



Business Process Matching Analytics

Katalin Ternai , Szabina Fodor , and Ildikó Szabó ^(✉) 

Corvinus University of Budapest, Fővám tér 13-15., Budapest 1093, Hungary
{katalin.ternai, szabina.fodor,
ildiko.szabo2}@uni-corvinus.hu

Abstract. Organizations study competitors to gain insights into what extent their processes are compliant to processes of other ones. Business process models are usually described in documents and their elements can be extracted by semantic technologies such as process-based text mining method. Business analytics provides a powerful tool to analyze competitors' identified processes through dashboards. Educational institutions gain competitive advantage by establishing international partnerships. They must improve their processes to become more attractive to students. This paper presents a general method using semantic technologies to enhance the examination of competitors' processes and their application in internationalization strategies.

Keywords: Process matching · Semantic business process management · Business analytics · Text mining · Internationalization

1 Introduction

Organizations must be aware of how their competitors behave on the market. Usually, it is difficult to investigate this behavior directly and explicitly. News, speeches, studies about companies can be applied to detect behavioral symptoms on the market (such as acquisition strategies, digitalization endeavors and so on). Organizational and operational rules contain information about business processes fit to organizational strategies. Nowadays, the digital transformation wave poses effects on every knowledge-intensive organization and makes them committed to improve their processes and modify their operational rules. New ideas, best practices can be captured from these kind of published documents. Business analytics toolset is a powerful instrument to analyze competitors' operations through organizational documents. Nevertheless, it would also provide us opportunities to look into what processes have already been transformed by others and in what extent.

This study addresses the important questions of process analysis and investigates whether and to what extent processes of an organization are matched with processes of their competitors. Different approaches are available to reach this goal. Business process management tools have built-in functions to compare business process models with each other. Competitor analysis can be carried on literature reviews, secondary data analysis out, but these activities require mainly human efforts to solve the problem. However, our approach is based on processing organizational documents with semantic technologies such as process ontologies, because we assume that business process

models can be extracted from their descriptions with using semantic text mining algorithms [1, 2]. Our method (see in Fig. 1) starts with business process modelling phase in Adonis.

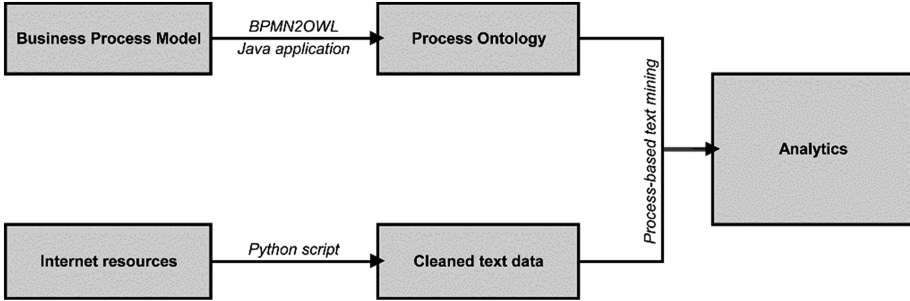


Fig. 1. Our method

In our approach the concept of process ontologies plays a key role on the scene of competitor analysis. Process ontologies have no precise definition in the literature. Some approaches refer to process ontology as a conceptual description framework of processes [3]. Process ontologies are abstract models from this point of view. Task ontologies determine a smaller subset of the process space, the sequence of activities in a given process. Meanwhile the domain ontologies focus on catching the essence of each object of the world and their connections. Process ontology identifies all the artefacts that describe a process, regardless of whether they are structured or not. It allows to build all process elements clearly and unambiguously, linked with the domain ontologies containing specific enterprise concepts.

Hence, they serve as an appropriate basis to capture implicit and explicit semantics of process models hidden in documents. Transformation of business process models into process ontologies makes it possible to process documents in semantic manner. XSLT transformation was used in our previous work, but it ensured us just a semi-automatic method, hence a new Java program was created to make this transformation more automated.

After the model transformation a process-based text mining stage in Python is responsible for identifying process elements in the documents. Text mining is usually considered as a specialized area of data mining but sometimes it just enhances data preparation steps in data mining projects. Its main purpose is to identify patterns within texts. Different approaches are differentiated based on the objective of text mining process and the nature of used methodology. General text mining process contains steps of preparing corpus, pre-processing texts, generating and selecting feature which are followed by data mining steps and interpretation of results [4]. Pre-processing methods usually include collecting multiple documents, tokenizing text contents into individual terms/words, eliminating stop words (e.g. pronouns), identifying root/stem of words and using statistical methods for calculating TF-IDF (term frequency-inverse document frequency) to determine the importance words in collections [5]. Text

clusters can be created based on the strengths of relationships between extracted expressions. Ontology learning generates elements of domain ontologies from various kinds of resources with applying natural language processing and machine learning techniques [6]. It relies on statistical, rule-based or mixed methods [7]. According to Kő and Gillani [8] process models provides the contexts, but domain ontologies are applied to extract knowledge by Our process-based text mining method uses preprocessing techniques, n-grams as most likely connected word pairs and the structure of process models to create a dashboard for visualizing the results of process matching. All phases will be described in more detail during presentation of our business case.

This paper presents this general method using semantic technologies to enhance the analysis of competitors' process in the respect of a given business process. Having presented our method, the second section introduces related theoretical works in the field of semantic business process modelling. The third section presents how our method works in practice. The case of the research grant application process was selected to illustrate the applicability of this method as a proof-of-concepts. Limitations and future research steps are highlighted in the fourth section.

2 Related SBPM Researches

The usage of Semantic Web technologies like reasoners, ontologies, and mediators add a completely new viewpoint to business process management. This approach is known as semantic business process management (SBPM) [9]. Process mining is one of its related subdomains.

Process mining deals with data of actual process execution stored in event logs, transactional data etc. "to discover process models, check the conformance of process models to reality and extend or improve process models" [3].

Three common classes of process mining techniques can be summarized as follows:

- **Process Analysis.** The goal of process analysis and process monitoring is to monitor process runs and to analyze their executions with using business intelligence tools. It supports business analysts in identifying deviations in processes and corrective measures to redesign suboptimal processes.
- **Semantic Process Mining.** Process mining includes techniques to extract process models from logs. It focuses on the automatic discovery of information from event logs without a predefined model. The importance of process mining in BPM is widely acknowledged as an important and unavoidable analysis tool to aid the (re-) design and the (re-)configuration of process models. Current process mining techniques are already quite powerful and mature. However, the analysis they provide are purely syntactic.
- **Cross-Organizational Conformance Checking.** Conformance Checking refers to algorithms for verifying whether logs follow the predefined behavior expressed in process model or not. These algorithms require the process model and its instances as well. The main advantage of ontologically defined process instances and models is that they improve the interoperability between information systems.

Organizations aspire to learn from others on how they adapt their processes towards improvement [10]. Process benchmarking, however, is mainly a manual process, requiring the involvement of experts to collect and interpret process-related data [11]. A main problem is that processes are often modeled on different levels of granularity.

Several approaches were elaborated to combine Business Process Management with Semantic Web technologies [12, 13].

In the context of BPR, organizations compare business process models to identify operational correspondences and differences. The approach for measuring the degree of similarity considers linguistic and behavioral aspects of process models to calculate a degree of similarity [14].

There are other works on measuring similarity between semantic business process models. A business process may be modeled in different ways by different modelers utilizing the same modeling language. An appropriate method for solving ambiguity issues in process models caused by the usage of synonyms, homonyms or different abstraction levels for process element names is the use of ontology-based descriptions of process models [15]. This method describes high-level Petri nets in OWL (Web Ontology Language) [16].

Our approach focuses on extracting information from documents instead of logs. Our process-based text mining uses a generic algorithm - extracting business processes from documents using predefined process structures as heuristics. Similarity measures of process models will be useful metrics in a later phase of this research.

3 Case Study: Research Grant Process

Higher education institutions want to rise to higher place in global rankings to become more famous and attract more international students. Internationalization is getting paid more attention because it provides competitive advantages on national and international scene. The number of participations on international research and conferences reflect the intensity of scientific activities at a university. The Research Committee's mission is to create and strengthen enabling conditions for research activities at the international academic level. The research grant application process at our university provides a good basis to apply our method to it and analyzing its matching with the process of other educational institutions. Application time period for this grant is continuous, the evaluation of any complete applications shall be made in five working days. Automatic acceptance is given up to the budget frame, if the application meets the requirements of the Research Committee.

3.1 Business Process Model

In the Process modeling phase the previously described business process have been implemented on the BOC ADONIS platform¹. The business process model, the working environment, the document model and the IT system model have been

¹ ADONIS is a BPM tool created at the University of Vienna.

specified. The logical shell of the business process model with the core objects (e.g. task) has been created. The input and output data, the IT system information and the responsible role from the organogram have been linked to the activities (Fig. 2).

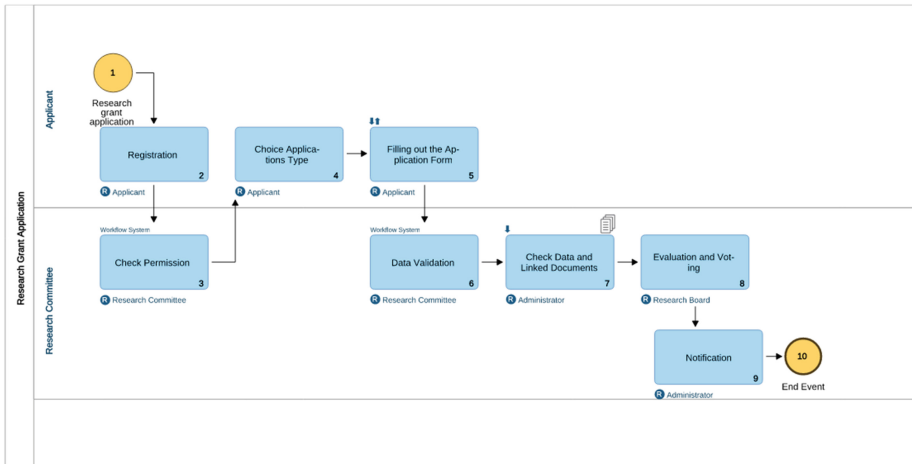


Fig. 2. The BPMN model of our case study

3.2 Transforming BPMN Models (BPMN to OWL)

To represent the business model in the ontology, the representation of ADONIS model language constructs and the representation of ADONIS model elements have to be differentiated. ADONIS model language constructs are created as classes and properties and the BPMN model elements can be represented through the instantiation of these classes and properties in the ontology. The linkage of the ontology and the ADONIS model element instances is accomplished by the usage of properties.

For the transformation process, a prototypical software tool was developed, which transforms a BPMN2.0 into an OWL format. The resulting file contains a partial ontology including classes and individuals of the input file. Every node of the BPMN diagram represents a class and has a parent and can have multiple attributes. All nodes are classes. The top-level class is **owl:Thing** and contains the six child-classes *Documents*, *IT_system*, *Roles*, *Research_Grant_Application*, *Start_* and *End_Event*. The *Research_Grant_Application* class contains the task elements of the BPMN 2.0 specification. The *Check_Data_and_Linked_Documents* class is one example of BPMN nodes. The developed BPMN ontology contains BPMN elements with their attributes and model associations (see Table 1).

Table 1. Overview of BPMN-OWL correspondence

BPMN	OWL	
	Structure	Current value
OWL classes		
process	super class of tasks	Research_Grant_Application
	subclass of <i>Start_Event</i>	
task	subclass of process	Check_Data_and_Linked_Documents Check_Permission Choice_Application_Type Data_Validation Evaluation_and_Voting Filling_out_the_Application_Form Notification Registration
adonis:target C_ROLE class	subclass of <i>Roles</i>	Administrator Applicant Research_Board Research-Committee
adonis:target C_DOCUMENT class	subclass of Documents	Application_Form Cost_Calculation Travel_Form
adonis:target C_INFRASTRUCTURE_ELEMENT C_SERVICE class	subclass of IT_system	Workflow_system
OWL object properties		
sequenceFlow	followed_by owl:ObjectProperty	
adonis:relation RC_REFERENCED_INPUT_DATA_OPTIONAL_M	input_data owl:ObjectProperty	
adonis:relation RC_REFERENCED_OUTPUT_DATA_OPTIONAL_M	out_data owl:ObjectProperty	
adonis:relation RESPONSIBLE_FOR_EXECUTION	responsible_for_execution owl:ObjectProperty	
adonis:attribute A_ORDER	order owl:ObjectProperty	

3.3 Process-Based Text Mining

The first step of our text mining process was the corpus creation. Descriptions about research grant application processes published by educational institutions were collected by a Python² scraper. The script uses beautifulsoup, urllib, re, time Python libraries to download hits provided by the Bing search engine. ‘Conference’, ‘Grant’, ‘Funding’, ‘University’ keywords were applied to identify related descriptions published on the Internet. Ten other announcements were downloaded manually besides this repository. After reviewing this collection containing 57 descriptions, we detected that our searching method and keywords would be refined to find more relevant hits in the future. At the end, the repository was separated into two datasets with 11 and 8 elements. These datasets contained announcements published by universities in the US and outside US to facilitate regional analysis as well.

² Python Programming Language: <https://www.python.org/>.

Text pre-processing steps were executed in the second stage. Process ontology resulted from the BPMN to OWL transformation was handled by our Python script uses owlready2 Python package. Class names of the process steps, documents and roles were collected from this ontology and textblob functions were used in preprocessing the corpus. Bigrams meaning a sequence of two adjacent elements were extracted from each description. The name of process ontology elements was split into words and lemmatized and the collections sets of n-grams were filtered by these terms. For example, in the case of Application Form document, all bigrams containing at least application or form word were gathered. This procedure was iterated in the case of each process steps, documents and roles. Having created these lists, the intersection of remained bigram lists of a given process step and a document was determined to generate and select feature e.g. ['evaluation', 'application'] bigram was identified by the 'Evaluation and voting' process step and by the 'Application form' document. All bigrams from the intersection list, their related process step and document were stored in a CSV file in the case of each announcing institutions. This step was repeated for process steps and roles as well.

3.4 Analytics

Having executed these pre-processing and feature generation procedures in the case of datasets about universities of the US and universities outside the US, Tableau³ business analytics tool was used to discover interesting information about the research grant application process of each institution. Corvinus Business School operates an online application platform⁴ to make this process more flexible. One aim of this process matching was to detect the degree of automatization at each university.

Three charts were created to investigate which documents are required, which the executive roles are and which documents are transmitted in each process steps at our competitors. A dashboard was composed from these diagrams to provide a toolset for stakeholders. All charts can be filtered on a given university. The following figure presents that application form is transmitted during this process in the case of five universities. After filtering data on Indiana University Bloomington (IUB) we can see that they use application website to enhance this process and a board makes the decisions. At Corvinus Business School online application platform is used and the Research Committee deals with the internal grant applications. These findings indicate us that the process of IUB might be similar to ours, it is worth checking later (Fig. 3).

We do scrutinize the following chart (see A part of Fig. 4) to gain deeper insights into what extent the processes of our competitors are automatized. It reveals that Cornell University requires PDF form and DAAD Office prefers online application. These universities are from US and we need to examine the cases of universities outside the US to get a broader overview about this area (see B part of Fig. 4). The results of this process-based text mining do not reveal any signs of automatization in the case of non-US universities.

³ Tableau: <https://www.tableau.com/>.

⁴ Corvinus Business School Online Application platform: <http://workflow.business.research.uni-corvinus.hu/>.



Fig. 3. Dashboard to analyze competitors' processes

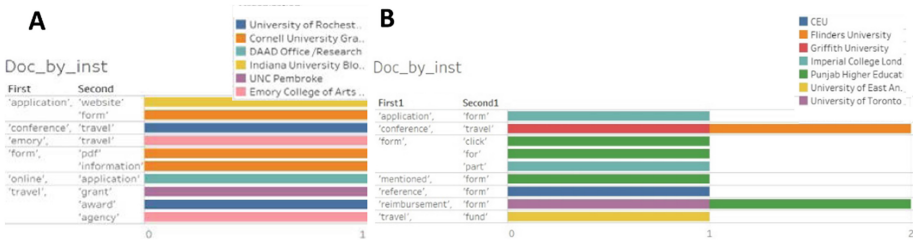


Fig. 4. Transmitted documents per US and outside US university

4 Limitations and Future Research

This paper highlighted a solution to analyze competitors' processes extracted from business documents with using BPMN to OWL Java-based transformation program and process-based text mining. A general algorithm was developed to transform business process models into process ontologies. It was written in Java and it can be parameterized hence it replaces the XSLT transformation in a more sophisticated way.

BPMN elements and their connected OWL model elements were also presented to enhance future development in this field.

A text mining algorithm written in Python showed how competitors' document can be processed based on a reference process model.

This research in the current stage has several limitations. Corpus collected by Bing search engine needs to be refined, cleansed in the future. Same words in the name of different process elements result the duplicates of bigrams - e.g. ['application', 'form'] was identified by the Choice Application Type process step and the Application Form document as well – hence the Count Distinct function was applied to present data on charts. Nevertheless, this paper presented that the concept is acceptable and applicable. However, it is necessary to improve this method to make competitor analysis more effective, accurate and automatized. We are planning to modify our corpus collection method, the names of process elements, validate our BPMN to OWL transformation procedure, extend our text mining algorithm with semantic-based relationship mining [1, 2] or SVM-based algorithms.

References

1. Szabó, I., Ternai, K.: Semantic audit application for analyzing business processes. In: Tjoa, A.M., Xu, L.D., Raffai, M., Novak, N.M. (eds.) CONFENIS 2016. LNBP, vol. 268, pp. 3–15. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-49944-4_1
2. Ternai, K., Szabó, I.: Semantic application for the internationalization audit of higher education institutions. In: Kő, A., Francesconi, E. (eds.) EGOVIS 2016. LNCS, vol. 9831, pp. 194–205. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-44159-7_14
3. Haller, A., Marmolowski, M., Gaaloul, W., Oren, E.: A process ontology for business intelligence. DERI, Technical report (2008)
4. Hashimi, H., Hafez, A., Mathkour, H.: Selection criteria for text mining approaches. *Comput. Hum. Behav.* **51**, 729–733 (2015). <https://doi.org/10.1016/j.chb.2014.10.062>
5. Vijayarani, S., Ilamathi, M.J., Nithya, M.: Preprocessing techniques for text mining-an overview. *Int. J. Comput. Sci. Commun. Netw.* **5**, 7–16 (2015)
6. Haase, P., Völker, J.: Ontology learning and reasoning—dealing with uncertainty and inconsistency. In: da Costa, P.C.G., et al. (eds.) URSW 2005-2007. LNCS (LNAI), vol. 5327, pp. 366–384. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-89765-1_21
7. Zhou, L.: Ontology learning: state of the art and open issues. *Inf. Technol. Manage.* **8**, 241–252 (2007). <https://doi.org/10.1007/s10799-007-0019-5>
8. Ko, A., Gillani, S.: Ontology maintenance through semantic text mining: an application for it governance domain. In: Innovations, Developments, and Applications of Semantic Web and Information Systems, pp. 350–371 (2018). <https://doi.org/10.4018/978-1-5225-5042-6.ch013>
9. Koschmider, A., Oberweis, A.: Ontology based business process description. In: Proceedings of the CAiSE, pp. 321–333 (2005)
10. Rolstadås, A.: Performance Management: A Business Process Benchmarking Approach. Chapman & Hall, London (1995)
11. Teuteberg, F., Kluth, M., Ahlemann, F., Smolnik, S.: Semantic process benchmarking to improve process performance. *Benchmarking Int. J.* **20**(4), 484–511 (2013)

12. van Dongen, B., Dijkman, R., Mendling, J.: Measuring similarity between business process models. In: APCCM 2007 Proceedings of the Fourth Asia-Pacific Conference on Conceptual Modelling, Australia, pp. 71–80 (2007)
13. Ehrig, M., Koschmider, A., Oberweis, A.: Measuring similarity between semantic business process models, pp. 1–4 (2007)
14. Maedche, A., Staab, S.: Measuring similarity between ontologies. In: Gómez-Pérez, A., Benjamins, V.R. (eds.) EKAW 2002. LNCS (LNAI), vol. 2473, pp. 251–263. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-45810-7_24
15. Kalogeraki, E.-M., Apostolou, D., Panayiotopoulos, T., Tsihrintzis, G., Theocharis, S.: A semantic approach for representing and querying business processes. In: Tsihrintzis, G.A., Virvou, M., Jain, L.C. (eds.) Intelligent Computing Systems. SCI, vol. 627, pp. 87–114. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-49179-9_4
16. Bakhshandeh, M., Pesquita, C., Borbinha, J.: An ontological matching approach for enterprise architecture model analysis. In: Abramowicz, W., Alt, R., Franczyk, B. (eds.) BIS 2016. LNBIP, vol. 255, pp. 315–326. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39426-8_25