

Towards Understandability Evaluation of Business Process Models using Activity Textual Analysis

Andrii Kopp, Dmytro Orlovskiy and Sergey Orekhov

National Technical University “Kharkiv Polytechnic Institute”, Kyrpychova str. 2, Kharkiv, 61002, Ukraine

Abstract

There are two purposes of business process modeling. Business process models are created by business analysts for understanding, analysis, and improvement of process scenarios, search, and elimination of weak spots and bottlenecks in organizational activities. Another purpose of business process models is the requirements engineering in software development projects. In both cases, the quality of created business process models is the core issue. Poor models are similar to text documents written with mistakes – they are not understandable, which may negatively impact the real processes they represent and the software workflows they describe. However, existing studies in the field of business process model quality mostly focus on the structural analysis of models using size, complexity, and other metrics with thresholds, while the textual analysis of activity labels is omitted. Therefore, in this paper, we propose an approach to the analysis of business process model understandability taking into account best practices of activity labeling. The proposed approach includes the use of natural language processing techniques, so the respective software tool was developed to perform experiments with a set of business process models. According to obtained results, we suggest considering both textual and structural qualities to achieve the understandability of business process models due to the bad correlation between these metrics (0.0171) – well-structured models can have unclear activity labels and vice versa.

Keywords

Business Process Model, Model Quality, Model Understandability, Textual Analysis.

1. Introduction: Related Work and Problem Statement

Business processes are organized sequences of activities that take different kinds of input and produce value for customers, e.g. goods or services. Nowadays Business Process Management (BPM) is the widely used management approach. This approach is based on the business process modeling technique – a visual representation of organizational activities, events, and decisions using graphical diagrams. Business process models are the most valuable assets of the BPM lifecycle. They help to design, analyze, improve, and automate organizational workflows [1]. Business process modeling helps stakeholders to understand, capture (i.e. document using graphical models), analyze, and improve the enterprise workflows. The analysis stage includes performance measurement and errors detection activities, which help to improve captured business processes [2].

1.1. Related Work

According to the analysis of the latest survey, there are various business process modeling notations used to document business operations in companies that practice the BPM approach [3]:

- 64% of respondents use BPMN (Business Process Model and Notation);

MoMLeT+DS 2022: 4th International Workshop on Modern Machine Learning Technologies and Data Science, November, 25-26, 2022, Leiden-Lviv, The Netherlands-Ukraine

EMAIL: kopp93@gmail.com (A. Kopp); orlovskiy.dm@gmail.com (D. Orlovskiy); sergey.v.orekhov@gmail.com (S. Orekhov)

ORCID: 0000-0002-3189-5623 (A. Kopp); 0000-0002-8261-2988 (D. Orlovskiy); 0000-0002-5040-5861 (S. Orekhov)



© 2022 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

- 18% of survey participants use EPC (Event-driven Process Chain);
- 4% of organizations use IDEF-based notations, e.g. IDEF0 and DFD (Data Flow Diagram).

Other survey participants use less popular business process modeling notations, however, the BPMN notation is a leader and currently the de-facto standard for business process modeling [3].

According to [4], BPMN models describe workflows as sequences of tasks and events connected using control flows (Fig. 1). Moreover, business processes described using the BPMN notation contain start events and end events to signalize their beginning and finishing (Fig. 1). Hence, the simplest BPMN business process consists of events and activities [4]:

- things that happen in an instant are represented by events;
- activities are work units that have a set duration.

Also, events and activities are logically related in a business process workflow using sequences. A sequence means that one event or activity is followed by another event or activity [4]. Fig. 1 shows the most basic business process structure, described using BPMN graphical notation, that consists of events (start and end) and activities connected using sequences (also referred to as arcs).

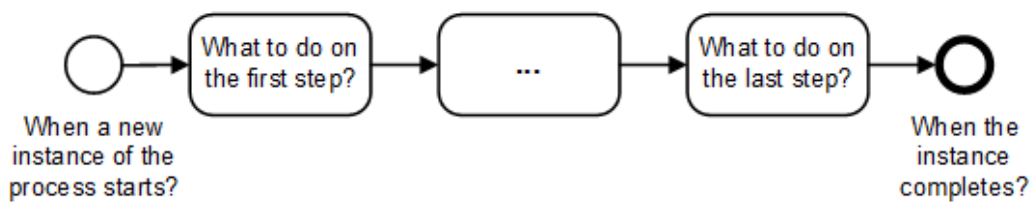


Figure 1: The most basic business process structure described using BPMN graphical notation [4]

According to Fig. 1, when describing a business process using BPMN graphical notation, the modeler should answer the following questions:

- “when a new instance of the business process starts?” – for the start event;
- “when the instance completes?” – for the end event;
- “what to do on the particular process step?” – for activities.

Thus, if events are usually named as combinations of nouns followed by verbs in past participle form (i.e. “order received”, “order fulfilled”), which is quite intuitive, empirical studies have shown that real-world business process models created by many practitioners do not always follow naming conventions for activities [5]. The verb-object labeling style (i.e. a verb in infinitive form followed by the noun: “submit order”, “confirm order”, etc.) is recommended for activity labels [5]. This rule is even included in the Seven Process Modeling Guidelines (7PMG) by Mendling et al. [6].

Fig. 1 demonstrates all the essential elements of BPMN graphical notation [7].

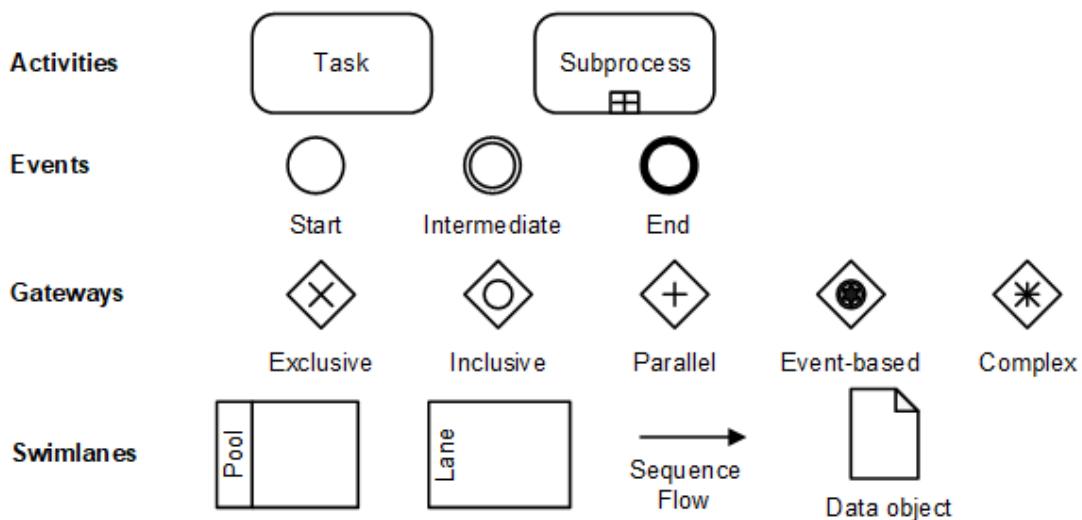


Figure 2: The essential elements of BPMN graphical notation [7]

Advanced business process models created using BPMN graphical notation may contain particular elements to demonstrate the branching and merging workflow scenarios, business process boundaries, and participants. Gateways (Fig. 2) are particular elements that define parallel (AND), inclusive (OR), or exclusive (XOR) branching within workflow scenarios. Pools describe the boundaries of business processes, while lanes define different roles of business process participants [7].

According to [8], there are various metrics and thresholds exist to evaluate BPMN models:

- size (i.e. the number of tasks, events, gateways, and control flows).
- gateway mismatch (the sum of gateway pairs of different types).
- connectivity coefficient (the number of arcs divided by the number of nodes).
- control flow complexity (the sum of gateways weighted by their possible combinations of states after the split).

Other studies are also focused mostly on size metrics for the evaluation of business process model efficiency from understandability and maintainability views:

- authors of [9] have analyzed a large collection of BPMN models created by practitioners and found that improper usage of splits and joins, message flows, decomposition, and labeling lead to the poor quality of business process models;
- in [10] authors propose control-flow complexity metrics and corresponding threshold values they have obtained using data mining techniques to help designers evaluate the quality of business process models;
- authors of [11] formulate the importance of having high-quality business process models as inputs for requirements engineering since the quality of BPMN models influences the software quality; however, this study proposes quality checklists for model reviewers instead of metric and formal approaches to verify the business process model quality.

We have discovered within the context of BPMN and quality assurance two more interesting studies [12] and [13] that consider the quality of the business process itself and do not analyze the quality of a business process model reflecting a particular process.

1.2. Problem Statement

Thus, poorly designed business process models are hard for understanding and maintenance, and they cannot be efficiently used to document business operations, measure business performance, or find workflow errors that may reduce organizational performance. However, existing studies mostly focus on structural analysis of BPMN model flow using the size and control-flow metrics, and thresholds, while relatively smaller attention is paid to the textual analysis of activity labels used in business process models.

Hence, in this study, we propose to pay more attention to labeling styles used for business process model activities (i.e. tasks and collapsed sub-processes) when analyzing the understandability of BPMN models. The soundness of the business process model structure is extremely important for the proper understanding of process scenarios, decisions, occurring events, and other important workflow elements by readers. However, improper naming of activities may mislead the essential understanding of which particular tasks should be completed on each step of the business process scenario or which exactly sub-processes should be initialized. This misunderstanding caused by invalid activity labels can negatively impact business processes and software guided by business process models with these poorly-described activities.

Let us formally describe a business process model as a coherent directed labeled graph [14]:

$$BP_{Graph} = (N, F, L, \lambda), \quad (1)$$

where:

- N is the set of business process elements, which includes subsets of activities A , events E , and gateways G ;
- A is the set of activities;
- E is the set of events, which includes subsets of start events E^s , intermediate events E^i , and end events E^e ;

- G is the set of gateways, which includes subsets of XOR gateways G^{xor} , AND gateways G^{and} , and OR gateways G^{or} ;
- F is the set of sequence flows between business process elements, $F \subseteq N \times N$;
- L is the set of labels defined for business process elements and sequence flows;
- λ is the mapping that assigns labels to business process elements and sequence flows, $\lambda: N \cup F \rightarrow L$.

Thus, the formal statement of a high-quality business process modeling to achieve understandable diagrams may be given as the following:

$$\begin{aligned} Q_{Structural}(BP_{Graph}) &\rightarrow \max, \\ Q_{Textual}(BP_{Graph}) &\rightarrow \max, \end{aligned} \quad (2)$$

where:

- $Q_{Structural}$ is the mapping that assigns respective structural quality values to business process models, $Q_{Structural}: BP_{Graph} \rightarrow [0,1]$;
- $Q_{Textual}$ is the mapping that assigns respective textual quality values to business process models, $Q_{Textual}: BP_{Graph} \rightarrow [0,1]$.

Equation (2) formally describes the problem of business process modeling, according to which created BPMN diagram should be of maximum structural and textual quality [5].

The demonstrated graph (1) can be built automatically, as the result of a BPMN file processing, which is the XML (eXtensible Markup Language) document created according to the specific schema of the BPMN 2.0 format [15].

Hence, we suggest the following workflow of the approach to understandability evaluation of BPMN 2.0 business process descriptions (Fig. 3).

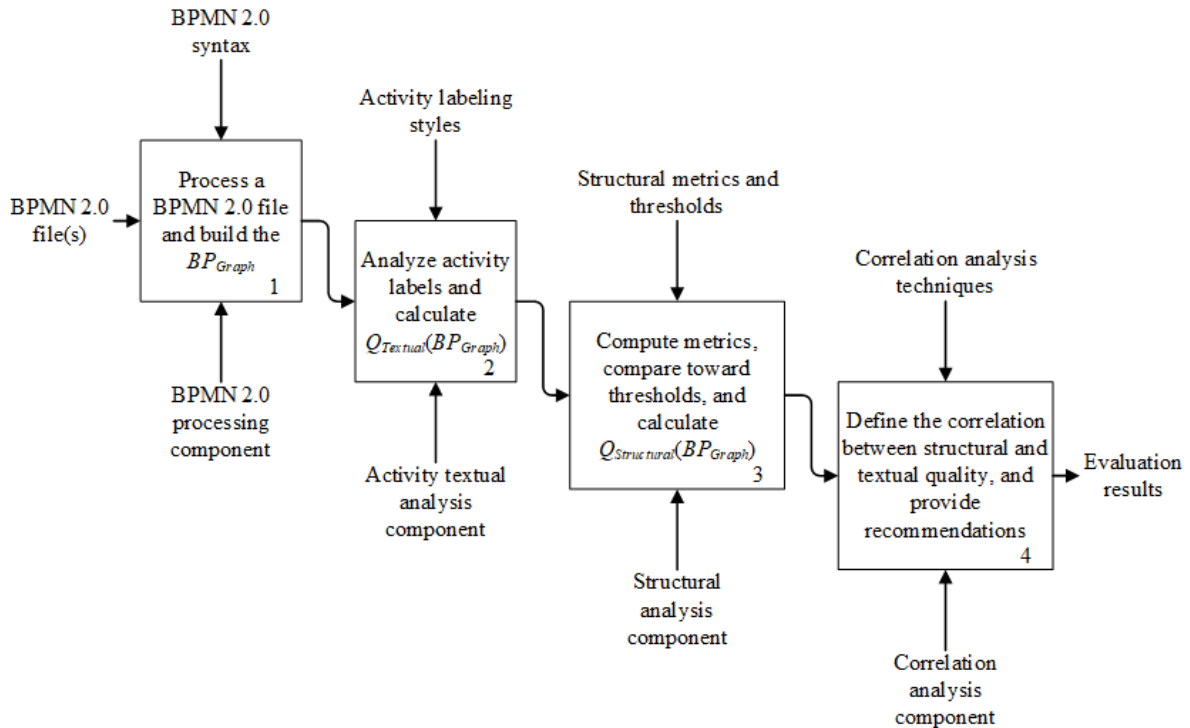


Figure 3: The BPMN 2.0 business process models understandability evaluation workflow

The proposed approach (Fig. 3) may not only allow evaluation of the understandability of BPMN models based on the textual analysis of business process activities but also answer the following question – “does the structural quality of business process models affects their textual quality?”. This may help to formulate recommendations for business process modelers to pay attention not only to the

structural soundness of created diagrams but also to the textual quality of described business process steps to achieve better understandability of models and make sure they serve their purpose.

Therefore, in this study, we need an approach to the textual analysis of business process model activity labels to elaborate the techniques of understandability evaluation of BPMN diagrams. We assume that our approach may include the use of Natural Language Processing (NLP) techniques and work with collections of BPMN 2.0 files, so the particular software tool should be developed to perform experiments with a set of business process models. In general, this study considers the process of business process modeling using BPMN graphical notation and aims at the improvement of created models' quality to assure their understandability by stakeholders for organizational activity analysis and software engineering.

The rest of this paper is organized as follows. Section 2 outlines the textual analysis approach for the evaluation of business process model understandability. Section 3 proposes the structural analysis of business process models based on metrics and thresholds. Section 4 includes experiments, analysis, and discussion of the obtained results.

2. Textual Analysis of Business Process Model Activity Labels

2.1. Activity Labels Extraction from BPMN Models

Before the proposed approach outline, let us demonstrate the sample BPMN 2.0 business process model and its file representation (Fig. 4). According to the example below (Fig. 4), the “process” tag includes all core business process items such as events (i.e. “startEvent” and “endEvent”), activities (i.e. “task”), and sequence flows (i.e. “sequenceFlow”) [16]. Thus, it is quite easy to read such an XML document and represent it formally using the coherent directed labeled graph (1).

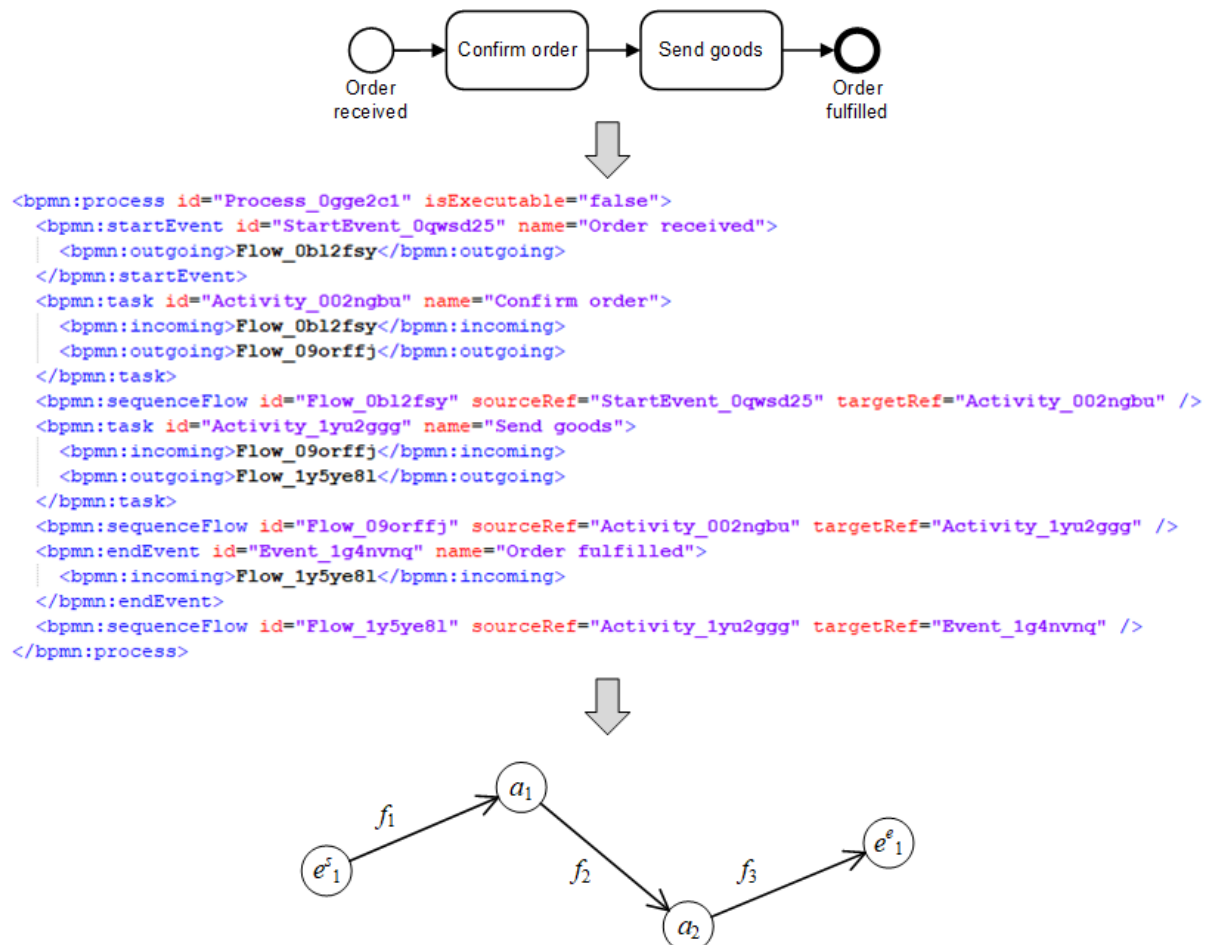


Figure 4: Example of BPMN 2.0 model translation into the graph (1)

Described graph (Fig. 4) consists of the following sets of business process items:

- start events $E^s = \{e_1^s\}$;
- end events $E^e = \{e_1^e\}$;
- activities $A = \{a_1, a_2\}$;
- sequence flows $F = \{f_1, f_2, f_3\}$.

In addition, the mapping λ assigns labels to business process elements and sequence flows, which can be extracted using the “name” attribute of respective tags (Fig. 4):

- $\lambda(e_1^s)$ = "Order received" – using the “name” attribute of the “startEvent” tag;
- $\lambda(a_1)$ = "Confirm order" – using the “name” attribute of the first “task” tag;
- $\lambda(a_2)$ = "Send goods" – using the “name” attribute of the second “task” tag;
- $\lambda(e_1^e)$ = "Order fulfilled" – using the “name” attribute of the “endEvent” tag.

Therefore, it is possible to obtain the set of activity labels $L^{activity} \subseteq L$:

$$L^{activity} = \{l_i^{activity}, i = \overline{1, |A|}\} \quad (3)$$

where $l_i^{activity}$ is the label assigned to the i -th activity $a_i \in A$, $i = \overline{1, |A|}$.

2.2. Activity Labels Analysis Method based on Natural Language Processing

Let us describe the proposed method of textual analysis of business process model activity labels extracted from BPMN 2.0 documents (3).

1. Tokenize each activity label $l_i^{activity} \in L^{activity}$, $i = \overline{1, |A|}$ to get bags of words that correspond to each of the business process activities.

$$\theta : L^{activity} \rightarrow W^{activity}, \quad (4)$$

where:

- θ is the mapping that assigns a bag of words $w_i^{activity} \in W^{activity}$ to each activity label $l_i^{activity} \in L^{activity}$, $i = \overline{1, |A|}$;
- $W^{activity}$ is the collection of bags of words $w_i^{activity} \in W^{activity}$ formulated for each activity label $l_i^{activity} \in L^{activity}$, $i = \overline{1, |A|}$.

2. For each word of tokenized activity labels (4) define one or several parts of speech to which it belongs:

$$\varphi : w_i^{activity} \rightarrow PoS, \quad (5)$$

where:

- φ is the mapping that assigns one or several parts of speech $PoS_i \subseteq PoS$ to each word that belongs to the bag of words $w_i^{activity} \in W^{activity}$ created for each activity label $l_i^{activity} \in L^{activity}$, $i = \overline{1, |A|}$;
- PoS is the set of all parts of speech that can be assigned to each of words in tokenized activity labels, $PoS = \{Noun, Verb, Adjective, Adverb\}$.

3. For each activity label check its length (i.e. the number of words it contains) and if the label consists of at least two words, check if the first and second words are verbs and nouns correspondingly (5):

$$\forall_{i=\overline{1, |A|}} l_i^{activity} \in L^{activity} : q_i^{activity}(l_i^{activity}) = \begin{cases} 0, & |w_i^{activity} = \theta(l_i^{activity})| < 2, \\ 1, & Verb \in \varphi(w_i^{activity}(0)) \wedge Noun \in \varphi(w_i^{activity}(1)), \\ 0, & else, \end{cases} \quad (6)$$

where $q_i^{activity}$ is the mathematical logic predicate that returns 1 for activity labels that match the verb-object labeling style and 0 for activity labels that do not match the verb-object labeling style, $q_i^{activity} \in \{0,1\}$.

4. Calculate the textual quality as the ratio between the number of activities, which labels match the verb-object labeling style (6), and the total number of business process activities:

$$Q_{Textual}(BP_{Graph}) = \frac{1}{|A|} \sum_{i=1}^{|A|} q_i^{activity}(I_i^{activity}). \quad (7)$$

Fig. 5 demonstrates the algorithm of the proposed activity labels analysis method.

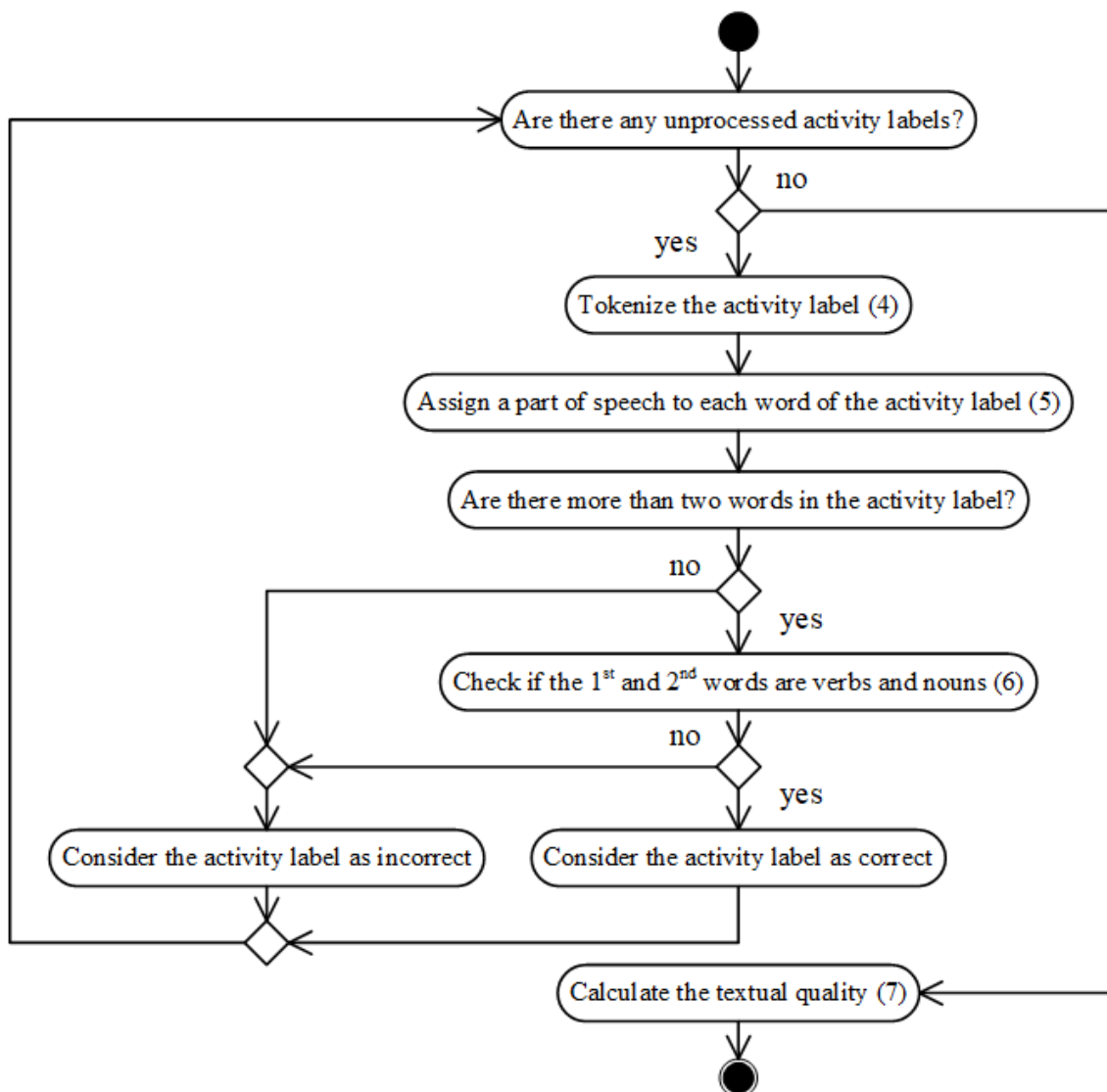


Figure 5: The algorithm of activity labels analysis method

Activity labels tokenization and part of speech assignment to extracted words can be achieved using particular NLP software components, which will be used for experiments in Section 4.

3. Structural Analysis of Business Process Models based on Metrics and Thresholds

Let us also describe the method for structural analysis of business process models to then answer the question of how the structural quality of business process models affects their textual quality.

1. Calculate values of the basic structural metrics proposed in [5] and [6] to manage the business process model's structural quality:

$$M_{Structural} = (|N|, |E^s|, |E^e|, |G^{or}|), \quad (8)$$

where:

- $|N|$ is the number of nodes;
- $|E^s|$ is the number of start events;
- $|E^e|$ is the number of end events;
- $|G^{or}|$ is the number of OR gateways.

2. Therefore, using business process modeling guidelines defined in [5] and [6], the following threshold values can be defined for the respective structural metrics (8):

$$T_{Structural} = (31, 2, 2, 0). \quad (9)$$

Given threshold values (9) reflect the business process modeling guidelines suggested by authors of [5] and [6], which say:

- do not use more than 31 nodes;
- do not use more than 2 start and end events;
- do not use OR gateways.

These threshold values (9) were also confirmed in the latest paper by Mendling et al. [17].

3. Then, using values of the basic structural metrics (8) and corresponding threshold values (9), calculate the structural quality as the average of inverse sigmoid function results:

$$Q_{Structural}(BP_{Graph}) = \frac{1}{|M_{Structural}|} \sum_{j=1}^{|M_{Structural}|} v(m_j, t_j), \quad (10)$$

where:

- m_j is the value of j -th structural metric (8);
- t_j is the threshold value for j -th structural metric (9);
- $v(m_j, t_j)$ is the function that returns values in the range $[0, 1]$:

$$v(m_j, t_j) = \begin{cases} 1, & m_j \leq t_j, \\ \frac{1}{1 + e^{(m_j - t_j - 1)}}, & m_j > t_j. \end{cases} \quad (11)$$

In (11) obtained $v(m_j, t_j) = 1$ values signalize that the value of j -th structural metric m_j completely corresponds to the respective threshold value t_j while smaller values $v(m_j, t_j) < 1$ signalize violations of thresholds (9) by the metric values (8).

4. Results and Discussion

Let us use the collection of BPMN diagrams created during business process modeling training sessions by Camunda company. This collection of BPMN 2.0 diagrams includes four subsets that describe four business processes: goods dispatch, insurance recourse, credit-scoring, and self-service restaurant flows. It is freely available in Camunda's GitHub repository for research purposes [18].

In general, this dataset includes 197 models in English:

- 67 models are alternative versions that describe the goods dispatch business process;
- 47 models are alternative versions that describe the insurance recourse business process;
- 34 models are alternative versions that describe credit-scoring business processes;
- 49 models are alternative versions that describe self-service restaurant business processes.

Hence, to perform experiments with such a collection of BPMN 2.0 files, the software tool was created. It was built using the Python programming language, which has a great tool NLTK (Natural Language Toolkit) for working with computational linguistics [19].

Fig. 6 below demonstrates the workflow and dependencies of the developed software tool, which will be used to perform experiments in this study.

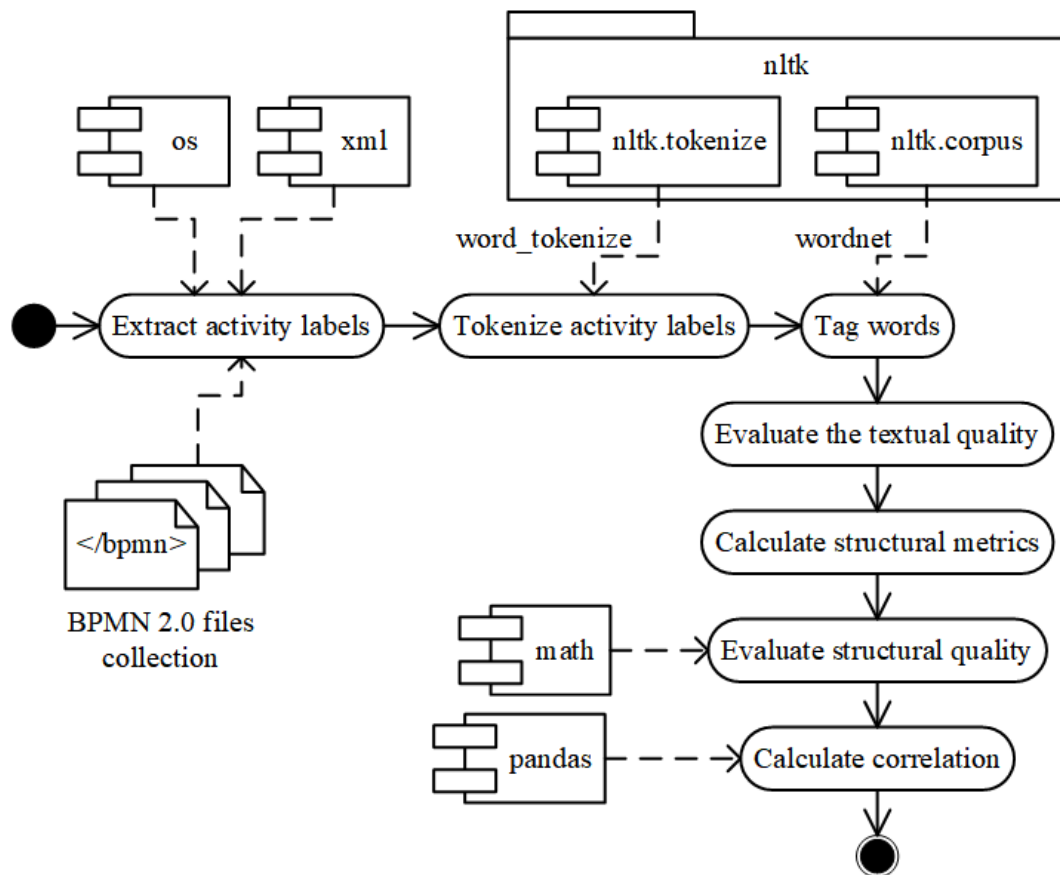


Figure 6: The software tool created to conduct experiments

According to Fig. 6, the developed software tool uses the following external packages:

- the “os” and “xml” packages for working with the file system and processing BPMN 2.0 models that are stored as XML files;
- the “nltk” package for tokenization of activity labels (the “word_tokenize” utility) and words tagging (the “wordnet” lexical database);
- the “math” package for calculations, e.g. exponentiation;
- the “pandas” package for the correlation analysis to study the relationship between business process models’ textual and structural quality.

Table 1 below shows correlation analysis results obtained using the Pandas package that allows the computation of the Pearson standard correlation coefficient [20].

Table 1

The correlation analysis results

| Metrics | Textual quality | Structural quality |
|--------------------|-----------------|--------------------|
| Textual quality | 1.0000 | 0.0171 |
| Structural quality | 0.0171 | 1.0000 |

Calculated correlation analysis results (Table 1) demonstrate bad correlation (0.0171) which means there is no relationship between textual (7) and structural (10) quality coefficients calculated for each of the experimental BPMN business process models [18].

All of these business process models were designed by different persons that were using textual descriptions of business processes they are supposed to create as part of BPMN training sessions. Thus, we may conclude that textual and structural quality dimensions of business process modeling

using BPMN graphical notation are not connected. For example, among the obtained calculation results we can discover perfect BPMN models from the textual quality point of view, but poor BPMN models from the structural quality point of view and vice versa.

Table 2 demonstrate such cases:

- the business process model of high textual quality (1.00) has structural issues (0.88) – the OR gateway is used (Fig. 7);

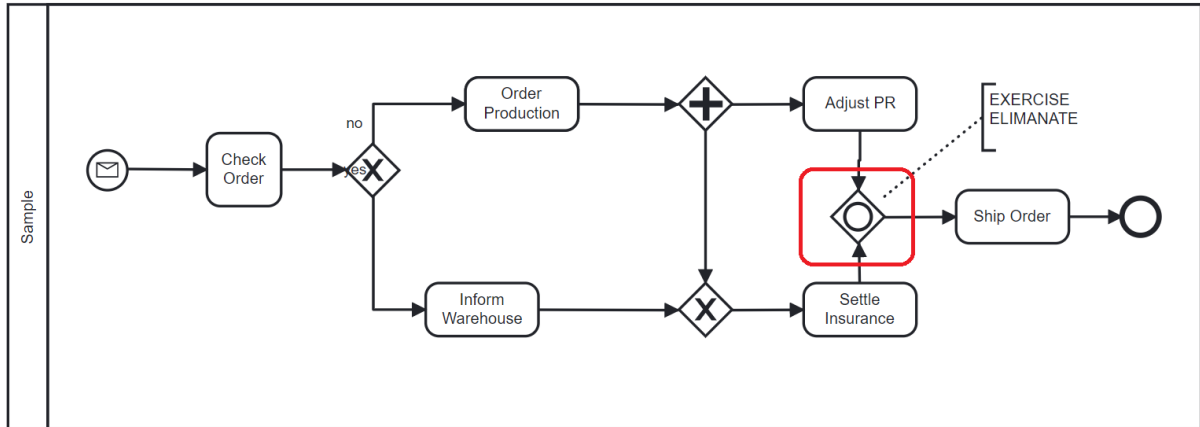


Figure 7: The model of high textual quality but with structural issues

- the business process model of high structural quality (1.00) has poor textual quality (0.43) – 4 of 7 activities has labelling style that does not match the recommended verb-object style.

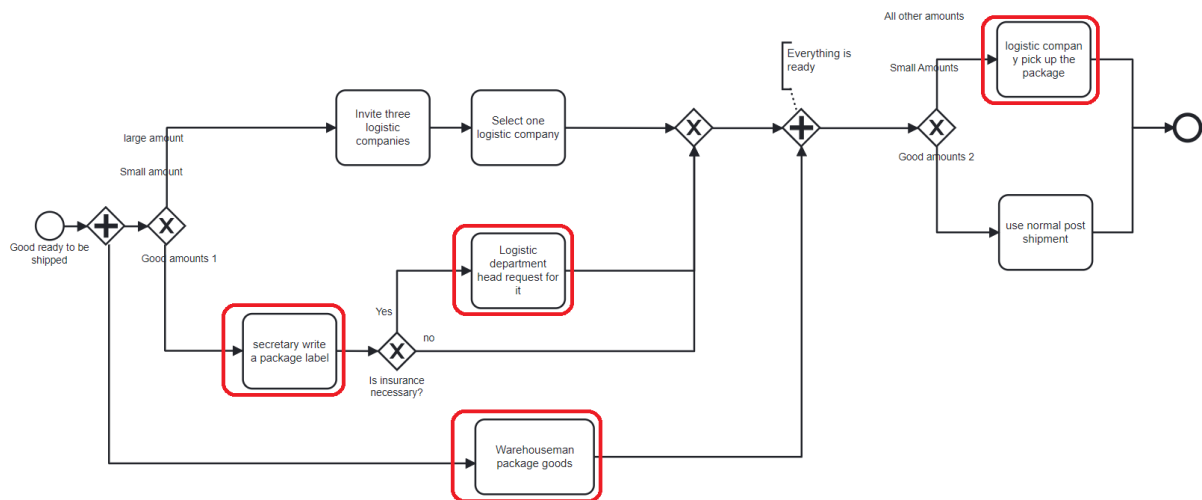


Figure 8: The model of high structural quality but poor textual quality

Table 2

Examples of business process models with opposite textual and structural quality indexes

| Business process model | Textual quality | Structural quality |
|--|-----------------|--------------------|
| Warenversand_035d8eef52bc4e36aac840bdd2feff21.bpmn | 1.00 | 0.88 |
| Exercise_1_21a36e3570ab48d59098702f4f8ad279.bpmn | 0.42 | 1.00 |

Indeed, the model can be perfectly structured but have uninformative activity labels (see 2nd row in Table 2), while there could be desired labeling style used (e.g. verb-object style as the recommended best practice) but the process scenario can be poorly structured so there will be barely understandable in which way activities and events follow each other (see 1st row in Table 2).

5. Conclusion and Future Work

In this paper, we addressed the problem of the understandability evaluation of business process models using the textual analysis of activity labels. We focused on the BPMN diagramming notation since it is the de-facto standard for business process modeling nowadays, which allows the creation of not only visual models but also machine-readable XML-alike files for interexchange between BPM suites and workflow automation. As it was discovered in the related work in the domain of business process model quality analysis, the structural-based approaches that use metrics and thresholds are much more elaborated than approaches based on textual analysis of BPMN activity labels. We identified this situation as a serious limitation – a business process model can have a perfect structure but can have poorly labeled activities making such a model hard to understand by involved stakeholders. Poor models that are not understandable can lead to errors in organizational improvement and software development projects, cause extra resource allocation to fix arising errors, and, therefore, more costs.

Therefore, in this paper, we proposed an approach to the analysis of business process models' understandability taking into account best practices of activity labeling. The proposed approach and the software tool created for experimental processing of the sample BPMN 2.0 files collection are based on particular NLP techniques such as tokenization and part of speech tagging.

Obtained results confirm that the structural quality of a business process model does not mean its understandability since there is a bad correlation between these metrics (0.0171). Provided examples (Fig. 7 and 8, Table 2) show how the models of high textual quality (1.00) can be of moderate structural quality (0.88) and vice versa – how the models of poor textual quality (0.42) can be of high structural quality (1.00). Therefore, understandable business process models, which are valuable for the stakeholders, should demonstrate high textual and structural quality.

Thus, we can recommend business process modelers pay for the textual quality and proper activity labeling as much attention as they pay to the structural quality of business process scenarios. Having a business process model both structurally and textually sound will make it serve its initial purpose to communicate knowledge about ongoing or planned business processes.

Future work in this field may include the use of advanced NLP and machine learning methods and techniques to allow the automatic correction of poorly named activity labels to ensure the understandability of business process models. Also, more advanced metrics of structural analysis can be applied to continue the study of the relationship between the textual and structural quality of business process models.

6. References

- [1] M. Hammer, J. Champy, *Reengineering the Corporation: A Manifesto for Business Revolution*, Zondervan, 2009.
- [2] W. M. P. van der Aalst, Business process management: a comprehensive survey, in: *International Scholarly Research Notices*, volume 2013, Hindawi, 2013, pp. 1–37. doi:10.1155/2013/507984
- [3] P. Harmon, The State of Business Process Management, in: *The State of the BPM Market*, volume 2016, BPTrends, 2016, pp. 1–50.
- [4] M. Dumas, M. La Rosa, J. Mendling, H. A. Reijers, *Fundamentals of business process management*, Springer, Heidelberg, 2013. doi:10.1007/978-3-642-33143-5
- [5] J. Mendling, Managing structural and textual quality of business process models, *International Symposium on Data-Driven Process Discovery and Analysis*, Springer, Berlin, Heidelberg, 2012, pp. 100–111. doi:10.1007/978-3-642-40919-6_6
- [6] J. Mendling, H. A. Reijers, W. M. van der Aalst, Seven process modeling guidelines (7PMG), *Information and software technology* 52(2) (2010) 127–136. doi:10.1016/j.infsof.2009.08.004
- [7] H.G. Ceballos, V. Flores-Solorio, J. P. Garcia, A Probabilistic BPMN Normal Form to Model and Advise Human Activities, in: *International Workshop on Engineering Multi-Agent Systems*, Springer, Cham, 2015, pp. 51–69. doi:10.1007/978-3-319-26184-3_4

- [8] F. Corradini, F. Fornari, S. Gnesi, A. Polini, B. Re, Quality assessment strategy: Applying business process modelling understandability guidelines, University of Camerino, Italy, 2015. URL: https://openportal.isti.cnr.it/data/2017/380283/2017_380283.pdf
- [9] L. Henrik, J. Mendling, O. Günther, Learning from quality issues of BPMN models from industry, *IEEE software* 4(33) (2015) 26–33. doi:10.1109/MS.2015.81
- [10] W. Kbaier, S. A. Ghannouchi, Determining the threshold values of quality metrics in BPMN process models using data mining techniques, *Procedia Computer Science* 164 (2019) 113–119. doi:10.1016/j.procs.2019.12.161
- [11] W. M. C. da Silva, A. P. F. Araújo, M. T. Holanda, R. T. de Sousa Jr., A Method for Quality Assurance for Business Process Modeling with BPMN, in: *Developments and Advances in Intelligent Systems and Applications*, Springer, Cham, 2018, pp. 169–179. doi:10.1007/978-3-319-58965-7_12
- [12] A. L. da Costa, S. A. F. Salles, R. L. Carvalho, A. S. C. Morais, S. V. and Silva, BPMN and quality tools for process improvement: a case study. *Gepros: Gestão da Produção, Operações e Sistemas* 14(4) (2019) 156–175. doi:10.15675/gepros.v14i4.2308
- [13] P. Peggy, H. Schlieter, Process-based quality management in care: adding a quality perspective to pathway modelling, in: *OTM Confederated International Conferences “On the Move to Meaningful Internet Systems”*, Springer, Cham, 2019, pp. 385–403. doi:10.1007/978-3-030-33246-4_25
- [14] M. T. Gómez-López, J. M. Pérez-Álvarez, A. J. Varela-Vaca, R. M. Gasca, Guiding the creation of choreographed processes with multiple instances based on data models, in: *International Conference on Business Process Management*, Springer, Cham, 2016, pp. 239–251. doi:10.1007/978-3-319-58457-7_18
- [15] M. Kurz, F. Menge, Z. Misiak, Diagram Interchangeability in BPMN 2, 2014. URL: https://www.omg.org/oceb-2/documents/BPMN_Interchange.pdf
- [16] Business Process Model and Notation (BPMN), Version 2.0, 2011. URL: <https://www.omg.org/spec/BPMN/2.0/PDF/changebar>
- [17] J. Mendling, L. Sanchez-Gonzalez, F. Garcia, M. La Rosa, Thresholds for error probability measures of business process models, *Journal of Systems and Software* 85(5) (2012) 1188–1197. doi:10.1016/j.jss.2012.01.017
- [18] BPMN for research. URL: <https://github.com/camunda/bpmn-for-research>
- [19] Natural Language Toolkit. URL: <https://www.nltk.org/>
- [20] pandas.DataFrame.corr – pandas 1.5.0 documentation URL: <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.corr.html>