# Intelligent information technology for business process model quality and error volume assessment

Olha Yanholenko[†], Andrii Kopp[*,†], Mykhailo Godlevskyi[†] and Dmytro Orlovskyi[†]

*National Technical University "Kharkiv Polytechnic Institute", Kyrpychova str. 2, Kharkiv, 61002, Ukraine*

**Abstract**

In the Business Process Management domain, business process modeling is the key technique used to bridge the gap between Information Technology and Management domains, by introducing graphical diagrams understandable by both technical and business stakeholders. Nowadays BPMN (Business Process Model and Notation) models are the most widely-used workflow mapping diagrams, used to document, analyze, improve, and automate organizational activities. Therefore, business process models must be of high quality as ones of the most valuable artifacts for organizational and information systems design and development. This study considers the application of intelligence theory by introducing and using predicates to assess the business process modeling rules fulfillment by BPMN elements. The proposed intelligent information technology assumes BPMN model processing, detection of incorrect elements that violate modeling rules, impact assessment of different element types, quality and error volume measurement, as well as the textual explanations generation for the detected incorrect BPMN elements. The experiments with the large set of business process models are performed, obtained results are analyzed and discussed, conclusions are drawn and the further research is formulated.

**Keywords**

Business Process Modeling, Model Quality, Modeling Rules, Error Detection, Intelligence Theory

## 1. Introduction

Let us define a business process as the sequence of structured tasks or activities, which are executed in the organizational and (or) information system to achieve specific targets and create value for a customer [1]. Usually, business processes consist of events, activities of various types, and decision making steps to assure organizational flows and provide services or products [2].

Business processes are visually described using graphical models and diagrams, which help all of the involved stakeholders to understand, analyze, and improve depicted workflows [3]. Therefore, being the most valuable assets of the Business Process Management (BPM) methodology, business process models should be of high quality. Only high-quality business process models may be useful for the workflow design, analysis, improvement, and automation activities [3].

The BPM is an interdisciplinary methodology, which combines best practices from Information Technology (IT) and Business Management domains, helping organizations to identify, document, analyze, design, implement, and monitor workflows [4]. Business process modeling is used within BPM to visually depict organizational activities, events, and decision making points in the form of graphical diagrams [5]. BPM serves to provide clients with the high-quality products and services, as well as to find ways for the organizational efficiency improvement [6]. Thus, business process modeling techniques help to represent organization activities in a convenient for the further analysis

*ISW-2025: Intelligent Systems Workshop at 9th International Conference on Computational Linguistics and Intelligent Systems (CoLInS-2025), May 15–16, 2025, Kharkiv, Ukraine*

[*] Corresponding author.

[†] These authors contributed equally.

olha.yanholenko@khpi.edu.ua (O. Yanholenko); andrii.kopp@khpi.edu.ua (A. Kopp); mykhailo.hodlevskyi@khpi.edu.ua (M. Godlevskyi); dmytro.orlovskyi@khpi.edu.ua (D. Orlovskyi)

0000-0001-7755-1255 (O. Yanholenko); 0000-0002-3189-5623 (A. Kopp); 0000-0003-2872-0598 (M. Godlevskyi); 0000-0002-8261-2988 (D. Orlovskyi)

way. Such an important BPM technique is used for graphical description of business processes, their analysis and improvement [6].

Business Process Model and Notation (BPMN) offers a set of graphical elements used to describe various workflow aspects, such as events, activities, gateways (i.e. AND, OR, and XOR connectors determining decisions within business process scenarios and the corresponding execution logic), and flows. These BPMN elements were created to bridge the gap between IT and management domains, understandable by both technical and business users [6]. BPMN models assume workflows triggered by start events, finished by end events. Other activities and events, occurring within the workflow, are represented by tasks (or sub-processes) and intermediate events respectively [6].

Hence, business process modeling is a key BPM technique, which is used to graphically describe organizational workflows as interrelated models and, therefore, simplify business and IT providers' communication for the enhancement of the IT systems design, development, and maintenance [7].

Thus, this study aims to improve the quality and reduce the volume of errors, which may occur in business process models, by introducing the corresponding intelligent information technology.

The research object of this study, is the procedure of business process models' quality and error volume assessment. While the research subject is the intelligent information technology for business process models' quality and error volume assessment.

## 2. Related Work

In [8] Panayiotou et al., while studying modeling architecture development for Industry 4.0, stressed on the importance of business process modeling for BPM, since it assumes visual description of the organizational activities, events, and decisions via graphical models.

Avila et al. in their study on the modeling guidelines to create high-quality business process models [9] summarized, that such visual representation helps stakeholders to understand, capture, analyze, and improve enterprise business processes.

Business process modeling also allows to improve events monitoring, information control, and assess organizational efficiency [10].

In [11] Haj Ayech et al. discuss the critical importance of business process modeling techniques for the success of any BPM project or its particular goals. Authors of [11] proposed to extend the BPM lifecycle to improve the maintainability of BPMN models, by introducing structural measures based on activities, gateways, network and control-flow features:

- NOAJS – Number of Activities, Joins, and Splits;
- CNC – Coefficient of Network Connectivity;
- CFC – Control Flow Complexity.

Moreover, the authors of [11] suggested the use of modeling guidelines (rules), understandability and maintainability measures, as well as validation tools and their own measures. These measures were also applied by Fotoglou [12] et al. in their study on complexity clustering of BPMN models. NOAJS and CFC were considered by Kbaier et al. in [13] to determine respective threshold values for these measures using data mining techniques.

Threshold values for BPMN modeling measures were considered in [14] by Augusto et al. Also in [15] Corradini et al., through the study on consistent BPMN modeling, suggested the use of size and complexity measures to get insights on business process models quality from the architectural perspective.

The use of quality criteria for business process model assessment is considered in [16], authors Dai et al. suggested expert judgement and software engineering measures usage within the study on business process models refactoring and redundancy elimination. In [17] Pavlicek et al. considered the use of modeling rules and best practices to measure business process model quality.

Authors of [18] suggested the gathering structural measures of BPMN models to assess modeling rules conformance and the overall model quality.

Let us consider the rules for consistent business process modeling proposed in [19] and further elaborated in [20]:

- Start events should have one outgoing flow.
- Intermediate events should have one incoming and one outgoing flow.
- Boundary events should have one outgoing flow.
- End events should have one incoming flow.
- Activities (i.e. task or sub-process) should have one incoming and one outgoing flow.
- Gateways should have either one incoming and two outgoing flows (i.e. for splits), or two incoming and one outgoing flow (i.e. for joins).

## 3. Problem Statement

As the artificial intelligence research trend is growing and our basic goal is to formalize the human-centric analysis of BPMN models and their certain elements, let us apply the intelligence theory's method of comparator identification [21]. This indirect identification method is based on predicate logic, taking any data (signals) as input and providing binary value (0 or 1) on output [21]:

$$\mathrm{P}(x_1, x_2, \dots, x_n) = \mathrm{K}\big(y_1 = f(x_1), y_2 = f(x_2), \dots, y_n = f(x_n)\big) = t, \tag{1}$$

where:

- $x_1, x_2, \dots, x_n$ are the input signals;
- $y_1 = f(x_1), y_2 = f(x_2), \dots, y_n = f(x_n)$ are the internal signals;
- K is the comparator with $n$ inputs and one boolean output $t \in \{0,1\}$.

Let us consider extracted from the models BPMN elements as identifiable objects in terms of the intelligence theory [22]. Fig. 1 below outlines the idea of BPMN elements' comparator identification.

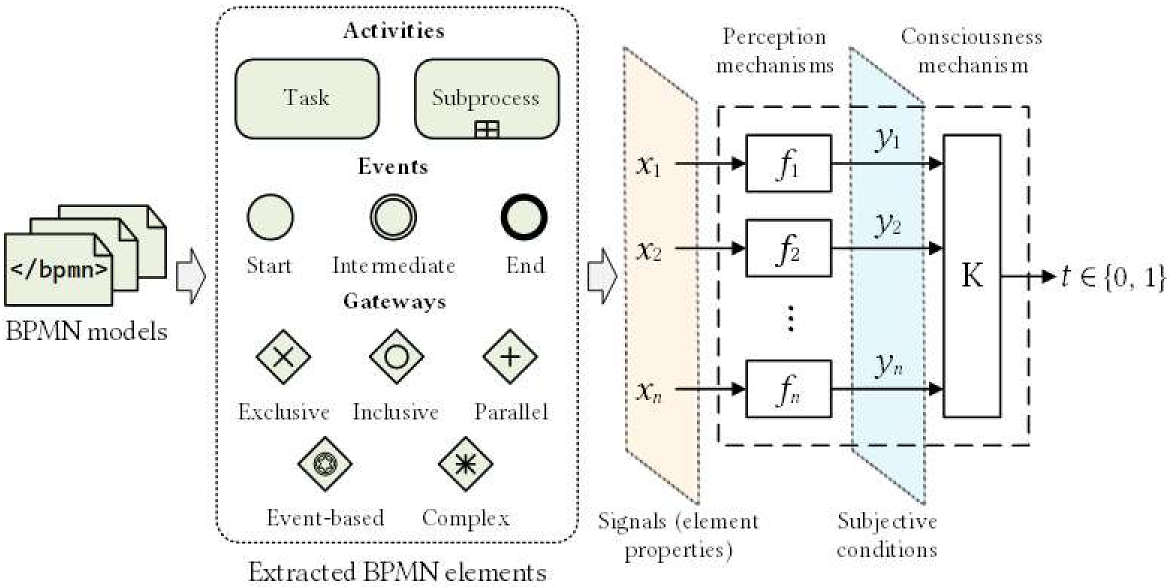


**Figure 1:** BPMN elements assessment using comparator identification.

Therefore, the problem includes the detection of incorrect elements that violate modelling rules using the comparator identification method, assessing the impact of different element types on the correctness of the BPMN models, measuring the quality and error volume, and the generation of textual explanations for the detected incorrect BPMN elements. The intelligent software should be developed to process large real-world collections of BPMN models.

## 4. Materials and Methods

Let us formally describe the set of business process element types as $U$ (i.e. start events, end events, intermediate events, boundary events, tasks, sub-processes, AND-gateways, OR-gateways, and XOR-gateways) [23], $|U| = 9$ (see Table 1).

**Table 1**
BPMN business process element types

| Business process element type | $i \in [1, \lvert U \rvert]$ | Brief description |
|---|---|---|
| OR-Gateway | 1 | Parallel execution of all actions that meet conditions |
| Sub-Process | 2 | A complex task that can be broken down into smaller tasks |
| Boundary Event | 3 | An event that may occur during the execution of a specific task |
| XOR-Gateway | 4 | Branching with a condition and a single course of action |
| Intermediate Event | 5 | An event that occurs between task executions |
| Task | 6 | A unit of work performed within the business process |
| End Event | 7 | An event that marks the business process completion |
| Start Event | 8 | An event that initiates the business process execution |
| AND-Gateway | 9 | Parallel execution of all actions without condition checking |

Typically, BPMN models are formalized using directed graphs [24]. Let us outline the following formal definition of a business process model:

$$BPrM = \langle N, P, SF, MF \rangle, \quad (2)$$

where:

- $N$ is the set of nodes that represents various business process elements, such as events, activities (i.e. tasks or sub-processes), and gateways, $N^k \in N$, $k \in [1, |N|]$;
- $P$ is the set of pools, each of which may contain nodes to define business process boundaries within a model, $P = \{P_q \subseteq N \big| P_1 \cap P_2 \cap \ldots \cap P_q = \emptyset, q \in [1, |P|]\}$;
- $SF \subseteq N \times N$ is the binary relationship that represents sequence flows between business process elements;
- $MF \subseteq N \times N$ is the binary relationship that represents message flows between business process elements.

Above, the definition of $P$ assumes the set of pools, which elements (i.e. single pools) are subsets of another set of nodes $N$ and that these subsets do not intersect – $P_1 \cap P_2 \cap \ldots \cap P_q = \emptyset$.

Let us introduce the tuple of business process model elements given in Table 1:

$$BPrM_{elements} = \langle E_{start}, E_{end}, E_{int}, E_b, T, SP, G_{and}, G_{or}, G_{xor} \rangle, \quad (3)$$

where:

- $E_{start}$ is the set of start events, $E_{start}^k \in E_{start}$, $k \in [1, |N|]$;
- $E_{end}$ is the set of end events, $E_{end}^k \in E_{end}$, $k \in [1, |N|]$;
- $E_{int}$ is the set of intermediate events, $E_{int}^k \in E_{int}$, $k \in [1, |N|]$;
- $E_b$ is the set of boundary events, $E_b^k \in E_b$, $k \in [1, |N|]$;
- $T$ is the set of tasks, $T^k \in T$, $k \in [1, |N|]$;
- $SP$ is the set of sub-processes, $SP^k \in SP$, $k \in [1, |N|]$;
- $G_{and}$ is the set of parallel (AND) gateways, $G_{and}^k \in G_{and}$, $k \in [1, |N|]$;

- $G_{or}$ is the set of inclusive (OR) gateways, $G_{or}^k \in G_{or}$, $k \in [1, |N|]$;
- $G_{xor}$ is the set of exclusive (XOR, as well as complex and event-based) gateways, $G_{xor}^k \in G_{xor}$, $k \in [1, |N|]$.

Let us apply the intelligence theory [22] and formulate the following predicates to assess business process elements' correspondence to the modeling rules [20]:

1. Start events:

$$R\left(E_{start}^k\right) = \begin{cases} 1, & if\ \mathrm{Out}\left(E_{start}^k\right) = 1, \\ 0, & else, \end{cases} \tag{4}$$

where $\mathrm{Out}\left(E_{start}^k\right)$ is the number of outgoing sequence flows of the start event $E_{start}^k \in E_{start}$.

2. End events:

$$R\left(E_{end}^k\right) = \begin{cases} 1, & if\ \mathrm{In}\left(E_{end}^k\right) = 1, \\ 0, & else, \end{cases} \tag{5}$$

where $\mathrm{In}\left(E_{end}^k\right)$ is the number of incoming sequence flows of the end event $E_{end}^k \in E_{end}$.

3. Intermediate events:

$$R\left(E_{int}^k\right) = \begin{cases} 1, & if\ \mathrm{In}\left(E_{int}^k\right) = 1\ \wedge \mathrm{Out}\left(E_{int}^k\right) = 1, \\ 0, & else, \end{cases} \tag{6}$$

where:

- $\mathrm{In}\left(E_{int}^k\right)$ is the number of incoming sequence flows of the intermediate event $E_{int}^k \in E_{int}$;
- $\mathrm{Out}\left(E_{int}^k\right)$ is the number of incoming outgoing flows of the intermediate event $E_{int}^k \in E_{int}$.

4. Boundary events:

$$R\left(E_{b}^k\right) = \begin{cases} 1, & if\ \mathrm{Out}\left(E_{b}^k\right) = 1, \\ 0, & else, \end{cases} \tag{7}$$

where $\mathrm{Out}\left(E_{b}^k\right)$ is the number of outgoing sequence flows of the boundary event $E_{b}^k \in E_{b}$.

5. Tasks:

$$R\left(T^k\right) = \begin{cases} 1, & if\ \mathrm{In}\left(T^k\right) = 1\ \wedge \mathrm{Out}\left(T^k\right) = 1, \\ 0, & else, \end{cases} \tag{8}$$

where:

- $\mathrm{In}\left(T^k\right)$ is the number of incoming sequence flows of the task $T^k \in T$;
- $\mathrm{Out}\left(T^k\right)$ is the number of incoming outgoing flows of the task $T^k \in T$.

6. Sub-processes:

$$R\left(SP^k\right) = \begin{cases} 1, & if\ \mathrm{In}\left(SP^k\right) = 1\ \wedge \mathrm{Out}\left(SP^k\right) = 1, \\ 0, & else, \end{cases} \tag{9}$$

where:

- $\mathrm{In}\left(SP^k\right)$ is the number of incoming sequence flows of the sub-process $SP^k \in SP$;
- $\mathrm{Out}\left(SP^k\right)$ is the number of incoming outgoing flows of the sub-process $SP^k \in SP$.

7. Gateways:

$$R(G_l^k) = \begin{cases} 1, & if\ [\mathrm{In}(G_l^k) = 1\ \wedge \mathrm{Out}(G_l^k) = 2] \vee [\mathrm{In}(G_l^k) = 2\ \wedge \mathrm{Out}(G_l^k) = 1], \\ 0, & else, \end{cases} \tag{10}$$

where:

- $\mathrm{In}(G_l^k)$ is the number of incoming sequence flows of the sub-process $G_l^k \in G_l$;
- $\mathrm{Out}(G_l^k)$ is the number of incoming outgoing flows of the sub-process $G_l^k \in G_l$;
- $l \in \{and, or, xor\}$ is the set of various gateway types.

Therefore, by processing BPMN files of business process models as XML documents [23], we may obtain $\mathrm{In}(N^k)$ and $\mathrm{Out}(N^k)$ numbers of incoming and outgoing sequence flows for each business process model element $N^k \in N$, as it is shown in Fig. 2.

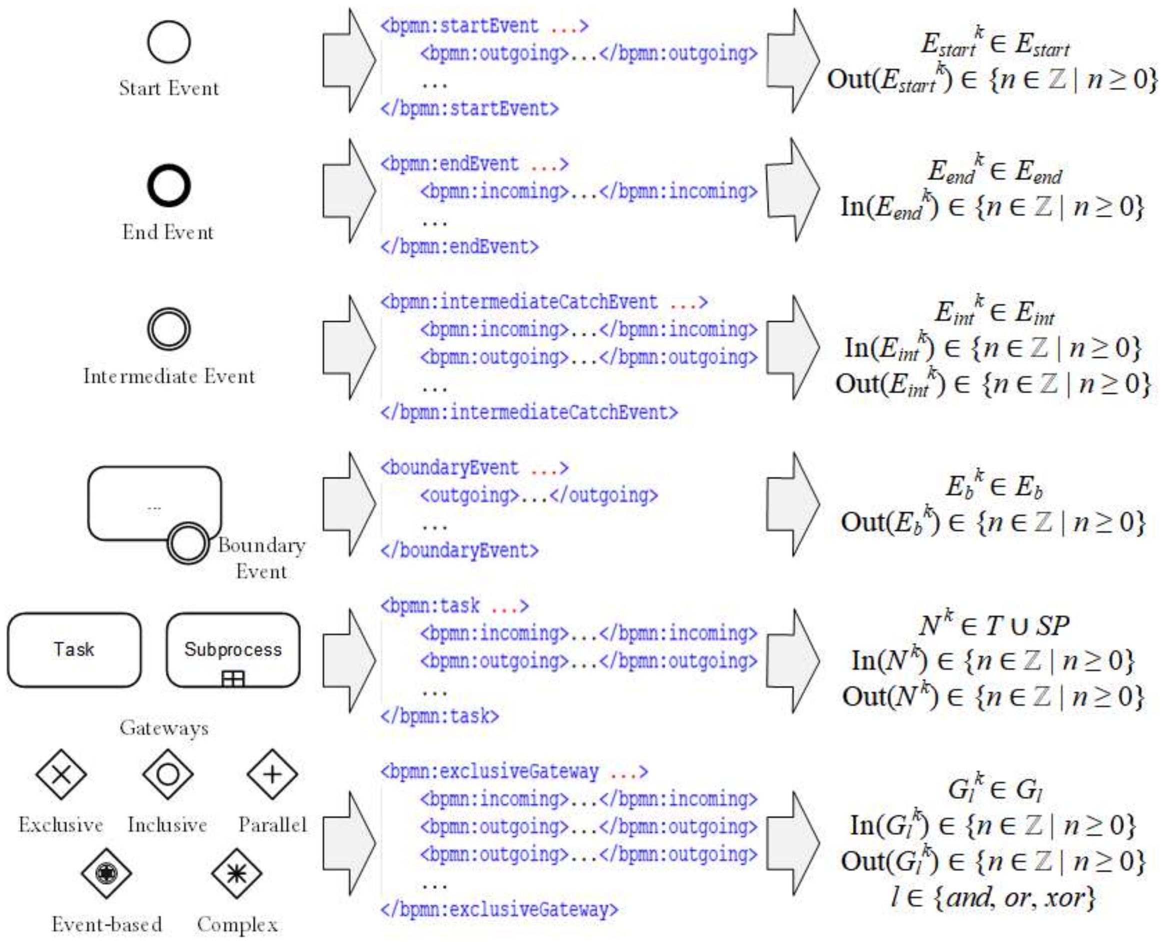


**Figure 2:** Extraction of business process elements of different types.

Having the collection of business process models $M$, the sets of business process elements are obtained for each processed BPMN model:

$$\forall m \in [1, |M|]: N_m = \bigcup_{N_{im} \in \{E_{start}^m, E_{end}^m, E_{int}^m, E_b^m, T^m, SP^m, G_{and}^m, G_{or}^m, G_{xor}^m\}} N_{im}. \tag{11}$$

where:

- $E_{start}^m$, $E_{end}^m$, $E_{int}^m$, $E_b^m$, $T^m$, $SP^m$, $G_{and}^m$, $G_{or}^m$, and $G_{xor}^m$ are the sub-sets of business process element types extracted from $m$-th BPMN model, $m \in [1, |M|]$;

- $N_{im}$ is the any sub-set of business process elements of a certain type $i \in [1, |U|]$, extracted from $m$-th BPMN model, $m \in [1, |M|]$.

Let us introduce the definition of a business process modeling error as the violation of a certain modeling rule (4) – (10) [20]:

$$\forall N_{im}^k \in N_{im} : \bar{R}(N_{im}^k) = 1 - R(N_{im}^k), m \in [1, |M|], i \in [1, |U|]. \quad (12)$$

where:

- $N_{im}$ is the sub-set of business process elements ($N_{im} \subseteq N_m$) of a certain type $i \in [1, |U|]$ extracted from $m$-th BPMN model;
- $N_{im}^k$ is the particular business process element $N_{im}^k \in N_{im}$ of a certain type $i \in [1, |U|]$ extracted from $m$-th BPMN model.

Hence, by processing the collection of business process models $M$, the following summarized measures can be found:

$$\forall i \in [1, |U|] : FE_i = \sum_{m=1}^{|M|} \sum_{N_i^k \in N_{im}} \bar{R}(N_{im}^k), \quad (13)$$

$$\forall i \in [1, |U|] : FR_i = \frac{FE_i}{|N_i|}, \quad (14)$$

where:

- $FE_i$ is the number of faulty business process elements (i.e. that do not follow the respective modeling rules (4) – (10)) of a certain type $i \in [1, |U|]$;
- $FR_i$ is the fault rate of business process elements of a certain type $i \in [1, |U|]$.

Let us apply Saaty's pairwise comparison method [25] to calculate weights of business process element types.

Obtained weights may describe the impact of each business process element type on the overall quality of business process models.

Thus, the following steps should be completed to calculate the weights of BPMN elements' impact on business process model quality:

1. For each pair $(i, j)$, $i, j \in [1, |U|]$ of business process element types calculate the ratio based on their fault rates:

$$A'_{ij} = \frac{FR_i}{FR_j}, i, j \in [1, |U|], j > i. \quad (15)$$

2. Order business process element types by the corresponding fault rates $FR_i$ in reverse order and map obtained fault rate ratios $A'_{ij}$ onto 1–9 scale [25] using logarithmic Min-Max scaling and round down transformation to obtain the elements placed above the main diagonal of the judgment matrix $\boldsymbol{A}$:

$$A_{ij} = \left\lfloor \alpha + \frac{\ln(A'_{ij}) - \ln\left(\min_{i,j \in [1,|U|]} \{A'_{ij}\}\right)}{\ln\left(\max_{i,j \in [1,|U|]} \{A'_{ij}\}\right) - \ln\left(\min_{i,j \in [1,|U|]} \{A'_{ij}\}\right)} \cdot (\beta - \alpha) \right\rfloor, i, j \in [1, |U|], j > i. \quad (16)$$

where $\alpha = 1$ and $\beta = 9$ according to Saaty's 1–9 scale.

3. Calculate the elements placed below the main diagonal of the judgment matrix $\boldsymbol{A}$:

$$A_{ji} = \frac{1}{A_{ij}}, i, j \in [1, |U|], j > i. \tag{17}$$

Diagonal elements of the judgment matrix $\boldsymbol{A}$ are equal to 1: $A_{ij} = 1$, $i, j \in [1, |U|]$, $i = j$.

4. Calculate weights of each business process element type $i \in [1, |U|]$ using the formula [25]:

$$W_i = \frac{\sqrt[|U|]{\prod_{j=1}^{|U|} A_{ij}}}{\sum_{i=1}^{|U|} \sqrt[|U|]{\prod_{j=1}^{|U|} A_{ij}}}, i \in [1, |U|]. \tag{18}$$

Let us define the business process model quality as the degree to which business process elements of different types fulfill the modeling rules (4) – (10) in accordance with the ISO 9001 definition [26]:

$$\forall m \in [1, |M|]: Q_m = \sum_{i=1}^{|U|} \left[ \frac{W_i}{|N_{im}|} \cdot \sum_{N_{im}^k \in N_{im}} R(N_{im}^k) \right]. \tag{19}$$

Hence, it becomes possible to assess the “volume” of business process modeling errors or, in other words, “errability” of a business process model:

$$\forall m \in [1, |M|]: V_m = \sum_{i=1}^{|U|} \left[ \frac{W_i}{|N_{im}|} \cdot \sum_{N_{im}^k \in N_{im}} \bar{R}(N_{im}^k) \right]. \tag{20}$$

Let us apply the Harrington’s scale and desirability function [27] to introduce business process models’ quality and errability levels (see Table 2).

**Table 2**
Assessment levels for business process models’ quality and errability

| Function value | Level of business process modeling rules fulfillment | Level of business process modeling errors volume |
|---|---|---|
| [0, 0.2) | Very bad | Very low |
| [0.2, 0.37) | Bad | Low |
| [0.37, 0.63) | Satisfactory | Moderate |
| [0.63, 0.8) | Good | High |
| [0.8, 1) | Very good | Very high |

The following linguistic variables can be obtained based on quality measures $Q_m$, $m \in [1, |M|]$:

$$\forall m \in [1, |M|]: Q_m^L = \begin{cases} Very\ bad, & Q_m < 0.2, \\ Bad, & 0.2 \leq Q_m < 0.37, \\ Satisfactory, & 0.37 \leq Q_m < 0.63, \\ Good, & 0.63 \leq Q_m < 0.8, \\ Very\ good & 0.8 \leq Q_m. \end{cases} \tag{21}$$

The following linguistic variables can be obtained based on errability measures $V_m$, $m \in [1, |M|]$:

$$\forall m \in [1, |M|]: V_m^L = \begin{cases} Very\ low, & V_m < 0.2, \\ Low, & 0.2 \leq V_m < 0.37, \\ Moderate, & 0.37 \leq V_m < 0.63, \\ High, & 0.63 \leq V_m < 0.8, \\ Very\ high & 0.8 \leq V_m. \end{cases} \tag{22}$$

Finally, the proposed procedure of BPMN models' analysis (Fig. 3) includes the following steps:

1. Extract business process elements and their properties from BPMN files by processing each model as the XML document (Fig. 2).
2. Assess extracted business process elements' correspondence to modeling rules (4) – (10).
3. Calculate summarized measures (13) – (14) by each type:

- the number of faulty business process elements;
- the fault rate of business process elements.

4. Calculate weights of business process element types using the Saaty's pairwise comparison method by filling the judgement matrix using corresponding fault rates (15) – (18).
5. Assess business process models' quality (19) and errors volume (20).
6. Provide linguistic variables for business process models' quality (21) and errors volume (22) to analyze obtained results.

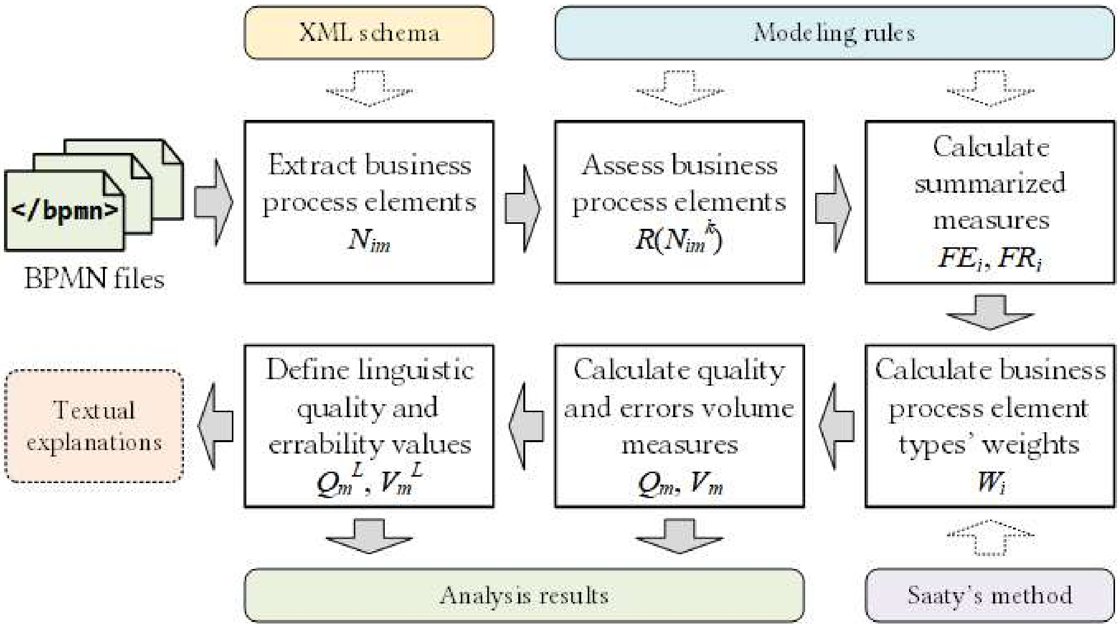

**Figure 3:** Proposed procedure of BPMN models' analysis.

For each analyzed business process model $m \in [1, |M|]$, the information about elements that do not fulfill the modeling requirements (i.e. that introduce errors) is provided as the set of explanations for various types of BPMN elements $\Theta$:

- $\exists x \in E^m_{start}: R(x) = 0$ having "$x$ with $\mathrm{Out}(x)$ outgoing flows found" for $\theta_1$;
- $\exists x \in E^m_{end}: R(x) = 0$ having "$x$ with $\mathrm{In}(x)$ incoming flows found" for $\theta_2$;
- $\exists x \in E^m_{int}: R(x) = 0$ having "$x$ with $\mathrm{In}(x)$ incoming, $\mathrm{Out}(x)$ outgoing flows found" for $\theta_3$;
- $\exists x \in E^m_b: R(x) = 0$ having "$x$ with $\mathrm{Out}(x)$ outgoing flows found" for $\theta_4$;
- $\exists x \in T \cup SP: R(x) = 0$ having "$x$ with $\mathrm{In}(x)$ incoming and $\mathrm{Out}(x)$ outgoing flows found" for $\theta_5$;
- $\exists x \in G^m_l: R(x) = 0, l \in \{and, or, xor\}$ having "$x$ with $\mathrm{In}(x)$ incoming and $\mathrm{Out}(x)$ outgoing flows found" for $\theta_6$.

Therefore, each business process model $m \in [1, |M|]$ containing incorrect elements is provided with the sub-set of respective explanations $\Theta_m \subseteq \Theta$.

## 5. Results and Discussion

The software implementation of the proposed approach is developed using Python programming language. Its activity diagram is demonstrated in Fig. 4.

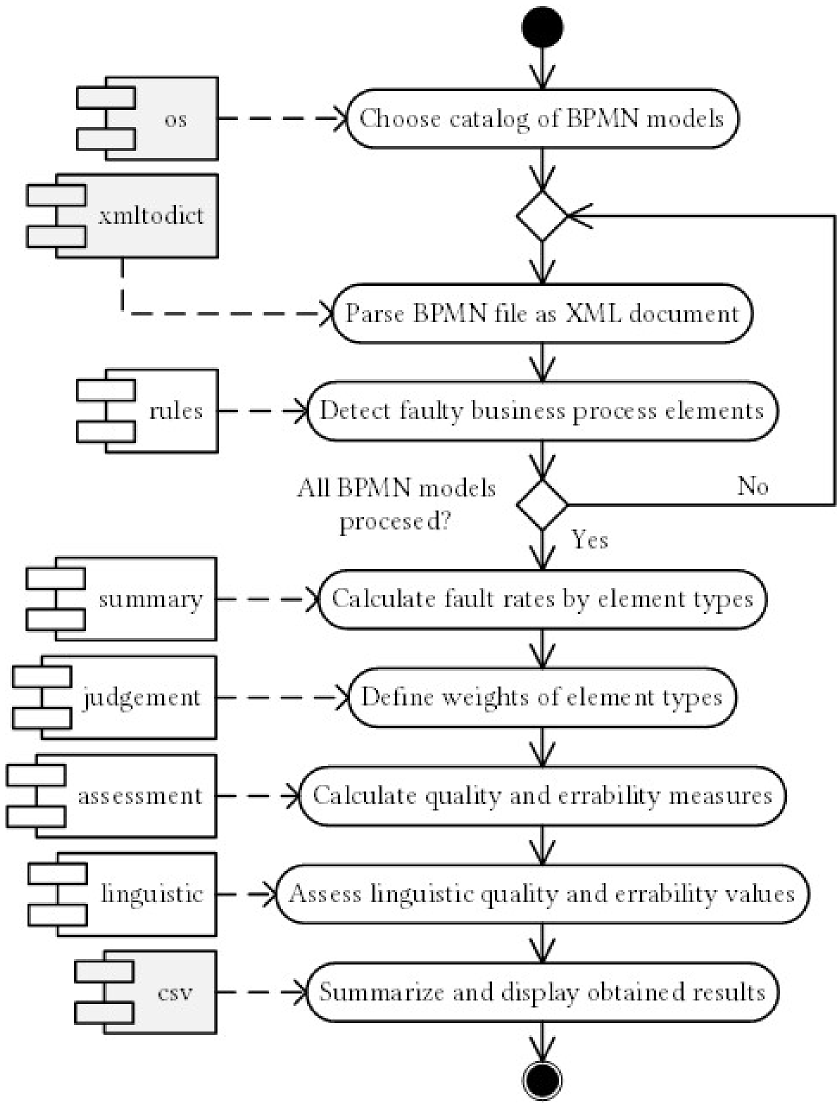


**Figure 4:** Software components for business process models' analysis.

Gray components on the activity diagram (see Fig. 4) demonstrate used Python modules:

- "os" [28] is used to walk over the catalog of BPMN models;
- "xmltodict" [29] is used to process BPMN model files as XML documents;
- "csv" [30] is used to prepare output results as CSV (Comma-Separated Values) documents.

Others are developed components used to implement particular method's steps (see Fig. 4):

- "rules" is used to detect incorrect business process elements using (4) – (10);
- "summary" is used to calculate measures (13) – (14) for analyzed BPMN models;
- "judgement" is used to find weights of BPMN element types using Saaty's method (15) – (18);
- "assessment" is used to calculate quality and errability measures (19) – (20) for BPMN models;
- "linguistic" is used to define linguistic values for quality and errability measures (21) – (22).

Experiments below are performed using the Camunda's repository of 3729 BPMN files [31], out of which 3722 were successfully processed and 7 remained with processing errors.

Fig. 5 demonstrates the distribution of extracted BPMN elements by different types.

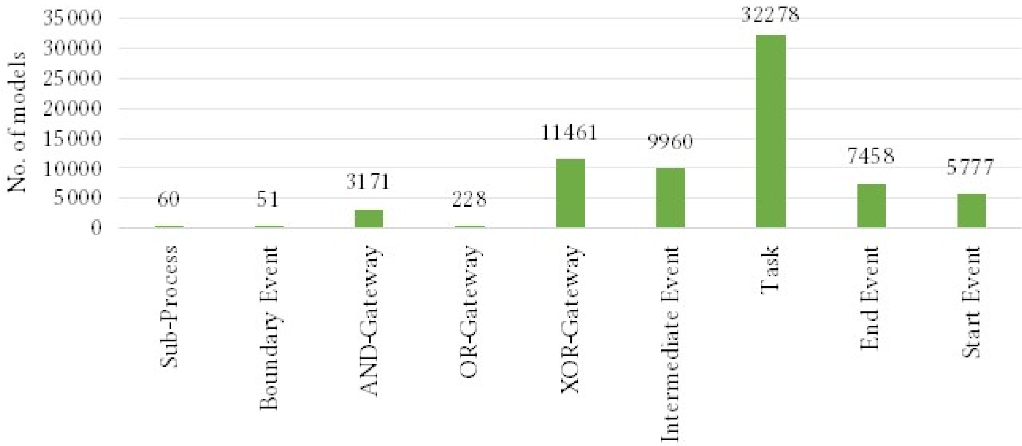


**Figure 5:** Distribution of business process element types.

Fig. 6 demonstrates the error frequency (13) of extracted BPMN elements by different types.

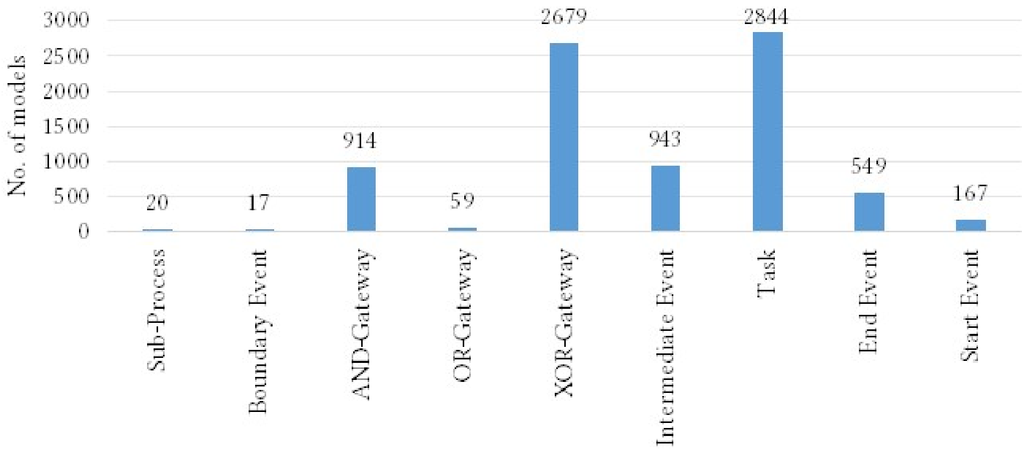


**Figure 6:** Error frequency by business process element types.

Fig. 7 demonstrates the fault rate (14) of extracted BPMN elements by different types.

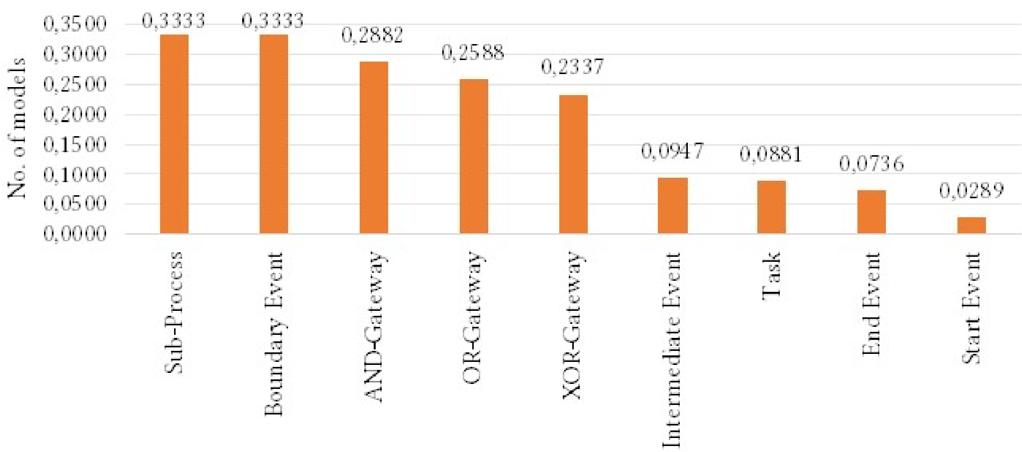


**Figure 7:** Fault rate by business process element types.

Using fault rate measures and the Saaty's pairwise comparison method (15) – (18), the following judgement matrix was obtained:

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 1 & 1 & 2 & 5 & 5 & 5 & 9 \\ 1 & 1 & 1 & 1 & 2 & 5 & 5 & 5 & 9 \\ 1 & 1 & 1 & 1 & 1 & 4 & 4 & 5 & 8 \\ 1 & 1 & 1 & 1 & 1 & 4 & 4 & 5 & 8 \\ 1/2 & 1/2 & 1 & 1 & 1 & 3 & 4 & 4 & 7 \\ 1/5 & 1/5 & 1/4 & 1/4 & 1/3 & 1 & 1 & 1 & 4 \\ 1/5 & 1/5 & 1/4 & 1/4 & 1/4 & 1 & 1 & 1 & 4 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/4 & 1 & 1 & 1 & 4 \\ 1/9 & 1/9 & 1/8 & 1/8 & 1/7 & 1/4 & 1/4 & 1/4 & 1 \end{pmatrix}, \tag{23}$$

And the following weights were calculated (Fig. 8) to assume the impact of BPMN element types on the models' quality.

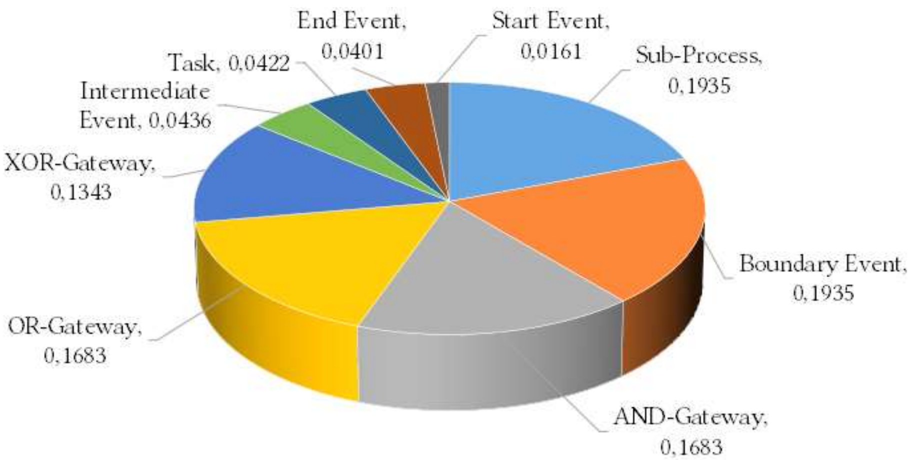


**Figure 8:** Weights of BPMN element types calculated using Saaty's pairwise comparison method.

Having the maximum eigenvalue of $\lambda_{max} = 9.13$, the Consistency Index (CI) is found [25]:

$$CI = \frac{\lambda_{max} - |U|}{|U| - 1} = 0.02. \tag{24}$$

The Consistency Rate (CR) is less than 0.1 (10%), hence, the judgment matrix (23) is considered to be consistent:

$$CR = \frac{CI}{RI} = 0.01 < 0.1, \tag{25}$$

where $RI = 1.45$ is the Random Consistency Index (RI) suggested by Saaty for $|U| = 9$ [25].

Fig. 9 demonstrates the distribution of analyzed BPMN models by quality measure (19) ranges.

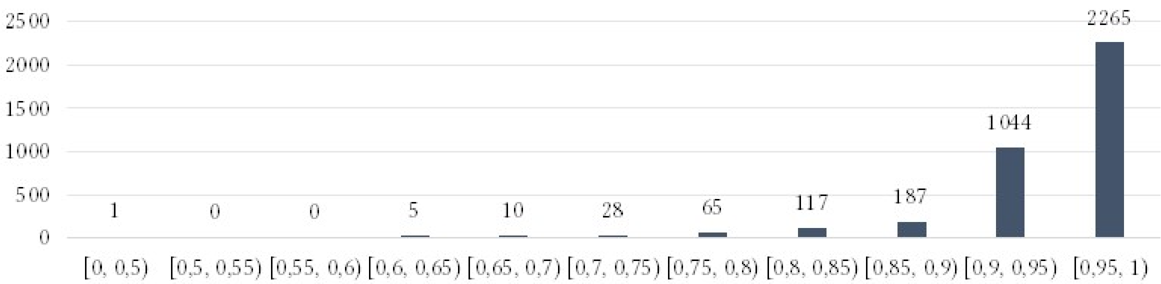


**Figure 9:** Distribution of business process models by quality value ranges.

Fig. 10 demonstrates the distribution of analyzed BPMN models by errability measure (20) ranges.

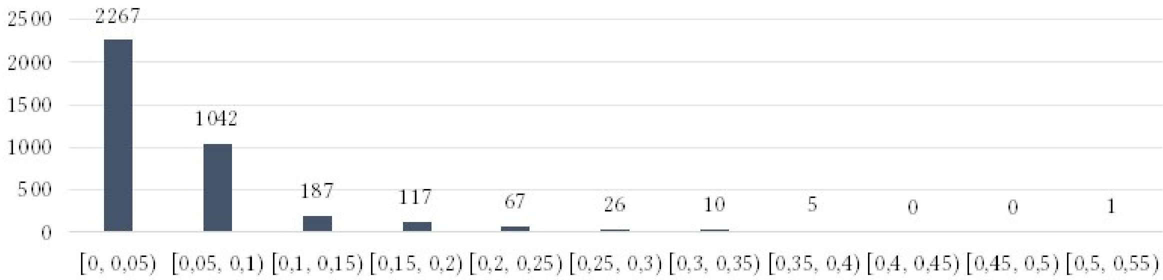


**Figure 10:** Distribution of business process models by errability (i.e. error volume) value ranges.

The exploratory analysis results of quality and errability measures are demonstrated in Table 3.

**Table 3**
Exploratory analysis results for quality and errability measures

| Value | Min | Q1 (25 percentile) | Median | Q3 (75 percentile) | Max |
|---|---|---|---|---|---|
| Quality | 0,49 | 0,94 | 0,96 | 0,99 | 1,00 |
| Errability | 0,00 | 0,01 | 0,04 | 0,06 | 0,50 |

Fig. 11 demonstrates the distribution of analyzed BPMN models using Harrington scale linguistic values for quality (21) and errability (22) measures.

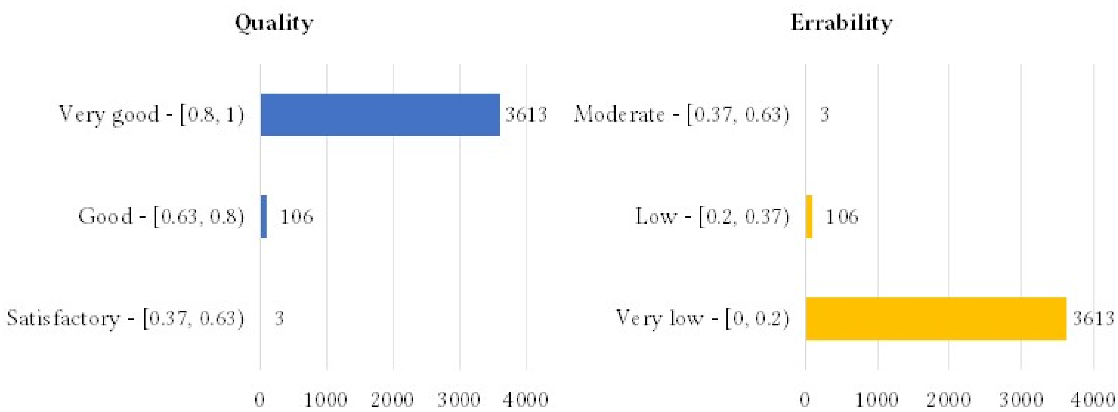


**Figure 11:** Distributions of business process models by quality and errability (i.e. error volume) levels.

Let us summarize the obtained results of BPMN models' processing and analysis:

- "Task" is the most widely-used element occurred 32278 times (Fig. 5) in the processed BPMN models;
- "Task" and "XOR-Gateway" are the most error-prone elements (Fig. 6) – faulty tasks occurred in the dataset 2844 times, while faulty XOR-Gateways occurred 2679 times;
- however, "Sub-Process" and "Boundary Event" are the most critical elements according to the fault rate measures, followed by gateways of all types – "AND-Gateway", "OR-Gateway", and "XOR-Gateway" (Fig. 7);
- much less critical BPMN elements according to the fault rate measures are "Intermediate Event", "Task", "End Event" (Fig. 7);
- much smaller impact on business process model quality show "Start Event" elements (Fig. 7);

- in general, sub-processes, boundary events, and gateways of all types (AND, OR, XOR) make over 85% impact of BPMN models' quality according to the weights (Fig. 8) calculated using Saaty's pairwise comparison method;
- the distribution by quality measure ranges demonstrates the most of BPMN models (3309 or 89%) have $Q_m > 0.9$, $m \in [1, |M|]$ (Fig. 9);
- while the distribution by errability (i.e. error volume) measure ranges demonstrates the same amount of BPMN models (3309 or 89%) have $V_m < 0.1$, $m \in [1, |M|]$ (Fig. 10);
- the exploratory analysis results demonstrate minimum quality value of 0.49 and maximum errability values of 0.50 (Table 3);
- the other interesting exploratory analysis results demonstrate (Table 3):
    a. 25% of BPMN models have $Q_m < 0.94$ and $V_m < 0.01$, $m \in [1, |M|]$;
    b. 50% of BPMN models have $Q_m < 0.96$ and $V_m < 0.04$, $m \in [1, |M|]$;
    c. 75% of BPMN models have $Q_m < 0.99$ and $V_m < 0.06$, $m \in [1, |M|]$;
- Harrington scale-based estimates of business process models' quality and errability (Fig. 11) demonstrate:
    a. 3613 (97.07%) BPMN models are of "Very good" quality – [0.8, 1) and, respectively, of "Very low" errability – [0, 0.2);
    b. 106 (2.85%) BPMN models are of "Good" quality – [0.63, 0.8) and, respectively, of "Low" errability – [0.2, 0.37);
    c. and only 3 (0.08%) BPMN models are of "Satisfactory" quality – [0.37, 0.63) and, respectively, of "Moderate" errability – [0.37, 0.63).

Finally, let us consider the example BPMN model from the Camunda's repository [31]. The one of BPMN models' version "warenversand_-_english_00f5b29d34c8482d9ec476f554c6dad0.bpmn" of the goods dispatch business process is shown in Fig. 12.

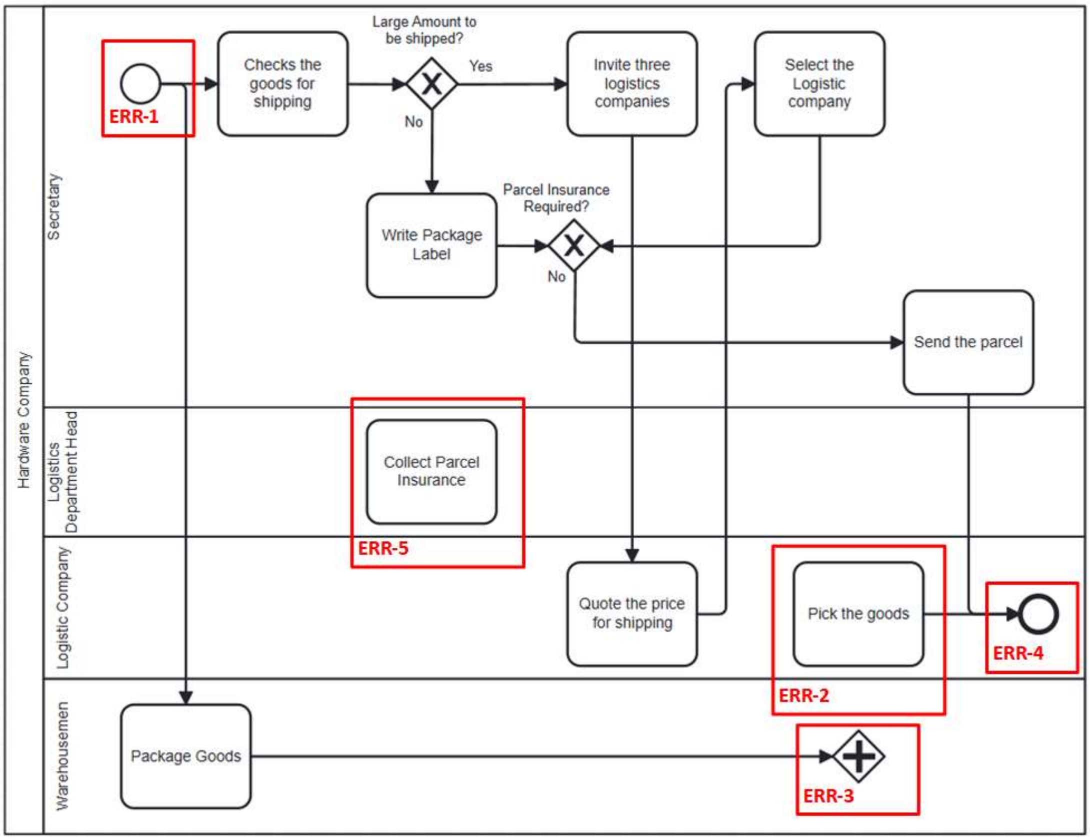


**Figure 12:** Goods dispatch business process model from the Camunda's repository.

The following explanations were generated for the detected incorrect business process elements highlighted in Fig. 12:

- ERR-1: Start Event with 2 outgoing flows found.
- ERR-2: Task with 0 incoming and 1 outgoing flows found.
- ERR-3: Parallel Gateway with 1 incoming and 0 outgoing flows found.
- ERR-4: End Event with 2 incoming flows found.
- ERR-5: Task with 0 incoming and 0 outgoing flows found.

Moreover, the BPMN model in Fig. 12 demonstrates inconsistent use of lanes and pools, e.g. the "Logistic Company" is modeled using lane, while the counterparties are expected to be modeled as pools, communicated via message flows [23].

Table 4 below outlines the following measures of the goods dispatch BPMN model (see Fig. 12).

**Table 4**
Quality measures of the sample business process model

| Measure | TNE | NCE | NEE | CER | IER | Quality | Quality level | Errability | Errability level |
|---|---|---|---|---|---|---|---|---|---|
| Value | 14 | 9 | 5 | 0.64 | 0.36 | 0.77 | Good | 0.23 | Low |

Here in Table 4, TNE is the total number of elements, NCE is the number of correct elements, NEE is the number of incorrect elements, CER is the correct elements rate

$$CER = \frac{NCE}{TNE}, \tag{26}$$

and IER is the incorrect elements rate:

$$IER = \frac{NEE}{TNE}. \tag{27}$$

Due to the introduced weights that differentiate the impact of various BPMN element types on the business process models' correctness, the sample model's (Fig. 12) quality and errability measures are estimated at 0.77 and 0.23 respectively, while the more straightforward measures CER and ERR are demonstrating values of 0.64 and 0.36.

## 6. Conclusion and Future Work

This paper addressed the problem of business process model quality improvement and error volume reducing using intelligent information technology. In particular, the intelligence theory's method of comparator identification based on predicate logic was applied to assess business process modeling rules fulfillment by different BPMN elements. The BPMN notation is considered as the standard for organization and information system workflow design, essential for BPM and IT domains due to its powerful capabilities of business process documenting, analyzing, improving, and automating.

The proposed approach and intelligent software, implemented as the Python tool, allow to:

- extract BPMN elements and detect incorrect ones that violate modeling rules;
- assess impact of different BPMN element types on business process model correctness;
- assess business process model quality and errability (i.e. error volume) as numeric measures and linguistic variables that correspond to different levels;
- provide textual explanations for the detected incorrect BPMN elements to assist responsible for business process modeling users.

The results of experiments with 3729 Camunda's BPMN models [31], generated and summarized by the developed intelligent tool, were analyzed and discussed. Therefore, it is possible to assume the ability of considered intelligent information technology to efficiently handle large enterprise business process model repositories, and further apply it to the analysis of real-world BPMN models.

Future work in this field includes the use of advanced intelligent techniques, such as fuzzy logic and comparator networks, to improve "thinking mechanisms" of the proposed technology. Deeper analysis of BPMN elements (e.g. various events' behavior, data stores, pools and lanes, message flows), as well as of business process semantics, should be conducted in the future work.

## Declaration on Generative AI

The authors have not employed any Generative AI tools.

## References


[1] S. Guerreiro, A. Vasconcelos, P. Sousa, Business Process Design, In: Enterprise Architecture and Cartography: From Practice to Theory; From Representation to Design, Cham, Springer International Publishing, 2022, pp. 185–210. doi:10.1007/978-3-030-96264-7_8.

[2] G. Rivera Lazo, R. Ñanculef, Multi-attribute Transformers for Sequence Prediction in Business Process Management, In: Discovery Science: 25th International Conference, DS 2022, Cham, Springer Nature Switzerland, 2022, pp. 184–194. doi:10.1007/978-3-031-18840-4_14.

[3] F. Vernadat, Enterprise modelling: Research review and outlook, Computers in Industry 122 (2020) 103265. doi:10.1016/j.compind.2020.103265.

[4] A. Gębczyńska, K. Vladova, Comparative analysis of selected process maturity assessment models applied in the public sector, Business Process Management Journal 29(3) (2023) 911–928. doi:10.1108/BPMJ-09-2022-0420.

[5] I. Beerepoot et al., The biggest business process management problems to solve before we die, Computers in Industry 146 (2023) 103837. doi:10.1016/j.compind.2022.103837.

[6] H. A. Reijers, Business Process Management: The evolution of a discipline, Computers in Industry 126 (2021) 103404. doi:10.1016/j.compind.2021.103404.

[7] A. Correia, F. Brito e Abreu, Enhancing the correctness of BPMN models, In: Information Resources Management Association (Ed.), Hershey: IGI Global, 2020, pp. 373–394. doi:110.4018/978-1-5225-9615-8.ch017.

[8] N. A. Panayiotou, K. E. Stergiou, V. P. Stavrou, Development of a Modeling Architecture Incorporating the Industry 4.0 View for a Company in the Gas Sector, In: Advances in Production Management Systems Towards Smart Production Management Systems: IFIP WG 5.7 International Conference, APMS 2019, Springer International Publishing, 2019, pp. 397–404. doi:10.1007/978-3-030-29996-5_46.

[9] D. T. Avila et al., An experiment to analyze the use of process modeling guidelines to create high-quality process models, In: Database and Expert Systems Applications: 30th International Conference, DEXA 2019, Springer International Publishing, 2019, pp. 129–139. doi:10.1007/978-3-030-27618-8_10.

[10] F. Bellini et al., Digital Identity: A Case Study of the ProCIDA Project, In: Exploring Digital Ecosystems: Organizational and Human Challenges, Springer International Publishing, 2020, pp. 315–327. doi:10.1007/978-3-030-23665-6_23.

[11] B. Haj Ayech, S. A. Ghannouchi, E. A. El Hadj Amor, Extension of the BPM lifecycle to promote the maintainability of BPMN models, Procedia Computer Science 181 (2021) 852–860. doi:10.1016/j.procs.2021.01.239.

[12] C. Fotoglou et al., Complexity clustering of BPMN models: initial experiments with the K-means algorithm, In: Decision Support Systems X: Cognitive Decision Support Systems and Technologies: 6th International Conference on Decision Support System Technology, ICDSST 2020, Springer International Publishing, pp. 57–69. doi:10.1007/978-3-030-46224-6_5.

[13] 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.

[14] A, Augusto, J. Mendling, M. Vidgof, B. Wurm, The connection between process complexity of event sequences and models discovered by process mining, Information Sciences 598 (2022) 196–215. doi:10.1016/j.ins.2022.03.072.

[15] F. Corradini, A. Polini, B. Re, L. Rossi, F. Tiezzi, Consistent modelling of hierarchical BPMN collaborations, Business Process Management Journal 28(2) (2022) 442–460. doi:10.1108/BPMJ-07-2021-0485.

[16] F. Dai et al., Refactor Business Process Models with Redundancy Elimination, In: Advances in Parallel & Distributed Processing, and Applications: Proceedings from PDPTA'20, CSC'20, MSV'20, and GCC'20 2021, Springer International Publishing, 2021, pp. 509–524. doi:10.1007/978-3-030-69984-0_37.

[17] J. Pavlicek, P. Pavlickova, A. Pokorná, M. Brnka, Business Process Models and Eye Tracking System for BPMN Evaluation-Usability Study, in: MOBA 2023: Model-Driven Organizational and Business Agility, Springer, Cham, 2023, pp. 53–64. doi:10.1007/978-3-031-45010-5_5.

[18] F. Corradini et al., Correctness checking for BPMN collaborations with sub-processes, Journal of Systems and Software 166 (2020) 110594. doi:10.1016/j.jss.2020.110594.

[19] A. Kopp, D. Orlovskyi, Towards Intelligent Technology for Error Detection and Quality Evaluation of Business Process Models, in: IntelITSIS'2023: 4th International Workshop on Intelligent Information Technologies and Systems of Information Security, Khmelnytskyi, Ukraine, 2023, pp. 1–14. URL: https://ceur-ws.org/Vol3373/keynote1.pdf.

[20] O. Yanholenko, A. Kopp, D. Orlovskyi, U. Litvinova, Towards algorithmic and software solution for business process model analysis and correction, in: MoDaST-2024: 6th International Workshop on Modern Data Science Technologies, Lviv-Shatsk, Ukraine, 2024, pp. 86-103. URL: https://ceur-ws.org/Vol-3723/paper6.pdf.

[21] V. Sokol, M. Tkachuk, M. Godlevskiy, M. Bilova, D. Studenikin, An Approach to ICT Professionals' Skills Assessment based on European e-Competence Framework, in: ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer, Kharkiv, Ukraine, 2020, pp. 677-692. URL: https://ceur-ws.org/Vol-2732/20200677.pdf.

[22] Z. V. Dudar, V. A. Chikina, S. Y. Shabanov-Kushnarenko, The modelling of some intelligence functions, Radioelectronics and informatics 3 (2003) 166–172.

[23] Business Process Model and Notation (BPMN). Version 2.0. URL: https://www.omg.org/spec/BPMN/2.0/PDF.

[24] F. Corradini et al., BPMN 2.0 OR-Join Semantics: Global and local characterisation, Information Systems 105 (2022) 101934. doi:10.1016/j.is.2021.101934.

[25] H. Liang, J. Ren, S. Gao, L. Dong, Z. Gao, Comparison of different multicriteria decision-making methodologies for sustainability decision making, Academic Press, 2023, pp. 233–271. doi:10.1016/B978-0-323-99514-6.00001-7.

[26] R. Tricker, ISO 9001:2015 for Small Business, Taylor & Francis, 2016.

[27] A. Voronin, Multi-Criteria Decision Making for the Management of Complex Systems, IGI Global, 2017. doi:10.4018/978-1-5225-2509-7.

[28] os — Miscellaneous operating system interfaces. URL: https://docs.python.org/3/library/os.html.

[29] xmltodict. URL: https://pypi.org/project/xmltodict/.

[30] CSV File Reading and Writing. URL: https://docs.python.org/3/library/csv.html.

[31] BPMN for research. URL: https://github.com/camunda/bpmn-for-research/.