ISSN 2079-0023 (print), ISSN 2410-2857 (online)

UDC 004.94 **DOI: 10.20998/2079-0023.2022.01.10**

A. M. KOPP, D. L. ORLOVSKYI, I. V. LIUTENKO

A SOFTWARE SOLUTION TO WORK WITH A DATABASE OF BUSINESS PROCESS MODELS AND ANALYZE THEIR STRUCTURAL MEASURES

Business process modeling is one of the main tools of the BPM (Business Process Management) concept. With the help of business process modeling, business scenarios can be presented in the form of graphical models that can be easily understood by both information technology (IT) professionals and non-IT professionals – business analysts, software customers, department heads, top managers, and other stakeholders interested in business process improvement. Business process improvement is usually done through the automation of activities, which were identified as “bottlenecks” after analysis. However, it is possible to analyze a business process model only if it is clear and correct in terms of compliance with both the notation used and the real business process it depicts. This paper considers the analysis of BPMN (Business Process Model and Notation) business process model structural measures. It is assumed that business process models, which by their structural features violate rules of business process modeling, are neither understandable nor suitable for further work with them, which also can lead to various errors occurring at the stage of business process analysis, as well as at the stage of its improvement and implementation of proposed changes, i.e., during development, testing and maintenance of distinct software components, information system modules or BPM-system scenarios that ensure business process execution. Therefore, in this paper, we propose to identify the main elements of BPMN business process models and their structural measures that affect models’ understandability and maintainability and could be sources of errors. Considering selected measures, it is proposed to calculate respective values for a large collection of BPMN business process models, and then study compliance with theoretical business process modeling guidelines on practice when real business process models are designed. In order to provide efficient storage and processing of a large collection of BPMN business process models data, there were developed a database, and a software component. Results of analysis of BPMN business process model structural measures obtained using developed database and software component are demonstrated and discussed. The conclusion is made, as well as future research directions in this field are formulated.

Keywords: business process model, modeling notation, structural measures, modeling guidelines.

А. М. КОПП, Д. Л. ОРЛОВСЬКИЙ, І. В. ЛЮТЕНКО

ПРОГРАМНЕ РІШЕННЯ ДЛЯ РОБОТИ З БАЗОЮ ДАНИХ МОДЕЛЕЙ БІЗНЕС-ПРОЦЕСІВ ТА АНАЛІЗУ ЇХ СТРУКТУРНИХ ХАРАКТЕРИСТИК

Моделювання бізнес-процесів є одним з основних інструментів концепції управління бізнес-процесами BPM (Business Process Management). За допомогою моделювання бізнес-процесів, сценарії діяльності підприємств та організацій можуть бути подані у формі графічних моделей, які можуть бути достатньо легко зрозумілі як спеціалістами з інформаційних технологій (IT), так і не IT-спеціалістами – бізнес-аналітиками, замовниками програмного забезпечення, керівниками відділів, топ-менеджерами та іншими стейкголдерами, зацікавленими в удосконаленні бізнес-процесів. Удосконалення бізнес-процесів зазвичай відбувається за рахунок автоматизації робіт, які в результаті аналізу були визначені як «вузькі місця». Проте, здійснити аналіз моделі бізнес-процесу представляється можливим лише за умови її зрозумілості та коректності з точки зору відповідності як нотації, що використовується, так і реальному бізнес-процесу, який вона зображує. В даній роботі розглядається аналіз структурних характеристик моделей бізнес-процесів, представлених за допомогою нотації BPMN (Business Process Model and Notation). Передбачається, що моделі бізнес-процесів, які за своїми структурними властивостями порушують правила моделювання бізнес-процесів, не є зрозумілими та придатними до подальшої роботи з ними, через що можуть виникати різні помилки як на етапі аналізу бізнес-процесу, так і на етапі його вдосконалення і впровадження запропонованих змін, зокрема, під час розробки, тестування та підтримки окремих програмних компонентів, модулів інформаційних систем або сценаріїв у BPM-системах, що забезпечують виконання бізнес-процесу. Для цього у роботі пропонується визначити основні елементи BPMN-моделей бізнес-процесів та їх структурні характеристики, які впливають на зрозумілість та придатність моделей до використання, та можуть бути джерелами помилок. Зважаючи на визначені характеристики, пропонується виконати розрахунок відповідних значень для великої колекції BPMN-моделей бізнес-процесів, після чого дослідити дотримання теоретичних правил моделювання бізнес-процесів на практиці при побудові моделей реальних бізнес-процесів. Для забезпечення ефективного зберігання та обробки великої колекції даних про BPMN-моделі бізнес-процесів було розроблено базу даних та програмний компонент. Продемонстровано та проаналізовано результати аналізу структурних характеристик BPMN-моделей бізнес-процесів, отримані з використанням розробленої бази даних та програмного компонента. Сформульовано висновки, а також визначено подальші напрямки досліджень.

Ключові слова: модель бізнес-процесу, нотація моделювання, структурні характеристики, правила моделювання.

Introduction. Business process management (BPM) and, respectively, business process modeling are used by modern organizations focused on continuous improvement of their performance [1].

Using business process modeling as the core tool of the BPM approach, small and large companies may capture their activities in the form of graphic diagrams that could be later brainstormed by business analysts to find ways for organizational workflows improvement.

Therefore, the understandability and modifiability of business process models are crucial for the success of future business process improvement projects [2]. It is natural, that error-prone business process models that are not clear for their readers could negatively affect the success of BPM projects, causing shortcomings starting from the design stage. Moreover, in the later stages of information system development or customization projects, business process models of poor quality could lead to literally a hundred times bigger and, therefore, expensive mistakes [3].

Problem statement. Nowadays, a BPMN (Business Process Model and Notation) is the de-facto standard notation for business process modeling [4].

This notation is quite similar to its predecessors EPC (Event-driven Process Chains) and IDEF3 [5], however, it is much more complex, due to the rich set of activities, events, and gateways (logical connectors used to manage process execution scenarios, e.g. “and”, “or”, “xor” etc.) of different types.

Deceptive “simplicity” could make inexperienced business analysts build BPMN models as simple workflow diagrams or vice versa, using complex BPMN constructs where it is not necessary. Both of these cases may lead to poor-structured business process models understandable only to their authors. It is basically the same as using a fancy text editor to write down a poem, but with a lot of grammar mistakes, or writing an essay or graduation thesis speech, but on a paper note.

Thus, it is relevant to study the main sources of BPMN modeling mistakes to provide recommendations for their efficient detection and elimination.

The research object includes an analysis of BPMN business process models structural measures.

The research subject includes the database and the software component to provide efficient storage and processing of a large collection of BPMN business process models data.

This research aims at the improvement of designed BPMN models by detection and elimination of modeling mistakes.

Proposed approach. In the proposed approach we suggest considering size measures of business process models, formulated at first for EPC models in [6] and later applied to measure BPMN models in [7]:

$$SM_{BPMN} = \langle SM_{OR}, SM_{SE}, SM_{EE}, SM_N \rangle, \qquad (1)$$

here SM_{OR} – the number of inclusive gateways (OR logic);

SM_{SE} – the number of start events;

SM_{EE} – the number of end events;

SM_N – the number of all nodes.

As the dataset of sample BPMN models, we propose to use the model collection of BPMAI (Business Process Management Academic Initiative) [8]. It provides JSON (JavaScript Object Notation) files with business process model metadata, including necessary measures (1).

Hence, the following algorithm (see fig. 1) is suggested to process input JSON files of BPMN models' properties to compare structural measures toward the thresholds defined in [7] and estimate error probabilities for BPMN models.

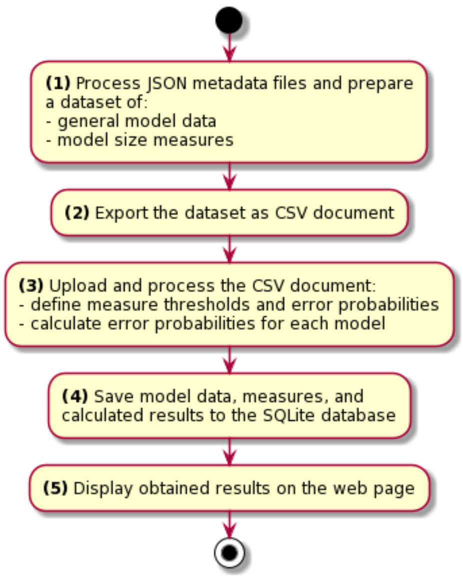

Fig. 1. The algorithm of BPMN models data analysis

Respective thresholds defined by authors of [7] for size measures (1) of business process models are following:

$$TM_{BPMN} = \langle TM_{OR}, TM_{SE}, TM_{EE}, TM_N \rangle, \qquad (2)$$

here TM_{OR} – the threshold to evaluate the number of inclusive gateways (OR logic), $TM_{OR} = 0.5$;

TM_{SE} – the threshold to evaluate the number of start events, $TM_{SE} = 2.5$;

TM_{EE} – the threshold to evaluate the number of end events, $TM_{EE} = 2.5$;

TM_N – the threshold to evaluate the number of all nodes, $TM_N = 31.5$.

Therefore, according to [7], there could be defined the following probabilities of finding errors in business process models when thresholds (2) are overcome:

- $P(SM_{OR} > TM_{OR}) = 0.09$;
- $P(SM_{SE} > TM_{SE}) = 0.07$;
- $P(SM_{EE} > TM_{EE}) = 0.05$;
- $P(SM_N > TM_N) = 0.09$.

Considering the multiplication rule [9], we would like to suggest the following equation to calculate the error probability of a BPMN model:

$$hasErrors = 1 - \prod_{i \in M} [1 - P(SM_i > TM_i)], \qquad (3)$$

here M – the set of business process model elements used to calculate measures (1), $M = \{OR, SE, EE, N\}$.

Results. Data extracted from JSON files of BPMAI collection [8] should be stored in the database for better performance and consistency. The suggested data model (see fig. 2) is similar to the "star" schema of data warehouses, suitable for analytical data processing [10].

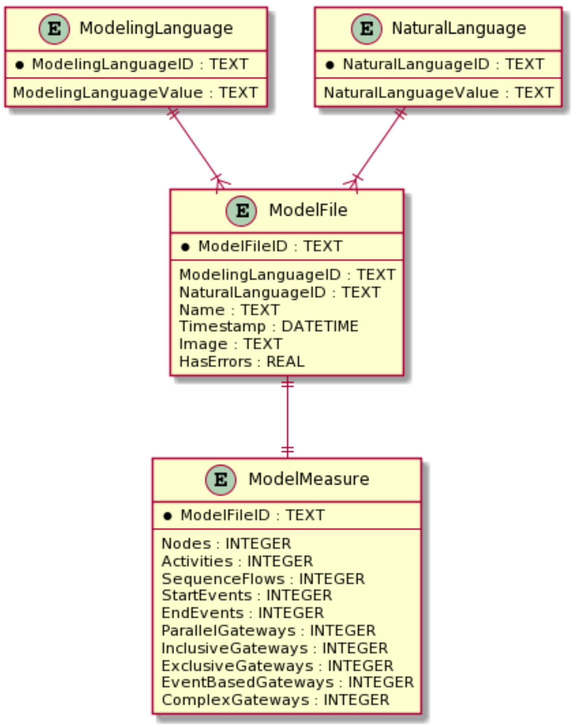

Fig. 2. The schema of the BPMN metadata database

The overall system for BPMN models analysis works as follows (see fig. 3):

- BPMAI JSON files are processed using NodeJS script and model names, and calculated measures (1) are stored in the CSV file (Comma-Separated Values);
- the CSV file is then processed using Python script to calculate error probabilities (3) and all obtained results are stored in the SQLite database;
- the Flask web application queries results from the database and transmits them to the web browser, where the JSON viewer plugin can display them (see fig. 4).

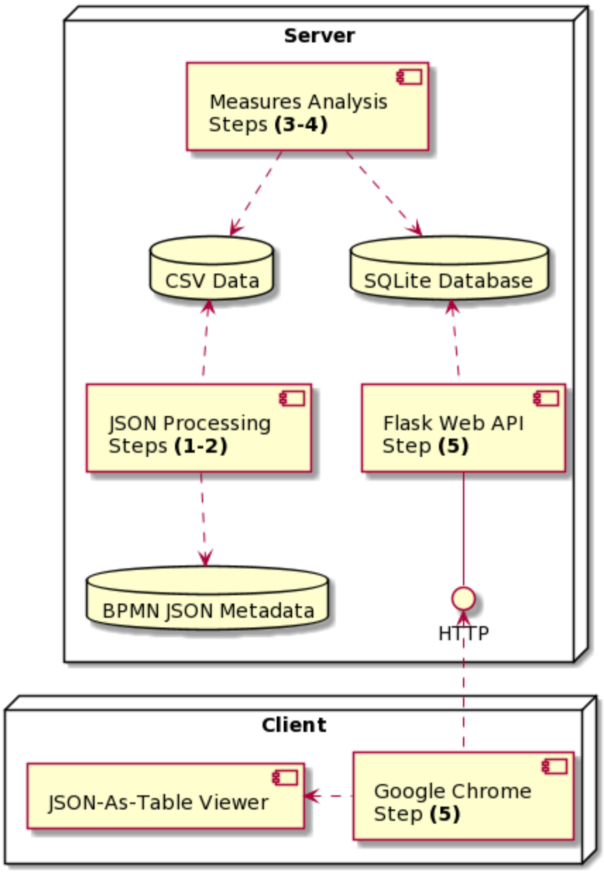

Fig. 3. Experimental system design

The example of obtained results is shown in fig. 4.

Discussion. Obtained statistical results (see fig. 5) demonstrate that redundant end events are the most often sources of errors in BPMN business process models – there are 4463 (23.72%) BPMN models with redundant end events that may mislead readers.

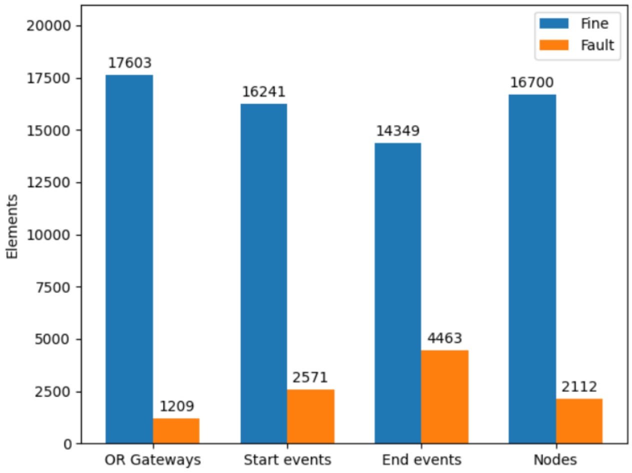

Fig. 5. The most common sources of errors in BPMN models

Other almost equal sources of errors are redundant start events and large models in general:

- 2571 (13.67%) BPMN models contain redundant start events that may mislead readers;
- 2112 (11.23%) BPMN models are considered too large and should be decomposed.

Usage of inclusive (i.e. "or") gateways is a less popular source of errors in comparison to other size measures – there are 1209 (6.43%) BPMN models with "or" gateways.

Anyway, inclusive gateways are recommended to be avoided because the semantics of these elements produces paradoxes and problems in the implementation of the business process model [11]. It is also stated that BPMN models with the exclusive (XOR) and parallel (AND) gateways tend to have a lower number of mistakes [11].

The results of the statistical analysis of calculated BPMN business process model measures, shown in (1), are demonstrated in Table 1.

Obtained statistical results demonstrate that 75% of analyzed BPMN models (18812 models in total) are well-

Show 10 entries Search: supply chain

Model File	Model Name	Start Events	End Events	Or Gateways	Nodes	Has Errors
1408075999.meta.json	AVERS_Supply_Chain_NEW	3	9	0	41	0.21
2019061664.meta.json	AVERS_Supply_Chain	4	10	0	39	0.21
1322223884.meta.json	Supply chain	3	3	0	16	0.12
643475489.meta.json	Customer Data Supply Chain Map	0	0	0	32	0.09
239343585.meta.json	Data Supply Chain	0	0	0	6	0

Showing 1 to 5 of 5 entries (filtered from 18,812 total entries) Previous 1 Next

Fig. 4. The example of results for "supply chain" BPMN models

structured since their measures (1) are not crossing the threshold values (2). Also, we can observe that the average BPMN model consists of around 17 elements in general, among which there is one start event (rarely two), two end events, and the presence of at least one inclusive gateway is highly unlikely.

Table 1 – The statistical analysis of BPMN measures.

Measure	SM_{OR}	SM_{SE}	SM_{EE}	SM_N
Mean	0.13	1.55	2.01	16.77
Std.	0.72	1.85	2.42	16.05
Min.	0.00	0.00	0.00	0.00
25%	0.00	1.00	1.00	7.00
50%	0.00	1.00	1.00	13.00
75%	0.00	2.00	2.00	21.00
Max.	23.00	103.00	49.00	262.00

We also complemented obtained results with the correlation analysis [12] of error probability and other size measures of BPMN models that were not initially included in the tuple (1), as well as ignored by authors of [7]:

- SM_A the number of activities;
- SM_{SF} the number of sequence flows.

Obtained results of the correlation analysis [12] are demonstrated in Table 2.

Table 2 – The correlation analysis results.

Variable	$hasErrors$	SM_A	SM_{SF}
$hasErrors$	1.00	0.62	0.67
SM_A	0.62	1.00	0.90
SM_{SF}	0.67	0.90	1.00

Since correlation coefficients are between 0.5 and 0.7, we may conclude that the error probability and the number of activities (0.62), as well as the error probability and the number of sequence flows (0.67), can be considered as moderately correlated [12].

It can be also noticed that the number of activities and the number of sequence flows are highly correlated (their correlation coefficient is 0.9) [12], however, it does not seem unusual for linked graphs such as BPMN diagrams.

These observations (shown in Table 2) could be used in the future to build a model for the estimation of BPMN business process models using numbers of activities and sequence flows.

Conclusion. In this paper were considered measures of BPMN business process models and their respective thresholds [7]. Using these thresholds [7] it is proposed to calculate error probabilities of analyzed BPMN models. To provide efficient storage and processing of experimental BPMN data, we have developed the database and software component. Using the software component BPMN data is extracted from JSON metadata files, provided in BPMAI research collection [8], then data is stored in a CSV file that later is processed by a Python script, and analysis results are stored in the SQLite database. Using the Flask web applications development framework, analysis results of BPMN models are displayed using the web page with search and pagination possibilities.

Using further statistical analysis of obtained results were defined most common sources of errors in BPMN models. Also, were defined typical measures of average BPMN models and correlations between other BPMN size measures and error probabilities. Future research may include the elaboration of models for early errors detection in BPMN models using essential measures, such as numbers of activities and sequence flows.

References

1. Dumas M. et al. *Fundamentals of Business Process Management*. Springer, 2018. 527 p.
2. Reijers H. A., Mendling J. A study into the factors that influence the understandability of business process models. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*. 2010. No. 3 (41). P. 449–462.
3. Ransome J., Misra A. *Core Software Security: Security at the Source*. CRC Press, 2018. 414 p.
4. Allweyer T. *BPMN 2.0: Introduction to the Standard for Business Process Modeling*. BoD – Books on Demand, 2016. 172 p.
5. Topi H., Tucker A. *Computing Handbook: Information Systems and Information Technology*. CRC Press, 2014. 1522 p.
6. Mendling J., Neumann G., van der Aalst W. On the correlation between process model metrics and errors. *Australian Computer Society*. 2007. P. 173–178.
7. Mendling J., Sánchez-González L., García F., la Rosa M. Thresholds for error probability measures of business process models. *Journal of Systems and Software*. 2012. No. 5 (85). P. 1188–1197.
8. Weske M., Decker G., Dumas M., La Rosa M., Mendling J., Reijers H. A. *Model Collection of the Business Process Management Academic Initiative (Version BPMAI-29-10-2019) [Data set]*. URL: https://doi.org/10.5281/zenodo.3758705 (accessed: 20.03.2022).
9. Weinberg S. L., Abramowitz S. K. *Statistics Using R: An Integrative Approach*. Cambridge University Press, 2020. 692 p.
10. Nordeen A. *Learn Data Warehousing in 24 Hours*. Guru99, 2020. 380 p.
11. Krogstie J. *Quality in Business Process Modeling*. Springer, 2016. 250 p.
12. *Correlation Coefficients*. URL: https://www.andrews.edu/~calkins/math/edrm611/edrm05.htm (accessed: 20.03.2022).

References (transliterated)

1. Dumas M. et al. *Fundamentals of Business Process Management*. Springer, 2018. 527 p.
2. Reijers H. A., Mendling J. A study into the factors that influence the understandability of business process models. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*. 2010, no. 3 (41), pp. 449–462.
3. Ransome J., Misra A. *Core Software Security: Security at the Source*. CRC Press, 2018. 414 p.
4. Allweyer T. BPMN 2.0: *Introduction to the Standard for Business Process Modeling*. BoD – Books on Demand, 2016. 172 p.
5. Topi H., Tucker A. *Computing Handbook: Information Systems and Information Technology*. CRC Press, 2014. 1522 p.
6. Mendling J., Neumann G., van der Aalst W. On the correlation between process model metrics and errors. *Australian Computer Society*. 2007, pp. 173–178.
7. Mendling J., Sánchez-González L., García F., la Rosa M. Thresholds for error probability measures of business process models. *Journal of Systems and Software*. 2012, no. 5 (85), pp. 1188–1197.
8. Weske M., Decker G., Dumas M., La Rosa M., Mendling J., Reijers H. A. *Model Collection of the Business Process Management Academic Initiative (Version BPMAI-29-10-2019) [Data set]*. Available at: https://doi.org/10.5281/zenodo.3758705 (accessed 20.03.2022).
9. Weinberg S. L., Abramowitz S. K. *Statistics Using R: An Integrative Approach*. Cambridge University Press, 2020. 692 p.

10. Nordeen A. *Learn Data Warehousing in 24 Hours*. Guru99, 2020. 380 p.
11. Krogstie J. *Quality in Business Process Modeling*. Springer, 2016. 250 p.
12. *Correlation Coefficients*. Available at: https://www.andrews.edu/~calkins/math/edrm611/edrm05.htm (accessed 20.03.2022).

Надійшла (received) 05.04.2022

Відомості про авторів / About the Authors

Копп Андрій Михайлович – доктор філософії, Національний технічний університет «Харківський політехнічний інститут», доцент кафедри програмної інженерії та інтелектуальних технологій управління; м. Харків, Україна; ORCID: http://orcid.org/0000-0002-3189-5623; e-mail: kopp93@gmail.com

Орловський Дмитро Леонідович – кандидат технічних наук, доцент, Національний технічний університет «Харківський політехнічний інститут», доцент кафедри програмної інженерії та інтелектуальних технологій управління; м. Харків, Україна; ORCID: http://orcid.org/0000-0002-8261-2988; e-mail: orlovskyi.dm@gmail.com

Лютенко Ірина Вікторівна – кандидат технічних наук, доцент, Національний технічний університет «Харківський політехнічний інститут», доцент кафедри програмної інженерії та інтелектуальних технологій управління; м. Харків, Україна; ORCID: http://orcid.org/0000-0003-4357-1826; e-mail: Iryna.Liutenko@khpi.edu.ua

Kopp Andrii Mykhailovych – PhD in Computer Sciences, National technical university «Kharkiv polytechnic institute», Associate Professor of the Department of Software Engineering and Management Intelligent Technologies; Kharkiv, Ukraine; ORCID: http://orcid.org/0000-0002-3189-5623; e-mail: kopp93@gmail.com

Orlovskyi Dmytro Leonidovych – Candidate of Technical Sciences (PhD), Docent, National technical university «Kharkiv polytechnic institute», Associate Professor of the Department of Software Engineering and Management Intelligent Technologies; Kharkiv, Ukraine; ORCID: http://orcid.org/0000-0002-8261-2988

Liutenko Iryna Vikrorivna – Candidate of Technical Sciences (PhD), Docent, National technical university «Kharkiv polytechnic institute», Associate Professor of the Department of Software Engineering and Management Intelligent Technologies; Kharkiv, Ukraine; ORCID: http://orcid.org/0000-0003-4357-1826; e-mail: Iryna.Liutenko@khpi.edu.ua