UDC 004.94 **DOI: 10.20998/2079-0023.2022.01.11**

A. M. KOPP, D. L. ORLOVSKYI

THE APPROACH AND THE SOFTWARE TOOL TO CALCULATE SEMANTIC QUALITY MEASURES OF BUSINESS PROCESS MODELS

Business process models are essential business process management artifacts that help describe visually ongoing business activities to facilitate communication between information technology and business stakeholders. Business process models are used to find inefficient spots within described workflows and resolve detected shortcomings by automation via configurable software solutions or unified workflow engines. However, this is impossible when using syntactically or semantically poor business process models. It is the same as building a house using the blueprint with windows on the floor and typos in text labels. Therefore, it is extremely important to keep created business process models clear and relevant to the actual workflows they describe. Hence, in this paper, we propose the approach and the software tool to calculate semantic quality measures of business process models. The proposed approach uses a special procedure to extract the modeling domain statements using natural language processing techniques. According to the proposed approach, the initial textual descriptions of business process models should be tokenized. Then obtained tokens should be turned to the lower case style and cleansed to remove non-alphabetic tokens and stop words. Finally, the remaining tokens should be stemmed and the existing duplicates should be removed. The same procedure is then repeated for text labels attached to the business process model activities. Then, tokens present in the result of textual description's processing but missing in the result of labels' processing are considered incomplete (i.e. incorrect in the modeling domain). Similarly, tokens present in the result of labels' processing but missing in the result of textual description's processing are considered invalid (i.e. irrelevant to the modeling domain). Therefore, respective semantic quality measures can be calculated. The software tool is created using the Python programming language because of its powerful natural language processing packages.

Keywords: business process modeling, semantic quality, quality measure, natural language processing, software tool.

А. М. КОПП, Д. Л. ОРЛОВСЬКИЙ

ПІДХІД ТА ПРОГРАМНИЙ ЗАСІБ ДЛЯ РОЗРАХУНКУ СЕМАНТИЧНИХ МІР ЯКОСТІ МОДЕЛЕЙ БІЗНЕС-ПРОЦЕСІВ

Моделі бізнес-процесів є важливими артефактами управління бізнес-процесами, які допомагають візуально описати поточну бізнес-діяльність, щоб полегшити взаємодію зацікавлених сторін бізнесу з інформаційними технологіями. Моделі бізнес-процесів використовуються для пошуку неефективних місць в описаних робочих процесах та усунення виявлених недоліків шляхом автоматизації за допомогою програмних рішень, що налаштовуються, або уніфікованих середовищ виконання робочих процесів. Однак це неможливо при використанні синтаксично або семантично некоректних моделей бізнес-процесів. Це те саме, що будувати будинок за кресленням із вікнами на підлозі та помилками в текстових написах. Тому надзвичайно важливо зберігати створені моделі бізнес-процесів чіткими та відповідними фактичним робочим процесам, які вони описують. Тому в цій роботі ми пропонуємо підхід та програмний інструмент для розрахунку семантичних показників якості моделей бізнес-процесів. Запропонований підхід використовує спеціальну процедуру для вилучення тверджень про предметну область з використанням методів обробки природної мови. Відповідно до запропонованого підходу початкові текстові описи моделей бізнес-процесів мають бути токенізованими. Потім отримані токени слід подати у нижньому регістрі та очистити, щоб видалити неалфавітні токени та стоп-слова. Нарешті, решту токенів слід стематизувати, а наявні дублікати слід видалити. Потім ця ж процедура повторюється для текстових міток, доданих до робіт моделі бізнес-процесу. Тоді лексеми, присутні в результаті обробки текстового опису, але відсутні в результаті обробки текстових міток, вважаються неповними (тобто неправильними щодо предметної області моделювання). Аналогічно, лексеми, присутні в результаті обробки текстових міток, але відсутні в результаті обробки текстового опису, вважаються недійсними (тобто не мають відношення до предметної області моделювання). Таким чином, можна розрахувати відповідні показники семантичної якості. Програмний засіб був створений за допомогою мови програмування Python завдяки її потужним пакетам обробки природної мови.

Ключові слова: моделювання бізнес-процесів, семантична якість, міра якості, обробка природної мови, програмний засіб.

Introduction. Today digital transformation is a trend in the enterprise management field. First of all, it considers Business Process Management (BPM) methodology and its main techniques, including business process modeling and workflow automation. These approaches rely on special BPM diagramming software or even complex business process automation suites. While workflow automation takes into account mostly routine documents and information flows, business process modeling, in general, allows representation of business activities visually as diagrams. Graphical business process models serve for identification and deeper understanding of organizational activities. They also help to find workflow shortcomings and facilitate interaction between Information Technology (IT) engineering experts and business stakeholders [1].

A business process is the sequence of manual or automated activities that aims at achieving organizational goals and providing products or services valuable for business users. Automated activities use specialized IT systems that implement required workflows fully or partially. These are BPM suites, Customer Relationship Management (CRM), Enterprise Resource Planning (ERP), and other IT solutions. The BPM methodology assumes workflow automation to improve the communication of business users with IT systems. In general, the business process management lifecycle includes the following repeated stages [2]:

- business process analysis;
- business process modeling;
- business process implementation (or deployment to the execution environment if the workflow is completely automated);
- business process monitoring;
- business process evaluation.

Graphical business process models are the most efficient in comparison to textual business process descriptions. They visually describe business process tasks, events that trigger process scenarios, transmitted data objects, organizational units or single persons

involved into the business process execution, and provided outcomes [3].

Business process modeling is the technique of visual representation of ongoing or planned business activities. Usually, business process models contain events that trigger various workflow scenarios while being driven by the control flow logic of parallel, exclusive, or inclusive branches. Business processes are pillars of process-centric IT solutions since they help understand and improve enterprise performance through re-engineering activities. Business processes are also subject to the service-oriented design in software engineering. Business process models are also knowledge-sharing assets. They help transfer best practices across organizations to improve industry standards and achieve better business performance [4].

In [5] author proposed the following classification of business process modeling aspects, including object-based, communication-based, role-based, topological, functional, and behavioral. The most popular business process modeling notations and languages that belong to the proposed taxonomy are the following [5]:

- Unified Modeling Language (UML);
- Integrated Definition Standards (IDEF0 and IDEF3);
- Data Flow Diagrams (DFD);
- Event-driven Process Chains (EPC);
- Business Process Modeling and Notation (BPMN).

The BPMN notation nowadays is the most popular and de-facto standard of business process modeling and workflow automation using BPM suites. BPMN combines behavioral and functional approaches [5].

Besides the BPMN which is a leader in the business process modeling field, the EPC notation is widely used. However, recently EPC is being replaced by BPMN [6].

Many EPC modeling software tools now support also the BPMN as an alternative business process diagramming approach [6]. For example, the "ARIS Express" application supports the BPMN notation as the supplementary business process diagramming capability.

Other standards (i.e. UML, IDEF, and DFD) are not such popular in today's BPM industry [6]. However, IDEF and DFD standards were widely used in the nineties by the U.S. Department of Defense (DoD) as the business process re-engineering methodology [7].

Currently, there are about 70 software tools that allow the building of BPMN diagrams [8]. Also, about 50 open-source tools are listed on the SourceForge [9].

Problem statement. Business process models are essential BPM artifacts that help describe visually ongoing business activities to facilitate communication between IT and business stakeholders.

Business process models are used to find inefficient spots within described workflows and resolve detected shortcomings by automation via configurable IT solutions or unified BPM engines. However, this is impossible when using syntactically or semantically poor business process models. Literally, it is the same as building a house using the blueprint with windows on the floor and typos in text labels.

Therefore, it is extremely important to keep created business process models clear and relevant to the actual workflows they describe. Hence, in this paper, we propose the approach and the software tool to calculate semantic quality measures of business process models.

Semantic quality measures of business process models. In [5] the author proposes the following definitions of syntactic and semantic qualities according to the "Semiotic Quality" (SEQUAL) approach application to the business process modeling domain:

- syntactic quality is the degree of correspondence between elements of a business process model and the syntax of the modeling notation;
- semantic quality is the degree of correspondence between elements of a business process model and the actual business process it describes.

Moreover, in [5] the author proposes the following phenomena, such as validity and completeness. Therefore, there could be defined semantic and syntactic validity and completeness degrees respectively [5]:

- a business process model is syntactically valid when all of its elements belong to the modeling notation;
- a business process model is syntactically complete when all of its elements obey the syntax of the modeling notation;
- a business process model is semantically valid when all of its elements belong to the modeling domain (i.e. the actual business process);
- a business process model is semantically complete when all of its elements are correct in the modeling domain.

While the syntactic quality was extensively studied in previous papers [10] and [11], let us now concentrate on the semantic quality. Formally, the semantic quality measures validity and completeness can be given using the following equations:

$$Semantic\ validity = 1 - \frac{\#M \backslash D}{\#M}, \quad (1)$$

$$Semantic\ completeness = 1 - \frac{\#D \backslash M}{\#D}. \quad (2)$$

Here in (1) and (2):

- $\#M \backslash D$ is the number of business process model M elements that do not belong to the modeling domain D (i.e. not relevant to the actual business process);
- $\#D \backslash M$ is the number of business process model M elements that are incorrect in the modeling domain D (i.e. poorly describing the actual business process);
- $\#M$ is the total number of business process model elements;
- $\#D$ is the total number of the modeling domain statements.

Calculation of the semantic quality measures. The proposed approach uses the following procedure to extract the modeling domain statements using Natural Language Processing (NLP) techniques.

According to this procedure:

- the textual description of a business process S is tokenized and the bag of tokens T is obtained:

$$T = \{t_i | i = \overline{1,n}\}; \quad (3)$$

- the obtained tokens are turned into the lower case style and the bag L is obtained:

$$lower: T \to L; \quad (4)$$

- non-alphabetic tokens are removed and the bag of alphabetic tokens A remains:

$$alpha: L \to A; \quad (5)$$

- stop words are removed and the bag of meaningful tokens W remains:

$$stop: A \to W; \quad (6)$$

- tokens are stemmed and the bag of reduced tokens R remains:

$$stem: W \to R; \quad (7)$$

- finally, all possible duplicates are removed from the bag of tokens and the set U is obtained:

$$unique: R \to U; \quad (8)$$

The sequence of the procedure's steps is demonstrated in fig. 1 below.

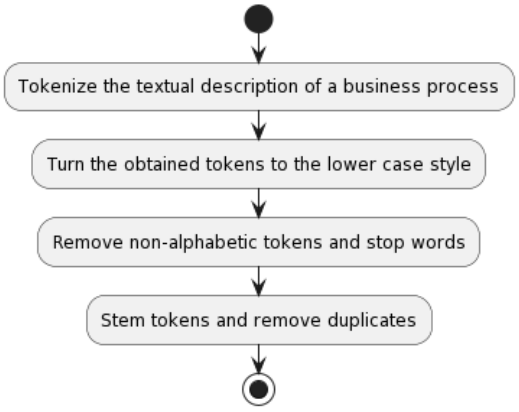

Fig. 1. The procedure for extraction of the modeling domain statements

This procedure (see fig. 1) can be applied not only to the textual description of a considered business process but also to text labels of the respective business process model activities.

Hence, using the tokens (3) transformed according to equations (4) – (8), two sets will be obtained:

- U_1 – the set of modeling domain tokens extracted from the textual description of a business process;
- U_2 – the set of tokens extracted from business process model element labels.

Therefore, we consider tokens that are present in U_2 but missing in the $U_1 \cap U_2$ as invalid – they do not belong to the modeling domain:

$$\#M \backslash D = \#U_2 - \#U_1 \cap U_2, \#M = \#U_2. \quad (9)$$

Similarly, we consider tokens that are present in U_1 but missing in the $U_1 \cap U_2$ as incorrect – they are incorrect in the modeling domain:

$$\#D \backslash M = \#U_1 - \#U_1 \cap U_2, \#D = \#U_1. \quad (10)$$

Finally, respective semantic quality measures can be calculated using equations (1) and (2).

The semantic quality measurement example. Let us consider the sample goods dispatch business process [12]. Its description in a text form is given in table 1 below, as well as the set of activity labels taken from its business process model given in the BPMN notation [12].

Table 1 – The input data for semantic quality measurement.

Business process description	If goods shall be shipped, the secretary clarifies who will do the shipping. If you have large amounts, special shipping will be necessary. In these cases the secretary invites three logistic companies to make offers and she selects one of them. In case of small amounts, normal post shipment is used. Therefore a package label is written by the secretary and a parcel insurance taken by the logistics department head if necessary. In the meantime the goods can be already packaged by the warehousemen. If everything is ready, the packaged goods are prepared for being picked up by the logistic company.
BPMN model's activity labels	Insure parcel Write package label Clarify shipment method Get 3 offers from logistic companies Select logistic company and place order Package goods Prepare for picking up goods

Using the developed Python software tool, we got the following sets (see table 2).

Table 2 – The obtained sets U_1 and U_2.

U_1	[good, shall, ship, secretari, clarifi, larg, amount, special, necessari, case, invit, three, logist, compani, make, offer, select, one, small, normal, post, shipment, use, therefor, packag, label, written, parcel, insur, taken, depart, head, meantim, alreadi, warehousemen, everyth, readi, prepar, pick]
U_2	[insur, parcel, write, packag, label, clarifi, shipment, method, get, offer, logist, compani, select, place, order, good, prepar, pick]

Using the obtained sets demonstrated in table 2 we can calculate the semantic quality measures demonstrated in (1) and (2) using assumptions made in (9) and (10):

$$Semantic\ validity = 1 - \frac{5}{18} = 0.72, \quad (11)$$

$$Semantic\ completeness = 1 - \frac{26}{39} = 0.33. \quad (12)$$

The obtained results are shown in (11) and (12) above can be interpreted as follows:

- the analyzed business process model [12] is rather semantically valid since 72% of its activities belong to the modeling domain;
- the analyzed business process model [12] is rather semantically incomplete since only 33% of its activities reflect the actual business process.

Using the semantic quality measures (11) and (12), the analyzed BPMN model can be improved by replacing activities that are not relevant to the modeling domain and adding activities that reflect the actual business process.

Conclusion. In this paper, we proposed the approach and the software tool to find semantic quality measures of business process models. This approach is based on the procedure driven by NLP techniques. It considers textual descriptions of business processes as reference domain descriptions and text labels of business process model activities as evaluation subjects. Thus, using the Python Natural Language Toolkit (NLTK), the sets of lexemes that describe the modeling domain and the analyzed business process model were obtained and used to find semantic quality measures. In the future, the software tool should be improved to automatically "read" business process models, i.e. using optical character recognition or other techniques, and calculate their semantic quality measures.

References

1. Kopp A., Orlovskyi D. Towards the Tokenization of Business Process Models using the Blockchain Technology and Smart Contracts. *CEUR Workshop Proceedings*. 2022. Vol. 3137. P. 274–287.
2. Gerth C. Business Process Modeling. *Business Process Models: Change Management. Lecture Notes in Computer Science*. Springer, 2013. P. 14–23.
3. Scheer A.-W. Strategic Business Process Analysis. *ARIS – Business Process Modeling*. Springer, 2013. P. 7–20.
4. Bitkowska A. The relationship between Business Process Management and Knowledge Management – selected aspects from a study of companies in Poland. *Business Process Management: Current Applications and the Challenges of Adoption*. Cognitione Foundation for Dissemination of Knowledge and Science, 2020. P. 169–193.
5. Krogstie J. *Quality in Business Process Modeling*. Springer, 2016. 250 p.
6. Allweyer T. BPMN – A Standard from Business Process Modeling. *BPMN 2.0: Introduction to the Standard for Business Process Modeling*. BoD – Books on Demand, 2016. P. 9–14.
7. *Deming, IT, and BPM IDEF0 Diagrams*. URL: https://www.bptrends.com/deming-it-and-bpm-idef0-diagrams/ (accessed: 25.03.2022).
8. *BPMN Tool Matrix*. URL: https://bpmnmatrix.github.io/ (accessed: 25.03.2022).
9. *SourceForge – Search results for "bpmn"*. URL: https://sourceforge.net/directory/?q=bpmn (accessed: 25.03.2022).
10. Kopp A., Orlovskyi D. A Method for Business Process Model Analysis and Improvement. *CEUR Workshop Proceedings*. 2019. Vol. 2403. P. 1–10.
11. Kopp A., Orlovskyi D. Intelligent Support of the Business Process Model Analysis and Improvement Method. *Information and Communication Technologies in Education, Research, and Industrial Applications*. Springer, 2020. P. 111–135.
12. *BPMN for research*. URL: https://github.com/camunda/bpmn-for-research (accessed: 25.03.2022).

References (transliterated)

1. Kopp A., Orlovskyi D. Towards the Tokenization of Business Process Models using the Blockchain Technology and Smart Contracts. *CEUR Workshop Proceedings.* 2022, vol. 3137, pp. 274–287.
2. Gerth C. Business Process Modeling. *Business Process Models: Change Management. Lecture Notes in Computer Science*. Springer, 2013, pp. 14–23.
3. Scheer A.-W. Strategic Business Process Analysis. *ARIS – Business Process Modeling*. Springer, 2013, pp. 7–20.
4. Bitkowska A. The relationship between Business Process Management and Knowledge Management – selected aspects from a study of companies in Poland. *Business Process Management: Current Applications and the Challenges of Adoption*. Cognitione Foundation for Dissemination of Knowledge and Science, 2020, pp. 169–193.
5. Krogstie J. *Quality in Business Process Modeling*. Springer, 2016. 250 p.
6. Allweyer T. BPMN – A Standard from Business Process Modeling. *BPMN 2.0: Introduction to the Standard for Business Process Modeling*. BoD – Books on Demand, 2016, pp. 9–14.
7. *Deming, IT, and BPM IDEF0 Diagrams*. Available at: https://www.bptrends.com/deming-it-and-bpm-idef0-diagrams/ (accessed 25.03.2022).
8. *BPMN Tool Matrix*. Available at: https://bpmnmatrix.github.io/ (accessed 25.03.2022).
9. *SourceForge – Search results for "bpmn"*. Available at: https://sourceforge.net/directory/?q=bpmn (accessed 25.03.2022).
10. Kopp A., Orlovskyi D. A Method for Business Process Model Analysis and Improvement. *CEUR Workshop Proceedings*. 2019, vol. 2403, pp. 1–10.
11. Kopp A., Orlovskyi D. Intelligent Support of the Business Process Model Analysis and Improvement Method. *Information and Communication Technologies in Education, Research, and Industrial Applications*. Springer, 2020, pp. 111–135.
12. *BPMN for research*. Available at: https://github.com/camunda/bpmn-for-research (accessed 25.03.2022).

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

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 – PhD in Technical Sciences, 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; e-mail: orlovskyi.dm@gmail.com