

UDC 004.9

*Otreshko Ivan, Postgraduate Student,
Luchnoi Roman, Postgraduate Student,
Litvinova Uliya, Ph.D.,
Kopp Andrii, Ph.D., Associate Professor
National Technical University «Kharkiv Polytechnic Institute»*

DATA-DRIVEN APPROACH TO SYNTACTIC GATEWAY QUALITY ANALYSIS OF BUSINESS PROCESS MODELS

Quality of BPMN models determines the accuracy of their interpretation and subsequent use in IT systems. Syntactic quality is of main importance, as errors at this level lead to incorrect process execution logic. Gateways, which determine the branching and merging of sequence flows, play a crucial role. A mismatch between the split and join gateways leads to structural defects in the BPMN model. The paper [1] investigates the ability of language models to analyze BPMN, confirming the relevance of automated approaches. Work [2] proposes intelligent methods for evaluating model quality. However, the problem of formalizing the syntactic gateway quality using BPMN models' data remains relevant. Thus, the aim of this study is to develop a data-driven approach to analyzing the syntactic gateway quality of BPMN models.

The syntactic gateway quality is defined as the average proportion of gateways in the BPMN model that correctly implement business process branching (splits) or merging joins) using the following formula:

$$Q_{syn}^G = \frac{1}{|G|} \left(\sum_{g \in G} \mathbf{1}_S(g) + \sum_{g \in G} \mathbf{1}_J(g) \right),$$

$$\mathbf{1}_S(g) = \begin{cases} 1 & |in(g)| = 1 \wedge |out(g)| > 1, \\ 0 & \text{else} \end{cases},$$

$$\mathbf{1}_J(g) = \begin{cases} 1 & |in(g)| > 1 \wedge |out(g)| = 1, \\ 0 & \text{else} \end{cases},$$

where:

$in(g)$ and $out(g)$ – are the sets of incoming and outgoing sequence flows of a gateway $g \in G$;

G – is the set of gateways defined in the BPMN model;

$\mathbf{1}_S(g)$ – is the indicator function, which takes 1 for split gateways $S \subseteq G$ and 0 for all the other gateway elements;

$\mathbf{1}_J(g)$ – is the indicator function, which takes 1 for join gateways $J \subseteq G$ and 0 for all the other gateway elements.

If the gateways are defined incorrectly (i.e. neither splits nor joins), the Q_{syn}^G value decreases. Thus, the metric Q_{syn}^G reflects the structural consistency level of a business process model.

The proposed data-driven approach is demonstrated in Fig. 1.

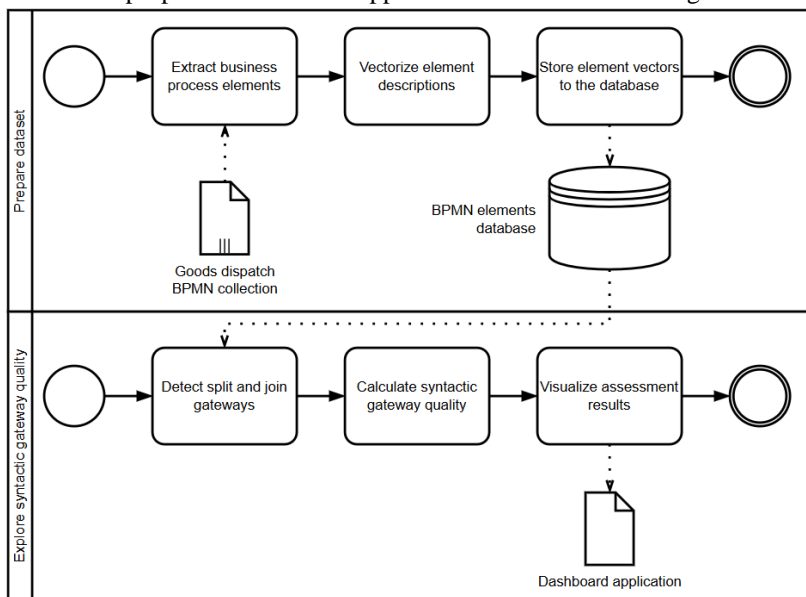


Fig. 1. Data-driven approach to syntactic gateway quality analysis

BPMN elements are extracted and converted into a vector representation, which is then stored in a database. The syntactic quality is calculated as an aggregated metric across all model gateways, reflecting the level of structural consistency. Then, analysis results are visualized using a dashboard and can be used for the further training of an intelligent BPMN evaluation system.

References:

1. Lauer C., Pfeiffer P., Rombach A., Mehdiyev N. Assessing the Business Process Modeling Competences of Large Language Models [Electronic resource]. – Access: <https://arxiv.org/abs/2601.21787>
2. Yanholenko O., Kopp A., Godlevskiy M., Orlovskiy D., Khatsko N. Towards information technology for BPMN models quality assessment based on intelligence theory [Electronic resource]. – Access: <https://ceur-ws.org/Vol-4004/paper7.pdf>