognition algorithm we measured the *precision* (number of correctly recognized labels divided by retrieved labels) and *recall* (number of correctly recognized labels divided by all labels) of the algorithm [2]. Fig. 2 presents the corresponding values. The evaluation of action/business object derivation makes use of precision. The precision value for action derivation is 88%, for business object derivation 85%, and for label refactoring 85%. As the precision values are reasonably high, we conclude that the proposed algorithms are capable of automatic derivation of actions and business objects from labels.

5   Related Work

The research reported in this paper relates to guidelines for process model labeling and natural language approaches for conceptual models. The verb-object style is widely promoted in the literature for *labeling activities of process models* [12]. It has been observed though that verb-object labeling in real process models is not consistently applied. For instance, the practical guide for process modeling with ARIS [4, pp.66-70] shows models with both actions as verbs and as nouns. Our work helps to automatically refactor such action-noun labels. It also complements prior work on the automatic identification of verb-object labels in [11]. The concept of part of speech tagging is also investigated for interactive process modeling support [3]. The relationship between *process models and natural language* has been discussed and utilized in various works. In [6] the authors investigate, how far the three steps of building a conceptual model can be automated using a model for pre-design. Further text analysis approaches have been used to link activities in process models to document fragments [9] and to compare process models from a semantic perspective [5]. Most beneficiary is the verb-object style for model verbalization (see [7]), as verb-object style labels can easily be verbalized using the *You have to* prefix.

6   Conclusions and Future Work

In this paper we have proposed a method for automatic refactoring of action-noun activity labels into labels of verb-object style. We performed an evaluation of the proposed approach with process models from the SAP Reference Model. There are two directions of the future work. On the one hand, it is the improvement of the existing refactoring technique and the development of alternative methods based on natural language processing tools, e.g., [8]. As a next step, we aim to evaluate the proposed refactoring technique against other test collections. On the other hand, a number of applications calls for algorithms deriving actions and business objects from activity labels. In particular, we want to integrate the derivation technique with automatic modeling support of [18].

References

1. W.M.P. van der Aalst. Workflow Verification: Finding Control-Flow Errors Using Petri-Net-Based Techniques, LNCS 1806, pages 161–183. Springer Verlag, 2000.
2. R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. May 1999.
3. J. Becker, P. Delfmann, S. Herwig, L. Lis, and A. Stein. Towards Increased Comparability of Conceptual Models - Enforcing Naming Conventions through Domain Thesauri and Linguistic Grammars. In *Proceedings of ECIS*, June 2009.
4. R. Davis. *Business Process Modelling with ARIS: A Practical Guide*. 2001.
5. M. Ehrig, A. Koschmider, and A. Oberweis. Measuring Similarity between Semantic Business Process Models. In *Proceedings of APCCM*, pages 71–80, 2007.
6. G. Fliedl, C. Kop, and H.C. Mayr. From Textual Scenarios to a Conceptual Schema. *Data & Knowl. Eng.*, 55(1):20–37, 2005.
7. P. Frederiks and T. van der Weide. Information Modeling: The Process and the Required Competencies of Its Participants. *Data & Knowl. Eng.*, 58(1):4–20, 2006.
8. Stanford Natural Language Processing Group. Stanford Log-linear Part-of-speech Tagger. http://nlp.stanford.edu/software/tagger.shtml. Accessed 8.04.2010.
9. J.E. Ingvaldsen, J. A. Gulla, X. Su, and H. Rønneberg. A Text Mining Approach to Integrating Business Process Models and Governing Documents. In *OTM*, volume 3762 of *LNCS*, pages 473–484. Springer, 2005.
10. G. Keller and T. Teufel. *SAP(R) R/3 Process Oriented Implementation: Iterative Process Prototyping*. Addison-Wesley, 1998.
11. H. Leopold, S. Smirnov, and J. Mendling. On Labeling Quality in Business Process Models. In *EPK*, volume 554 of *CEUR*, pages 42–57, 2009.
12. Th. W. Malone, K. Crowston, and G. A. Herman. *Organizing Business Knowledge: The MIT Process Handbook*. The MIT Press, 2003.
13. J. Mendling. *Metrics for Process Models: Empirical Foundations of Verification, Error Prediction, and Guidelines for Correctness*, volume 6 of *LNBIP*. 2008.
14. J. Mendling, H. A. Reijers, and J. Cardoso. What Makes Process Models Understandable? *Business Process Management*, pages 48–63, 2007.
15. J. Mendling, H. A. Reijers, and J. Recker. Activity Labeling in Process Modeling: Empirical Insights and Recommendations. *Information Systems* 2009.
16. G.Miller. WordNet: a lexical database for English. *CACM*, 38(11):39–41, 1995.
17. M. Rosemann. Potential Pitfalls of Process Modeling: Part A. *Business Process Management Journal*, 12(2):249–254, 2006.
18. S. Smirnov, M. Weidlich, J. Mendling, and M. Weske. Action Patterns in Business Process Models. In *ICSOC*, Stockholm, November 2009.