BPMN Inspector: A Tool for Extracting Features from BPMN Models

Ivan Compagnucci[1,*,†], Flavio Corradini[1,†], Fabrizio Fornari[1,†] and Barbara Re[1,†]

[1]*University of Camerino, School of Science and Technology, Computer Science Department, Via Madonna delle Carceri, 7, Camerino, Italy*

Abstract

Leveraging BPMN models for validating novel approaches and tools is a common practice among researchers in the BPM field. To ensure the results validation accuracy of the such research activities, it is important to have an in-depth understanding of the sample models' features. This mitigates the avoidance of any biased factors that could potentially impact the validation process. In this work, we present BPMN Inspector, a web application designed to streamline the inspection process of BPMN models. The inspection process effectively distinguishes various model types (i.e. collaboration, choreography, and conversation) while eliminating the need for manual effort in identifying duplicate models, validity issues, and non-English models. In addition, BPMN Inspector provides detailed insights into the collection of models by investigating the usage of BPMN notation elements, their combinations, syntactic violations of the standard and the adherence to established good modeling practices.

Keywords

Business Process Model and Notation, Notation usage, Model validation, Validation practices.

1. Introduction

BPMN has been widely acknowledged as the standard modeling notation for designing business process models. It holds the promise of providing a standardized *lingua franca* to represent business processes, which can be easily understood by all stakeholders involved in the business process life cycle [1]. Since its release, BPMN models have often been used to validate research activities in the BPM field. The selection of models used for validation plays a crucial role in determining the result of novel research approaches. In fact, the validation process of the research activities is directly influenced by the specific features of the models used. Understanding these features provides a solid foundation for a wide range of research activities. As an example, in [2], researchers rely on precise and well-defined features of BPMN models to develop formal methods and verification techniques. Similarly, in studies that focus on the BPMN language

Proceedings of the Demonstration Resources Track, Best BPM Dissertation Award, and Doctoral Consortium at BPM 2023, BPM 2023, Utrecht, Netherlands, September 11-15, 2023

*Corresponding author.

†These authors contributed equally.

ivan.compagnucci@unicam.it (I. Compagnucci); flavio.corradini@unicam.it (F. Corradini); fabrizio.fornari@unicam.it (F. Fornari); barbara.re@unicam.it (B. Re)

0000-0002-1991-0579 (I. Compagnucci); 0000-0001-6767-2184 (F. Corradini); 0000-0002-3620-1723 (F. Fornari); 0000-0001-5374-2364 (B. Re)

[3, 4], understanding the features of BPMN models becomes crucial for analyzing and interpreting the collected data to analyze trends and usage of the notation. Therefore, a reliable benchmark is needed that outlines the characteristics of the models.

In this paper, we introduce BPMN Inspector, a web application designed to extract features from a collection of BPMN models. By examining and interpreting the information obtained from the models, researchers can improve the efficiency and effectiveness of their research validation process, thereby ensuring the reliability and accuracy of their research findings. Additionally, the tool provides users with the opportunity to access detailed reports. These reports offer comprehensive and structured information about BPMN models, catering to the specific needs of educators, researchers, and standardization bodies. Users can tailor their analyses based on their research objectives, teaching requirements, or standardization goals to obtain relevant implications and suggestions.

2. BPMN Inspector Main Functionalities

BPMN Inspector is made available through a RESTful web application, which enables users to upload a set of BPMN models, apply optional filters, and explore their features by performing a set of analyses. The process of inspecting the model collection is depicted in Figure 1 and can be divided into three main phases: *BPMN Collection Upload*, *BPMN Collection Filtering*, and *BPMN Collection Analysis*.

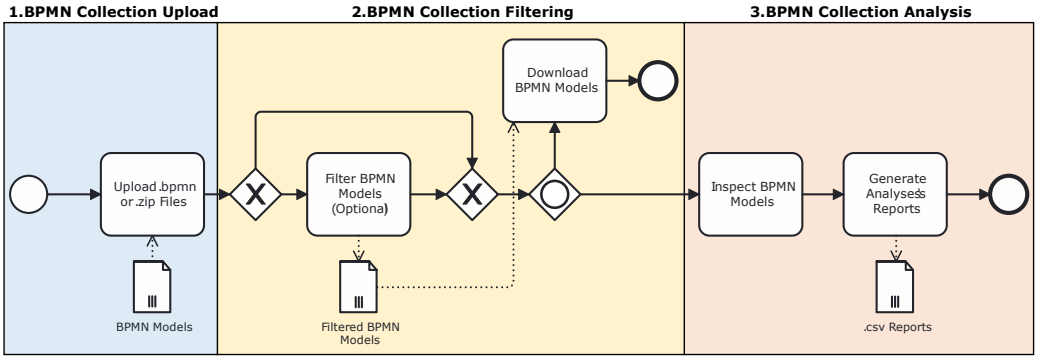

Figure 1: BPMN Inspector usage phases.

BPMN Collection Upload. This is the initial phase where the collection of models is uploaded. One or more BPMN models can be uploaded in file formats such as *.bpmn* or compressed into a *.zip* archive. If the file is in *.zip* format, the content is extracted, and the extensions of the file are examined to exclude any file that does not conform to the *.bpmn* format.

BPMN Collection Filtering. After uploading the models, the preview window is shown. Figure 2 depicts the preview page that allows users to filter and download the models as a *.zip* file, or inspect the collection by applying optional filters. Each model presents a set of general information such as name, model type (i.e., Process Collaboration, Choreography, or Conversation), whether the model is valid, whether it is duplicated within the collection,

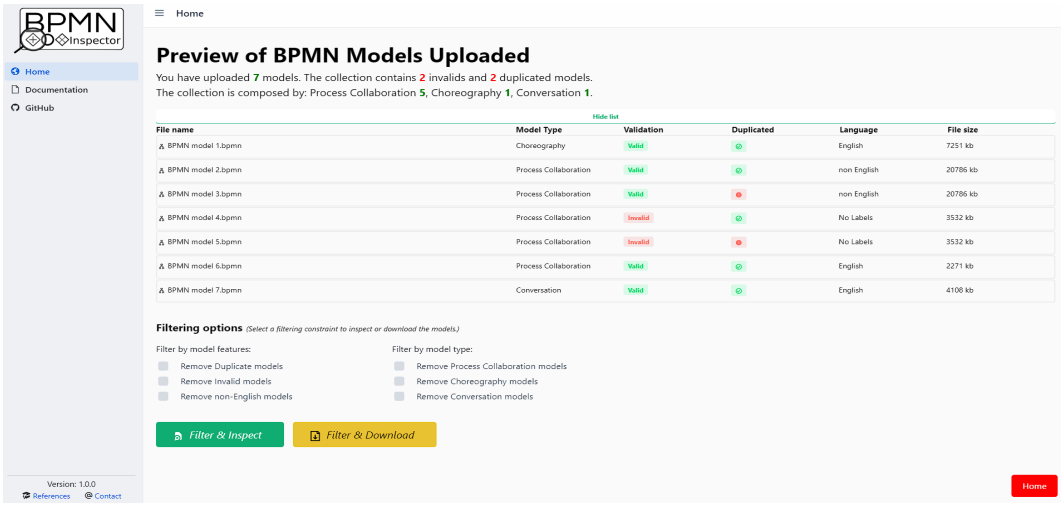

Figure 2: From BPMN collection upload to BPMN Collection Filtering.

whether the label's model is in English language, and the file size. The collection can be filtered based on two aspects: model features or model type. Regarding model features, the following filters can be applied: (i) removal of duplicate models; (ii) removal of syntactically invalid models; and (iii) removal of models whose labels are in a language different from English. Duplicate models are identified by reading and comparing all the bytes from each model's file. To determine syntactically invalid patterns, BPMN Inspector compares each model with the standard BPMN XML schema provided by the OMG. The language of the models is extracted using "Lingua",[1] a natural language detection library for Java. Considering the model's type, they can be filtered differently based on Process Collaboration, Choreography, and Conversation models. The model's type is detected by analyzing the XML definition of each *.bpmn* file.

BPMN Collection Analysis. This is the main phase of the tool and consists of a set of analyses performed on the collection of BPMN models previously uploaded. The inspection process involves four main analyses: *BPMN Element Usage*, *BPMN Element Combined Usage*, *BPMN Syntactic Validation*, and *BPMN Good Modeling Practices*. In the following, we report a description of each analysis.

BPMN Element Usage. In this analysis, each model of the collection is analyzed by counting the occurrences of each BPMN element. BPMN Inspector is able to distinguish and count up to 267 different BPMN elements [4]. Each element is treated as a separate entity based on its unique combination of markers. The analysis generates a *.csv* report providing detailed information about the element usage enabling a dynamic generation of the graphs. The first graph represents the model size, indicating the total number of elements used in each model's design. It offers a comprehensive view of the collection's complexity. Additionally, a practical complexity graph shows the number of distinct types of elements used in each model, providing valuable visual insights into the variety of BPMN elements utilized across the collection. The remaining analysis, focus respectively on the number of occurrences of each BPMN element in

[1] https://github.com/pemistahl/lingua

the collection, providing insights into the frequency of element usage, and the distribution of elements across the collection, indicating the total number of models containing each element.

BPMN Element Combined Usage. This analysis aims to identify possible relationships between the BPMN elements within the collection of models. BPMN Inspector performs an analysis of both pairs and groups of BPMN elements to determine the most frequently used combinations. To gather information on possible combinations, the tool calculates the Pearson correlation coefficient [5] for each pair of elements. This coefficient measures the strength of the relationship between two elements. The correlation index ranges from -1 to 1, where 0 indicates no correlation, 1 suggests that the elements are used consistently together, and -1 indicates that the elements are used alternatively. The analysis also explores groups of elements, starting with pairs of elements that are frequently used together. Then it expands the groups by adding the element that is most commonly used in combination with that set. This process iterates to automatically generate sets of elements. The tool calculates the percentage of occurrence of these element sets within the model collection using a graph to represent these combinations. Finally, two reports are generated: one containing the Pearson coefficients for each pair of elements, and the second one containing the percentage values of the most used sets of elements.

BPMN Syntactic Validation. This analysis focuses on identifying and handling syntactic violations in the BPMN models. Therefore, it will be conducted only if invalid models were not excluded during the filtering procedure. To determine whether the models include syntactic errors, BPMN Inspector compares each model with the standard BPMN XML schema provided by the OMG. The analysis generates a report in the form of a *.csv* file, which includes the list of the analyzed models. For invalid models, the report also provides an array of errors associated with each model. In the web application, is shown a chart to display the types of errors and their frequency. Additionally, is defined a table, presenting the error codes and providing a link to the XMLdation wiki site.[2] This link directs users to a detailed description of each error.

BPMN Good Modeling Practices. This analysis investigates whether the collection of models adheres to a set of established good modeling practices, as defined by [6]. To perform this analysis, BPMN Inspector integrates a tool called BEBoP (understandaBility vErifier for Business Process models).[3] The analysis results in a *.csv* report that provides boolean values indicating whether the model meets each guideline or not. In the web application, a radar graph is shown to indicate the percentage of violations considering each good modeling practice.

3. Maturity of the Tool

BPMN Inspector was specifically developed as a supporting tool for investigating trends in the usage of the BPMN notation language, enabling research on BPMN development, adoption, usage patterns, and impact. It has been adopted in a study conducted by [3], and then in the Business & Information Systems Engineering journal, in the special issue titled "*The Impact of the Business Process Model and Notation*"[4].

Maturity of the Tool. The maturity of BPMN Inspector was assessed in [4] where the analysis involved 54,500 BPMN models obtained from seven online repositories. The evalua-

[2]https://wiki.xmldation.com

[3]BEBoP tool: https://pros.unicam.it/bebop

tion process of the tool involved the four analyses described above. Starting from the usage of notation elements, the findings revealed that certain subsets of the BPMN notation were predominantly used, emphasizing the gap between theoretical (i.e., the total number of elements in the modeling notation) and practical complexity [7] (i.e., the number of different element types used in BPMN models) of the BPMN language. Regarding the combined usage of elements, the study identified strong correlations among certain elements, while no inverse correlation was found. Furthermore, an examination of syntactic errors in the models revealed that the predominant errors were associated with the connection of objects that lacked source or target elements. Furthermore, the study examined the adherence to good modeling practices and found that designers generally don't follow them. These observations have led to several implications and suggestions for educators, researchers, and standardization bodies. Based on the insights from the study, it is recommended that educators focus on commonly used BPMN elements and combinations in their teaching. Researchers are encouraged to investigate the reasons behind the underutilization of certain BPMN elements. Finally, the study has highlighted the need for better explanations from standardization bodies.

4. Resources

The source code for BPMN Inspector is available at https://github.com/PROSLab/BPMN-Inspector. Additional information on the tool are available at http://pros.unicam.it/bpmn-inspector/. Screencast showing a case study is available at https://youtu.be/pLDZyC60oRY. Finally, a docker image of the tool is made available at https://hub.docker.com/r/proslab/bpmn-inspector.

References

[1] D. Bork, D. Karagiannis, B. Pittl, A survey of modeling language specification techniques, Information Systems 87 (2020).

[2] F. Corradini, F. Fornari, A. Polini, B. Re, F. Tiezzi, A. Vandin, A formal approach for the analysis of bpmn collaboration models, Journal of Systems and Software 180 (2021) 111007.

[3] I. Compagnucci, F. Corradini, F. Fornari, B. Re, Trends on the Usage of BPMN 2.0 from Publicly Available Repositories, in: International Conference on Perspectives in Business Informatics Research, LNBIP, volume 430 of *LNBIP*, Springer, 2021, pp. 84–99.

[4] I. Compagnucci, F. Corradini, F. Fornari, B. Re, A Study on the Usage of the BPMN Notation for Designing Process Collaboration, Choreography, and Conversation Models, Business & Information Systems Engineering (2023).

[5] J. Benesty, J. Chen, Y. Huang, I. Cohen, Pearson correlation coefficient, in: Noise reduction in speech processing, volume 2, Springer, 2009, pp. 1–4.

[6] F. Corradini, A. Ferrari, F. Fornari, S. Gnesi, A. Polini, B. Re, G. O. Spagnolo, A Guidelines framework for understandable BPMN models, Data Knowledge Engineering 113 (2018) 129–154.

[7] J. Erickson, K. Siau, Theoretical and practical complexity of modeling methods, Communications of the ACM 50 (2007) 46–51.