On Labeling Quality in Business Process Models

Henrik Leopold†, Sergey Smirnov ‡, Jan Mendling†
†Humboldt-Universität zu Berlin, Germany
henrik.leopold@student.hu-berlin.de|jan.mendling@wiwi.hu-berlin.de
‡Hasso Plattner Institute, Potsdam, Germany
sergey.smirnov@hpi.uni-potsdam.de

Abstract: Quality assurance is a serious issue for large-scale process modeling initiatives. While formal control flow analysis has been extensively studied in prior research, there is a little work on how the textual content of a process model and its activity labels can be systematically analyzed. It is a major challenge to classify labels according to their quality and consequently assure high label quality. As a starting point we take a recent research on the activity labeling style, which establishes superiority of a so-called verb-object labeling style. Together with the labeling style, the length of an activity label is related to its quality.

In this paper, we investigate how various natural language processing techniques, e.g., part of speech tagging and analysis of phrase grammatical structure, can be used to detect an activity labeling style in an automatic fashion. We also study how ontologies, like WordNet, can support the solution. We conduct a thorough evaluation of the developed techniques utilizing about 20,000 activity labels from the SAP Reference Model.

1 Introduction

Modeling of business processes in large enterprises usually implies a team work of numerous specialists. Such teams may span several organizational departments and even several geographical locations. The staff who takes part in the projects can have different professional background. Hence, organization of efficient work in these teams and assuring an appropriate quality of the produced models become real challenges. This situation has motivated researchers and practitioners in business process management to discuss various aspects of model quality [GL07]. First, there are techniques and tools which assure formal properties like behavioral soundness of process models [Aal97]. Second, there are works on how process model characteristics affect model comprehension by human model readers. The importance of this aspect is motivated by the fact that most process models are created for documentation purposes [DGR^{+}06]. It has been shown that large and complex models are more likely to contain errors and are less understandable.

Relatively small attention has been paid to the problem of labeling quality. For instance, a study in [MRR09] showed that the current labeling practice is conducted rather arbitrarily. Meanwhile, labels are the key to understanding the process models by humans. The significance of label quality can be motivated by the dual coding theory [Pai69]. The theory states that humans grasp information more easily, if it is provided via the auditory and the visual channels. In the context of process modeling the visual channel is represented by the graphical constructs of a modeling language, while the auditory channel—by textual model element labels. As process models use only a few graphical constructs, the use of informative and unambiguous labels improves an overall understanding of a process model. In this way, labels contribute to semantic and pragmatic usefulness of a model (see [KSJ06]).

As real world process model repositories can easily include thousands of process descriptions [Ros06], an efficient quality assurance mechanism has to rely on an automatic classification of models according to a specific quality aspect. In terms of label quality the challenge is to identify the labels of poor quality. From the human user perspective, the designer can be supported by a modeling tool identifying poor labels. This functionality can be extended further to giving suggestions on label reengineering.

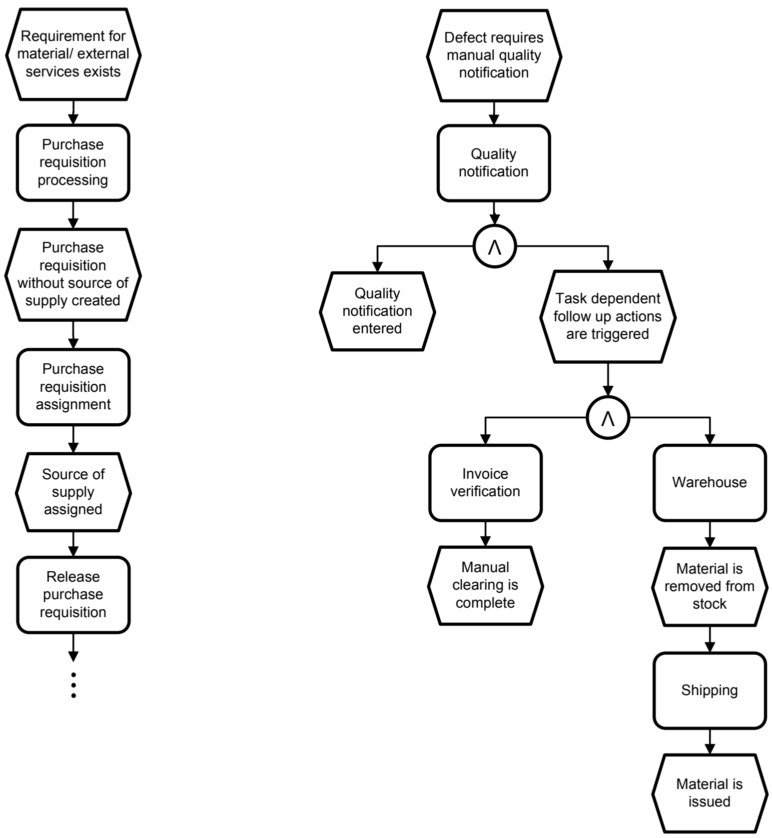
The quality of labels in process models can be discussed along two orthogonal dimensions. On the one hand, appropriate and consistent terms have to be chosen, which is related to a thesaurus. On the other hand, these terms have to be composed according to a particular structure of labels, which can be related to grammatical styles. As the strong influence of label structure on model understanding has been shown in [MRR09], in this work we study the structure of labels. Furthermore, we narrow the scope of our research to the investigation of activity labels. Although a model semantics is defined by the whole constellation of the model elements, activities accumulate its largest part.

The contribution of this paper is the specification and validation of two criteria for assessment of quality of activity labels in process models: frequency of activity labeling style and activity labels length. We propose two metrics on the basis of these criteria and present an automatic method for their quantitative evaluation. While quantitative evaluation of labels length is straightforward, frequency of labeling style requires sophisticated analysis of labels. Furthermore, we perform an evaluation of the SAP Reference Model using these metrics. The SAP Reference Model is a publicly available business process model collection, capturing the processes using Event-driven Process Chains (EPCs) [KT98].

The structure of the paper is as follows. Section 2 motivates this work using the examples of real-world process models. Section 2 discusses what labeling styles can be found in practice and how they correlate with quality and process model understandability. Section 3 introduces the quality metrics for measuring different aspects of structural quality. Subsequently, Section 4 discusses the implementation approach for computing one structural metric automatically. Section 5 presents the results of an evaluation involving the SAP Reference Model. Section 6 discusses the findings in its relationship to the related work. Section 7 concludes the paper and gives on outlook on future research.

2 Labeling Styles

In this section we discuss different styles of labeling. Section 2.1 discusses labeling in the SAP Reference Model, and Section 2.2 the classification of labels. Section 2.3 highlights



(a) Fragment of process model *Purchase Requisition*

(b) Model of process *Return Deliveries*

Figure 1: Labeling in two business process model examples from the SAP Reference Model

implications.

2.1 Labeling in the SAP Reference Model

Fig. 1 shows two examples of business process models from the SAP Reference Model, a process model collection that has been used in several works on process model analysis. Fig. 1(a) depicts a fragment of a business process where a *purchase requisition* is handled. Within this model fragment we observe activity labels *purchase requisition processing*, *purchase requisition assignment* and *release purchase requisition*. In the first two labels the actions are denoted with the nouns *processing* and *assignment*. In the third one the verb *release* corresponds to the action. Obviously, the modelers used several styles for activity labeling. Ambiguity is a potential threat to label understanding. For instance, consider the word *purchase*, which can be both a noun and a verb. This source of ambiguity is called zero derivation, since a verb is linguistically created from a noun without adding a postfix like -*ize* in *computerize*. It has been pointed out that different styles are prone to different degrees of ambiguity [MRR09], which emphasizes the importance of labeling styles for human understanding. If an action noun is used, there is likely an ambiguity, when it is combined with a zero derivation noun. If we consider the *purchase requisition processing* label, it is hard to tell if *purchase* or *processing* stands for an action. As zero derivation is an essential part of the language that cannot always be avoided, it is a useful strategy to employ and enforce a suitable labeling style.

Fig. 1(b) highlights further potential problems of labels. It captures a model of the *return deliveries* business process. In this model, we can observe activity labels that do not signify any action (e.g. *warehouse*) or activities with actions, but without any object (e.g. *shipping*). Again understanding of such activities requires the reader to interpret the context of the model.

2.2 Classes of Labeling Styles

In order to assess the structural label quality of a process model, it is essential to distinguish different labeling styles, which are found in practice. An analysis of the SAP Reference Model was conducted to identify such different styles and their relationship to potential causes of ambiguity [MRR09]. The activity labels are classified into three major structural categories: verb-object style, action-noun style, and the rest category. Table 1 provides categorization of labels used in models in Fig. 1.

The classification approach is based on the grammatical representation of the action in an activity label. For labels belonging to the verb-object style, the action is captured as a verb and used in the imperative form of the verb at the beginning of the label. Examples are labels like *enter count results* or *compare value dates*. Although the name verb-object style suggests the necessity of an object, also labels like *process* or *follow-up* are subsumed to the verb-object category. In a label following the action-noun style the activity is described in terms of a noun, as in *printing notification* or *check of order*. These nouns reflecting the action are either derived from a verb like in *analysis* or are a gerund like *analyzing*. All remaining labels, referred as the rest category, do not contain a word from which an action can be inferred. This applies to labels like *basic settings*.

A deeper analysis of these labeling styles reveals that all of them are prone to specific types of ambiguity. For example, verb-object labels could potentially be misinterpreted if they are affected by a zero derivation ambiguity. This is the case when a word can be both a verb and a noun without adding any suffixes. An example is *measure* in the label *measure processing*. Thus, this label could either refer to the measurement of a processing or the processing of a measure. But also action-noun labels are affected by ambiguity, which is referred as action-object ambiguity. The label *printing notification* emphasizes this problem, since it could either advice to print a notification or to notify somebody to conduct a printing job. As rest labels do not contain a word representing an action, they

Table 1: Activity labels used in the models in Fig. 1

Activity Label	**Labeling Style**	**Action**	**Object**
Release purchase requisition	verb-object	release	purchase requisition
Purchase requisition processing	action-noun	process	purchase requisition
Purchase requisition assignment	action-noun	assign	purchase requisition
Quality notification	action-noun	notify	quality
Invoice verification	action-noun	verify	invoice
Warehouse	rest		
Shipping	rest		

suffer from the verb-inference ambiguity. This means, that a reader of the process model may not be able to evaluate what to perform when reading a label of the rest category.

2.3 Implications for Labeling Practices in Process Models

Knowing the different styles and their ambiguities, the resulting question is which of these styles should be preferred in practice. Several guidelines for conceptual modeling propose to follow a verb-noun convention in which an action is grammatically captured as a verb [KDV02]. This suggestion is also supported in [MRR09]. On the one hand, the authors counted the ambiguity cases in the SAP Reference Model. They uncovered that among verb-object labels 5.1% and among action-noun labels 9% were affected by ambiguity. Due to the verb-inference ambiguity all labels from the rest category were considered as ambiguous. On the other hand, the same research group conducted an experiment studying the impact of labeling styles on perceived ambiguity. A group of students was asked to assess labels in a given process model regarding their ambiguity and their provided usefulness. The result was that verb-object labels were considered to be the least ambiguous and the most useful in comparison to labels of the other styles. Congruent with the frequency of the ambiguity cases, the rest labels were regarded to be most ambiguous and having the lowest usefulness.

Based on these findings, it can be stated that labeling activities according to the verb-object style is the most desirable from a quality point of view. While the authors of [MRR09] manually inspected the labels, this is hardly an option in industry practice. The SAP Reference Model with its 600 process models already contains about 20,000 activity labels, and it is still much smaller than other model collections with several thousand models. Therefore, we need to investigate how label quality in terms of compliance to a particular labeling style can be determined automatically.

3 Metrics for Measuring Structural Label Quality

In this section we present the metrics for measuring the model quality based on the properties of activity labels. We distinguish two main metrics groups: those considering the length of activity labels and those considering the style.

3.1 Metrics Based on the Length of Activity Labels

An activity label has to capture the essential information about the activity on one hand, but should not overload the reader with unnecessary information on the other hand. Among several other means to achieve this, a label length is a natural regulator for the amount of information in the label. In [MS08] the authors investigated the relation between the label length and label understandability. The study has shown that shorter labels facilitate proper understanding of a process model. However, the terms *short* or *long* are relative and imprecise. In the context of this work we need concrete numbers to tell *short* from *long*.

In [Fle51] Flesch argues that the understandability of a text with sentences of length eight or fewer words can be claimed as *very easy*. The author reports that sentences of length 14 still can be understood *fairly easy*. We adapt these findings to the problem of activity labels and claim that activity labels with the length of eight or less words are easy to interpret. We call the labels with the length greater than eight words labels with excess length. Hence, one metric is the fraction of activities which have labels of excess length. We refer to this metric as excess length fraction and denote it as L_e. The lower the value of excess length fraction, the higher is model understandability.

Although short labels improve model understandability, it is vindicable to state that a label should contain at least two words. In particular, a label should point to an action and an object on which the action is performed. Labels of one word length provide readers too small amount of information (see labels *shipping* and *warehouse* in Section 2). Considering this issue, we propose to treat the fraction of activity labels of one word length as a metric as well. We refer to this metric as one word fraction and denote it as L_o. The lower is L_o, the better is the model understandability.

Finally. a natural metric is the average length of activity labels in the model. We denote it with L_a. Again, the lower is L_a, the better is the model understandability. As a result, we distinguish the following three metrics based on the activity label length:

- excess length fraction L_e,
- one word fraction L_o,
- average label length L_a.

3.2 Metrics Based on the Frequency of Activity Labeling Styles

The second group of metrics considers fractions of activity labels adhering to one labeling style. In Section 2 we have identified three labeling styles for activities: verb-object, action-noun and rest. The fractions of activities with labels of particular labeling style result in three metrics for the quality of labels:

- verb-object style fraction S_{vo},
- action-noun style fraction S_{an},
- rest category fraction S_r.

Activities with labels of verb-object style are easy to comprehend for humans. Hence, high values of S_{vo} imply good model understandability. On the opposite, labels of the rest category are the source of high ambiguity. They harden model comprehension and witness of low model quality. As a result, the lower the value of S_r, the better the model understandability is. Thereafter, the fractions of activities with verb-object style and rest style provide sufficient information about labeling quality in a model. For instance, if the model has a large fraction of activities which adhere to verb-object style and small number of activities with rest style, the model has high labeling quality. At the same time, the number of labels with action-noun style depends on the fractions of activities with labels in verb-object and rest styles.

4 Automatic Identification of Labeling Style

In this section we describe different strategies towards an automatic identification of labeling styles. Every label is analyzed independently from others and from the structure of the process model containing the label. We focus on so-called part of speech tagging [JM08] as a tool to identify the grammatical form of the words of an activity label. Part of speech (or grammatical) tagging is a technique from computational linguistics that assigns the part of speech like verb, subject, or object to words of a text based on the syntactical form of the word and its context within a sentence.

Nowadays, there are powerful algorithms available to automatically perform part of speech tagging. The automatic determination of the grammatical structure of the labels is conducted with part of speech tagging tools. Such tools, referred as tagger or parser, are developed for natural language processing. They provide functionality to assign the according part of speech for each word in a given text. Thus, for instance the input string *process order* will lead to the tagging result *process/VB order/NN*. The tag *VB* indicates that the word *process* is a verb in the base form and the tag *NN* represents a singular noun. There exist about 50 different part of speech tags for providing differentiated information about a word [MMS93]. For example, it is possible to distinguish between different verb types. There are part of speech tags for verbs in the past tense (VBD or VBN), for gerunds (VBG) and for verbs in the third person form (VBZ).

In order to provide the process modeler with the best obtainable results, different part of speech tagging tools are considered regarding their accuracy. In this context, the following tools are further investigated: the Stanford Parser, the Stanford Tagger, and WordNet. The tagger and the parser from the Stanford University [TM00, KM03] both provide the functionality to assign part of speech tags to a given text. The Stanford Parser does additionally analyze the structure and relations between the words in each sentence and is thus able to use more information for its part of speech assessment. WordNet is a lexical database and provides different functionality for analyzing semantic relations between words [Mil95]. Amongst others, WordNet provides a function for determining the most likely part of speech for a single word without considering context.

The *Stanford Tagger* requires its input in terms of a text file, which can include all relevant activity labels separated with a "." as punctuation mark. As a result the tagger returns a string where each word is enriched with the according part of speech tag. In initial experiments, we observed some problems with tagging activity labels. In the general case, the Stanford Tagger has been tested to work with an accuracy of about 97% [TM00, TKMS03]. Apparently, the structure of the majority of the labels seems to be unappropriate for it as they are not really natural language sentences. We have seen that about one third of all labels (action-noun style plus rest category) do not even contain a verb, such as for instance *asset maintenance* or *order processing*. Even if the labels are proper sentences, they tend to be rather short like *perform posting* or *edit classes*. All these factors might contribute to a poor tagging result.

In order to improve the results the original labels were extended with the prefix *You have to*. First, it increases the label length, and therefore the grammatical context. Second, it yields a proper English sentence for verb-object languages. For instance, the label *process order* is extended to *You have to process order*. While the word *process* plays still the role of a verb, it might be detected with a higher probability since there is now more information for the tagging tool which can be evaluated. Applying this approach, we observed a considerable increase in tagging performance.

The *Stanford parser* expects the input sentences wise. Hence, the parser has to be fed with all activity labels sequently. Therefore, each label was extended with a "." as punctuation mark. Experiments by the original authors have shown that the tagger tends to work with an accuracy of about 86% [KM03]. The parser considers the relations between the words. Therefore, the accuracy does not only depend on the part of speech tagging but also on the accuracy of the correct detection of grammatical relations. Accordingly, we expect it to provide a higher recall. However, some initial experiments pointed to some weaknesses. These might stem from the fact that activity labels are not really sentences, which might cause problems.

Beyond the mentioned tagger and parser, we considered the lexical database *WordNet* [Mil95] via its corresponding Java implementation called Rita. WordNet provides a function called *getBestPos* that returns the best part of speech for a given word based on its polysemy count. This means that the function will return the part of speech for a given word that captures the most different senses. For instance, *getBestPos* will return *verb* for a word having 8 different senses as a verb and only 6 meanings as a noun. Consequently, WordNet was provided with the first word of each label, as verb-object labels would start with a verb, and we observed quite good performance in initial experiments. We consider a further improvement by combining WordNet with the tagging approach using the extended labels with the prefix. But as the combination of both requires the union of both result sets, also the amount of incorrectly detected labels increases in the combined result set. This implies a decreasing precision value.

5 Evaluation of the SAP Reference Model

In this section we present an evaluation of a real world collection of business process models against the proposed labeling quality metrics. First we introduce the collection, describing its properties relevant for the experiment. Further, we present the evaluation results.

5.1 SAP Reference Model

The experiment studies the SAP Reference Model [KT98], a process model collection that has been used in several works on process model analysis [Men08]. The collection captures business processes that are supported by the SAP R/3 software in its version from the year 2000. It is organized in 29 functional branches of an enterprise, like sales or accounting, that are covered by the SAP software. The SAP Reference Model includes 604 EPCs.

5.2 Metrics Overview for the SAP Reference Model

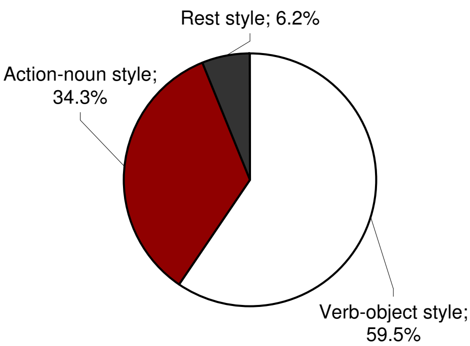


Figure 2: Distribution of activity labeling styles in the SAP Reference Model

Let us first discuss those metrics that are based on the labeling styles. [MRR09] performed an analysis of activity labeling styles employed in the SAP Reference Model. The 19,839 activity labels of the model collection were manually inspected in order to reveal frequencies of labeling styles. About 60% of labels follow the verb-object style, 34% were classified as action-noun labels and only about 6% of the labels belong to the rest category (see Fig. 2). This distribution is quite favorable as a majority of 94% of labels refer an action, while two thirds are verb-object labels. Nevertheless, there are still 6% of all labels which definitely suffer from the verb-inference ambiguity and

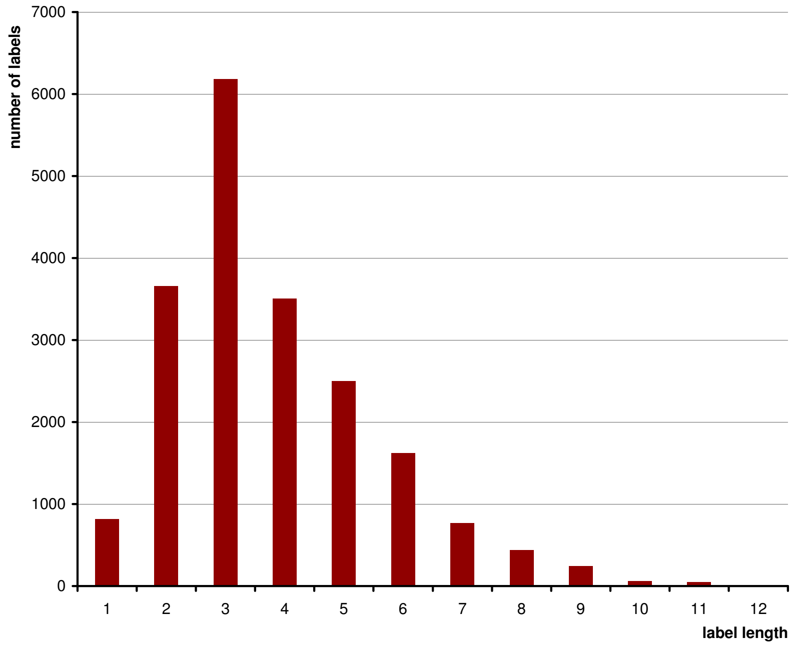


Figure 3: Label length distribution in the SAP Reference Model

might cause misinterpretations.

Figure 3 depicts the length distribution showing that most labels have length of only three words and that there is no label having more than 12 words. The average label length L_a equals to 3.78. Hence, according to [Fle51], most labels can be pronounced *fairly easy* to understand and even *very easy* to understand. The excess length fraction value is $L_e = 3.98\%$. Meanwhile, one word fraction L_o is 4.12%. This fact is important, as these activity labels do not point either to an action or an object, decreasing model label quality.

Summarizing the findings on the SAP Reference Model, the majority of the labels are action orientated or even follow the verb-object style. Moreover, the average label length is below 4 and is thus very short which also supports comprehension. Less favorable are those labels containing only one word or no word referring to an action since those will most likely cause misunderstandings. Table 2 summarizes all introduced metrics for the SAP Reference Model.

5.3 Automatic Part of Speech Identification Results

In this section, we analyze the accuracy of the different strategies to automatic part of speech identification. The manual classification of labels of [MRR09] serves us as a bench-

Table 2: Calculated Metrics for the SAP Reference Model

Metric	Value
Excess Length Fraction (%)	3.98
One Word Label Fraction (%)	4.12
Average Label Length	3.78
Verb-Object-Style Fraction (%)	60
Action-Noun-Style Fraction (%)	34
Rest-Style Fraction (%)	6

mark. We use standard precision (ratio of *found relevant labels* to *all found labels*) and recall (ratio of *found relevant labels* to *all relevant labels*) measurements to assess accuracy.

Figure 4 depicts the results of the different part of speech tagging techniques. It can be seen that the Stanford Parser achieved the lowest recall values. This rather weak performance is likely to be caused by its dependence on accurate contextual information, which is hardly available in short activity labels. The Stanford Tagger showed better results with a recall of about 51%. It was considerably improved by using the *You have to* prefix, which extends verb-object labels to correct English sentences. The rather simple approach of using the probable part of speech function offered by WordNet worked surprisingly well. Combining it with the prefixed tagger yielded almost 99% recall.

Besides the already discussed recall values the Figure shows also the precision value for each approach. This precision value was optimized using two practices. First of all, only those labels were assigned to the verb-object category where the first word in the label has a verb tag indicating a base form. Since only the base form of a verb matches with the imperative form, this approach is reasonable. Thus, for instance labels like *determining protocol proposal/NNP* with a gerund tag *VBG* at the beginning and the label *fixed price billing* starting with a past tense tag *VBN* are excluded from the result set. As Wordnet only differentiates between verb, noun, adjective and adverb, this practice was not applicable for WordNet. By contrast, the second practice for improving the precision value was applicable for all techniques. This approach aims for excluding labels that have a base form verb at the beginning but still do not belong to the verb-object category. Examples are the labels *check of order* or *release of process order*. Apparently, the first word of each label suffers from the zero-derivation ambiguity since considered in isolation they can both be verb and noun. But the preposition *of* uncovers that the first word simply cannot play the role of verb. Therefore, labels with the sequence *verb + of* are also excluded from the result set. These practices excluded up to 5% of all detected labels.

The results suggest two conclusions. If recall and precision are considered in combination, the tagging approach using the prefixes for extending the labels might be an option for an implementation approach. But still, WordNet obtained the best results. Thus, the WordNet approach should be preferred. Although, the combination with the tagger resulted in slightly increased results, the combination is not worthwhile for two reasons. First of all, the precision value decreases. Secondly, the effort for the additional tagging is not reflected in a significant increase of the results.

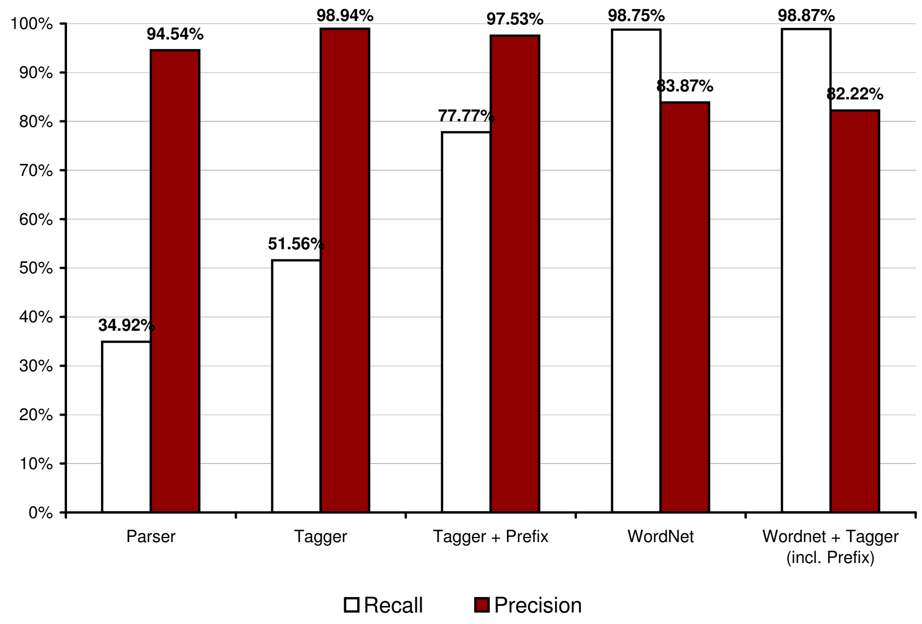


Figure 4: Result of the Analysis on Part of Speech Tagging Techniques

6 Related Work

Our work can be related to three major streams of related work: quality frameworks, process model labeling, and natural language approaches for models.

Process model quality is discussed in different works on *quality frameworks*. The SEQUAL framework builds on semiotic theory and defines several quality aspects [LSS94, KSJ06]. In essence, syntactic quality relates to model and modeling language, semantic quality to model, domain, and knowledge, and pragmatic quality relates to model and modeling and its ability to enable learning and action. The semantic and pragmatic quality clearly point to the relevance of labeling activities. The Guidelines of Modeling (GoM) define an alternative quality framework that is inspired by general accounting principles [BRU00]. The guidelines include the six principles of correctness, clarity, relevance, comparability, economic efficiency, and systematic design, where several of them have implications for good labeling. Also the ISO 9126 [ISO91] quality standard has been suggested as a starting point for model quality [Moo05, GD05].

The verb-object style is widely promoted in the literature for *labeling activities of process models* [Mil61, SM01, MCH03], but rather as informal guidelines. Similar conventions are advocated as guidelines for the creation of understandable use case descriptions, a widely accepted requirements tool in object-oriented software engineering [PVC07]. But in contrast to its promotion in the process modeling domain, it has been observed that verb-object labeling in real process models is not consistently applied. For instance, the practical guide for process modeling with ARIS [Dav01, pp.66-70] shows models with both actions as verbs and as nouns. It has also been shown that shorter activity labels improve model understanding [MS08], which is consistent with readability assessments on sentence length [Gre00, Fle51]. The concept of part of speech tagging is also investigated for interactive process modeling support. In a recent paper, the authors employ it for auto-completion [BDH$^+$09].

The enforcement of the verb-object style might help to close the gap between *natural language and formal language processing*. And indeed, the relationship between process models and natural language has been discussed and utilized in various works. In [FKM05] the authors investigate in how far the three steps of building a conceptual model (linguistic analysis, component mapping, and schema construction) 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 [IGSR05] and to compare process models from a semantic perspective [EKO07]. Most beneficiary is the verb-object style for model verbalization and paraphrasing, see [HC06, FW06]. Such verbalization is an important step in model and requirements validation [NE00]. For instance, verb-object style labels can easily be verbalized using the *You have to* prefix, which we also used in our analysis. In this way, automatic parsing enables a better validation of process models.

7 Conclusion and Future Work

Recent research revealed the high impact of the labeling quality in business process models on the overall model understanding. These findings motivated us to the work reported in this paper. First, we introduced several metrics that help to assess the quality of activity labels in the model. The metrics are categorized into two groups: metrics based on the activity label length and metrics based on the activity labeling style. Among the proposed metrics, we focused on verb-object fraction, since it is crucial for assessment of model labeling quality. We developed and implemented an approach enabling evaluation of verb-object fraction metrics. For evaluation of the techniques we proposed, we use the SAP Reference Model. The results show that high precision and recall can be achieved automatically by using part of speech tagging techniques and available tools.

In this paper we focused only on the labeling quality of activities in business process models. However, there are other model elements like events and data objects that should be subject to label quality assurance. It is part of our future research agenda to identify how part of speech tagging techniques can be applied for those labels as well. Our results also depend on using English as a language for labeling activities. It will be an interesting task to analyze other languages, like German or Russian, to see whether part of speech tagging can be utilized with the same accuracy. Finally, we plan to use tagging information for building taxonomies for process model collections. Identifying nouns from the label will be a crucial step for this application.

References

[Aal97] W.M.P. van der Aalst. Verification of Workflow Nets. In Pierre Azéma and Gianfranco Balbo, editors, *Application and Theory of Petri Nets 1997*, volume 1248 of *Lecture Notes in Computer Science*, pages 407–426. Springer Verlag, 1997.

[BDH+09] Jörg Becker, Patrick Delfmann, Sebastian Herwig, Lukasz Lis, and Armin Stein. Towards Increased Comparability of Conceptual Models - Enforcing Naming Conventions through Domain Thesauri and Linguistic Grammars. page (forthcoming), 2009.

[BRU00] J. Becker, M. Rosemann, and C. von Uthmann. Guidelines of Business Process Modeling. In W.M.P. van der Aalst, J. Desel, and A. Oberweis, editors, *Business Process Management. Models, Techniques, and Empirical Studies*, pages 30–49. Springer, Berlin et al., 2000.

[Dav01] R. Davis. *Business Process Modelling With Aris: A Practical Guide*. Springer, 2001.

[DGR+06] I. Davies, P. Green, M. Rosemann, M. Indulska, and S. Gallo. How do practitioners use conceptual modeling in practice? *Data & Knowledge Engineering*, 58(3):358–380, 2006.

[EKO07] M. Ehrig, A. Koschmider, and A. Oberweis. Measuring Similarity between Semantic Business Process Models. In J.F. Roddick and A. Hinze, editors, *Conceptual Modelling 2007, Proceedings of the Fourth Asia-Pacific Conference on Conceptual Modelling (APCCM 2007)*, volume 67, pages 71–80, Ballarat, Victoria, Australia, 2007. Australian Computer Science Communications.

[FKM05] G. Fliedl, C. Kop, and H.C. Mayr. From textual scenarios to a conceptual schema. *Data and Knowledge Engineering*, 55(1):20–37, 2005.

[Fle51] R. Flesch. *How to Test Readability*. Harper & Brothers, New York, NY, USA, 1951.

[FW06] P.J.M. Frederiks and T.P. van der Weide. Information Modeling: The Process and the Required Competencies of Its Participants. *Data & Knowledge Engineering*, 58(1):4–20, 2006.

[GD05] A. Selçuk Gücegliogu and O. Demirörs. Using Software Quality Characteristics to Measure Business Process Quality. In W.M.P. van der Aalst, B. Benatallah, F. Casati, and F. Curbera, editors, *Business Process Management, 3rd International Conference, BPM 2005, Nancy, France, September 5-8, 2005, Proceedings*, volume 3649 of *Lecture Notes in Computer Science (LNCS)*, pages 374–379. Springer Verlag, 2005.

[GL07] V. Gruhn and R. Laue. What Business Process Modelers Can Learn from Programmers. *Sci. Comput. Program.*, 65(1):4–13, 2007.

[Gre00] H. Gretchen. Readability and Computer Documentation. *ACM J. Comput. Doc.*, 24(3):122–131, 2000.

[HC06] T.A. Halpin and M. Curland. Automated Verbalization for ORM 2. In R. Meersman, Z. Tari, and P. Herrero, editors, *On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops, Montpellier, France, October 29 - November 3. Proceedings, Part II*, volume 4278 of *Lecture Notes in Computer Science*, pages 1181–1190. Springer, 2006.

[IGSR05] J.E. Ingvaldsen, J.A. Gulla, X. Su, and H. Rønneberg. A Text Mining Approach to Integrating Business Process Models and Governing Documents. In R. Meersman et. al, editor, *On the Move to Meaningful Internet Systems 2005: OTM 2005 Workshops, OTM Confederated International Workshops and Posters, AWeSOMe, CAMS, GADA,* *MIOS+INTEROP, ORM, PhDS, SeBGIS, SWWS, and WOSE 2005, Agia Napa, Cyprus, October 31 - November 4, 2005, Proceedings*, volume 3762 of *Lecture Notes in Computer Science*, pages 473–484. Springer, 2005.

[ISO91] International Standards Organisation ISO. Information Technology - Software Product Evaluation - Quality Characteristics and Guide Lines for their Use. Iso/iec is 9126, 1991.

[JM08] D. Jurafsky and J.H. Martin. *Speech and language processing*. Prentice Hall, 2008.

[KDV02] H. Koning, C. Dormann, and H. van Vliet. Practical Guidelines for the Readability of IT-architecture Diagrams. In *In: Proceedings of the 20th Annual International Conference on Documentation, ACM SIGDOC 2002*, pages 90–99, 2002.

[KM03] D. Klein and Ch. D. Manning. Accurate Unlexicalized Parsing. *41st Meeting of the Association for Computational Linguistics*, pages 423–430, 2003.

[KSJ06] J. Krogstie, G. Sindre, and H.D. Jørgensen. Process models representing knowledge for action: a revised quality framework. *European Journal of Information Systems*, 15(1):91–102, 2006.

[KT98] G. Keller and T. Teufel. *SAP(R) R/3 Process Oriented Implementation: Iterative Process Prototyping*. Addison-Wesley, 1998.

[LSS94] O.I. Lindland, G. Sindre, and A. Sølvberg. Understanding Quality in Conceptual Modeling. *IEEE Software*, 11(2):42–49, 1994.

[MCH03] T.W. Malone, K. Crowston, and G.A. Herman, editors. *Organizing Business Knowledge: The MIT Process Handbook*. The MIT Press, 2003.

[Men08] Jan Mendling. *Metrics for Process Models: Empirical Foundations of Verification, Error Prediction, and Guidelines for Correctness*, volume 6 of *Lecture Notes in Business Information Processing*. Springer, 2008.

[Mil61] L.D. Miles. *Techniques of value analysis and engineering*. McGraw-hill, 1961.

[Mil95] G. A. Miller. WordNet: a Lexical Database for English. *Commun. ACM*, 38(11):39–41, 1995.

[MMS93] M. P. Marcus, M. A. Marcinkiewicz, and B. Santorini. Building a Large Annotated Corpus of English: The Penn Treebank. *Computational Linguistics*, 1993.

[Moo05] D.L. Moody. Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions. *Data & Knowledge Engineering*, 55(3):243–276, 2005.

[MRR09] J. Mendling, H. A. Reijers, and J. C. Recker. Activity Labeling in Process Modeling: Empirical Insights and Recommendations. *Information Systems*, 2009.

[MS08] J. Mendling and M. Strembeck. Influence Factors of Understanding Business Process Models. In W. Abramowicz and D. Fensel, editors, *Proc. of the 11th International Conference on Business Information Systems (BIS 2008)*, volume 7 of *Lecture Notes in Business Information Processing*, page 142–153. Springer-Verlag, 2008.

[NE00] B. Nuseibeh and S.M. Easterbrook. Requirements engineering: a roadmap. pages 35–46, 2000.

[Pai69] A. Paivio. Mental Imagery in Associative Learning and Memory. *Psychological Review*, 76:241–263, 1969.

[PVC07] Keith Thomas Phalp, Jonathan Vincent, and Karl Cox. Improving the Quality of Use Case Descriptions: Empirical Assessment of Writing Guidelines. *Software Quality Journal*, 15(4):383–399, 2007.

[Ros06] M. Rosemann. Potential pitfalls of process modeling: part A. *Business Process Management Journal*, 12(2):249–254, 2006.

[SM01] A. Sharp and P. McDermott. *Workflow Modeling: Tools for Process Improvement and Application Development*. Artech House Publishers, 2001.

[TKMS03] K. Toutanova, D. Klein, Ch. D. Manning, and Y. Singer. Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. *HLT-NAACL*, pages 252–259, 2003.

[TM00] K. Toutanova and Ch. D. Manning. Enriching the Knowledge Sources Used in a Maximum Entropy Part-of-Speech Tagger. *EMNLP*, pages 63–70, 2000.