7. Larsson A., Viitaoja Y. Building customer loyalty in digital banking: A study of bank staff's perspectives on the challenges of digital CRM and loyalty. *International Journal of Bank Marketin.* 2017. Vol. 35, № 6. P. 858–877. URL: https://www.emerald.com/insight/content/doi/10.1108/IJBM-08-2016-0112/full/html

8. The 5 Hottest Technologies In Banking For 2022. URL: https://www.forbes.com/sites/ronshevlin/2021/12/31/the-5-hottest-technologies-in-banking-for-2022/

9. Commercial bank branches (per 100,000 adults). URL: https://data.worldbank.org/indicator/FB.CBK.BRCH.P5

10. Regulatory governance and independence. URL: https://digitalregulation.org/regulatory-governance-and-independence/

Rudskyi O.,
Master's Student at the Software Engineering and Management Intelligent Technologies Department,
National Technical University "Kharkiv Polytechnic Institute"
Kopp A.,
Ph.D., Associate Professor,
Head of Software Engineering and Management Intelligent Technologies Department,
National Technical University "Kharkiv Polytechnic Institute"

USING MACHINE LEARNING TO ASSESS THE COMPLIANCE OF BUSINESS PROCESS MODELS WITH TEXTUAL DESCRIPTIONS

This article discusses the problem of comparing business process models with their textual descriptions. This article employs a Systematic Literature Review (SLR) methodology to explore methods for comparing textual descriptions with business process models, aiming to enhance accuracy, identify discrepancies, and foster a common understanding among stakeholders.

Keywords: machine learning, assess, compliance, business process, models, textual descriptions

Business process modeling emerged to provide a better understanding of business processes in organizations. The result of business process modeling is a process model, which consists of a set of activity models and the execution constraints between them. It is usually illustrated by activities and events that are associated with management flows. Such processes can be modeled using various process modeling languages, also known as techniques or notations [1].

Business Process Model and Notation (BPMN) is a standard for business process modeling that provides a graphical notation for specifying business processes in the form of a Business Process Diagram (BPD) based on traditional flowcharting methods [2].

The main goal of developing BPMN was to create an understandable notation for creating business process models, while providing semantics and underlying mechanisms to handle the complexity inherent in business processes. The approach taken to address these two conflicting requirements was to organize the graphical aspects of the notation into specific categories. This provides a small set of notation categories so that the reader of a BPMN diagram can easily recognize the basic types of elements and understand the diagram [2].

However, creating business process models is a complex and resource-intensive task, sometimes leading to situations where the model itself does not match the textual description of the business process. This can lead to time and financial losses. Therefore, there is a need to analyze the compliance of business process models with their textual descriptions.

Comparing business process models with their textual descriptions has several advantages. First, it can help ensure that the process model accurately reflects the textual description of the process. Second, it can help identify discrepancies between the two representations, which can be used to improve the quality of the process model. Third, it can help identify areas where the textual description of the process is ambiguous or incomplete, which can be used to improve the quality of the textual description [3].

In addition, comparing business process models with their textual descriptions can help ensure that all stakeholders have a common understanding of the process. This is because different stakeholders may have different levels of familiarity with the BPMN notation and therefore may interpret the process model differently [3].

A Systematic Literature Review (SLR) method was used to investigate current methods for comparing texts.

The SLR method identifies, selects, and critically appraises studies to answer a clearly defined question. A systematic review should follow a well-defined protocol or plan that clearly defines the criteria for the review. It is a comprehensive, transparent search that is conducted in multiple databases and gray literature that can be replicated and reproduced by other researchers. It involves planning a well-designed search strategy that has a specific focus or answers a specific question. A systematic literature review identifies the type of information that has been sought, critiqued, and reported over a known period of time. Search terms, search strategies (including database names, platforms, search dates), and limitations should all be included in the review [1].

To answer the research questions, the following SLR objectives were identified:

1) to review articles to identify existing methods for comparing texts;

2) to identify weaknesses in the methods in order to eliminate them through additional research;

3) to gain new knowledge about text comparison methods that can be used for further research.

For the purposes of the study, the following search string was used: ("allintitle:" + "text" + "similarity" + "site:" + "ieeexplore.ieee.org").

The initial keyword search yielded 107 scientific articles related to text similarity in different languages. After a thorough review of the articles, we excluded articles that were not directly related to the research topic but appeared in the search results due to the coincidence of keywords. In addition, articles were also excluded due to duplication, lack of full text, and if the studies did not address any of the research questions. After all the exclusions, 8 scientific articles remained (Fig. 1).

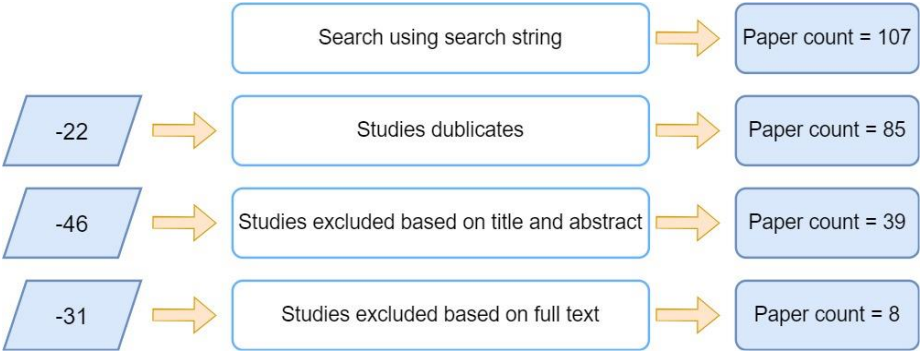

Fig. 1. General systematic literature review scheme

The following methods were discussed in these articles: Bag-of-Words (BOW), Word2vec, Word Vector Distance Decentralization (WVDD), Bidirectional Encoder Representation from Transformers (BERT), Global Vectors for Word Representation (GloVe), LDA model, FastText, and LexVec. Among these methods, the BERT model was chosen to process texts to evaluate business process models by their textual descriptions.

BERT is an abbreviation for Bidirectional Encoder Representation from Transformers, which is a transformer-based machine learning technique for pre-training natural language processing (NLP) developed by Google [4].

BERT can be defined as a function:

$$B : P \rightarrow R^{N \times h},$$

where h is the size of the hidden level;

$N = 512$ is the maximum sequence length supported by the model.

As an output, BERT receives a paragraph $\rho \epsilon P$ and decomposes it into a sequence $q \in N$ tokens $(p^j)_{j=1}^{q}$. After that, the sequence is supplemented with N elements by adding special CLS (Classification), SEP (Separator), and PAD (Padded) tokens.

This token sequence can be written in the form:

$$I^p = (CLS, (p^j)_{j=1}^{q}, SEP, \dots, PAD).$$

In BERT, all tokens are embedded using three functions: embedded tokens, positions, and segments, denoted as T, O, and G, respectively. Token embedding converts unique token values into intermediate vectors $T(I^p) \epsilon R^{N \times h}$. Position embedding encodes the token positions into a single space, $O(I^p) \epsilon R^{N \times h}$. Segment embedding is used to associate each token with one of two sequences $G(\{0,1\}^N) \epsilon R^{N \times h}$ [4]. The block diagram of the BERT model is given in Fig. 2.

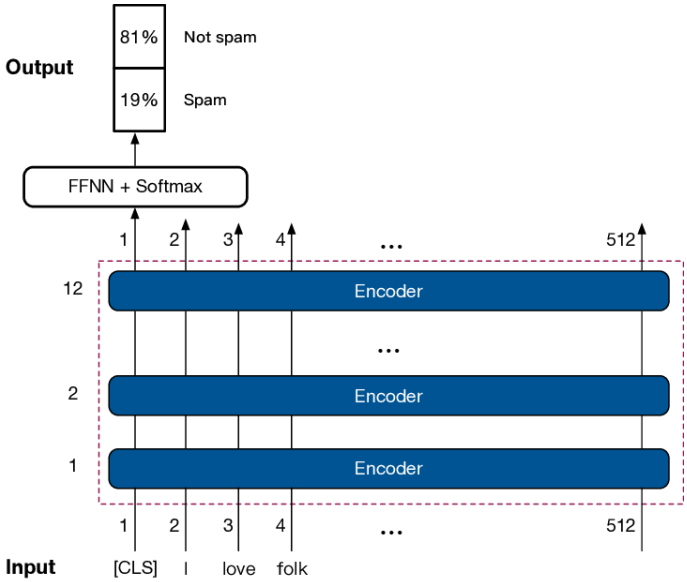

Fig. 2. Block diagram of the BERT model

To solve the problem of assessing the compliance of business process models with their textual descriptions, the software application under development must be able to separate from the BPMN file all the names of elements of the “task” type and related action elements:

– "Service Task" uses a web service, automated program, or other type of service to perform an activity;

– "Send Task" sends a message to another pool; the task is completed as soon as the message is sent;

– "Receive Task" indicates that the process must wait for the message to be received in order to continue working; the task will be completed as soon as the message is received;

– "User Task" means that a person performs the task using the software;

– "Manual Task" is performed without the help of any business process execution mechanism or any program;

– "Business Rule Task" provides a mechanism by which a process can provide input to a business rule engine and then receive output provided by the business rule engine;

– "Script Task" defines a script that the engine can interpret; when the task starts, the engine executes the script.

So, the software application must be able to process two files: a BPMN file and a text file describing the business process that this model should represent. From all the names of the elements of the "task" type, you need to generate the text T_1, and you also need to generate the text T_2, which will be obtained from the text file. The BERT model will be used to process these texts in order to evaluate business process models by their textual descriptions.

References

1. Jošt G., Polančič G., Heričko M., Kocbek M. Business process model and notation: The current state of affairs. *Computer Science and Information Systems*. 2015. 12(2). P. 509–539. DOI: 10.2298/CSIS140610006K.

2. Von Rosing M., White S., Cummins F., De Man H. Business process model and notation-BPMN. The Complete Business Process Handbook: Body of Knowledge from Process Modeling to BPM. 2014. Vol. 1. P. 429–453. DOI: 10.1016/B978-0-12-799959-3.00021-5.

3. Ottensooser A., Fekete A., Reijers H., Mendling J., Menictas C. Making sense of business process descriptions: An experimental comparison of graphical and textual notations. *Journal of Systems and Software*. 2012. Vol. 85(3). P. 596–606. DOI: 10.1016/j.jss.2011.09.023.

4. Zhengfang H, MacHica I, Zhimin B. Textual Similarity Based on Double Siamese Text Convolutional Neural Networks and Using BERT for Pre-training Model. IEEE 5th International Conference on Artificial Intelligence and Big Data. 2022. P. 107–111. DOI: 10.1109/ICAIBD55127.2022.9820371.