EACH | USP LESTE
Escola de Artes, Ciências e Humanidades
Universidade de São Paulo

Graduate Program in Information Systems
PPgSI

6th PPgSI's Dissertations Workshop 2019

Similarity Between BPMN 2.0 Business Processes Models with Loops and the Full Set of Elements

Author:	Carolina Lima de Melo			
Advisor:	Marcelo Fantinato, PhD.			
Co-advisor:				
Research lines:	[X] Systems Development and Management		[ ] Systems Intelligence	
Research areas:	[ ] Database [X] Information technology management	[ ] Software engineering [ ] Human-Computer Interaction	[ ] Artificial intelligence [ ] Pattern recognition	[ ] Graphics processing [ ] Optimization
Application areas:	[X] Enterprise environments / Business processes [ ] Economy [ ] Games / Serious games [ ] Health	[ ] Bioinformatics [ ] Education / Distance learning [ ] Linguistics / Natural Language [ ] Other Which? _____	[ ] Biometrics [ ] E-government [ ] Cheminformatics	[ ] Mobile devices [ ] Internet / Social Networks [ ] Robotics [ ] General*
Period in the program (at the workshop date):	[ ] 2nd semester	[X] 3rd semester	[ ] 4th semester	[ ] 5th semester
Qualifying:	[ ] Qualifying held in: dd/mm/yyyy		[ ] Plan for qualifying in: 01/11/2019	
Defense:	Deadline for deposit: 01/02/2021		Plan for defending in: 01/10/2020	
Publications associated with the master's project:	• No publications to date			

EACH | campus capital USP LESTE
Escola de Artes, Ciências e Humanidades
Universidade de São Paulo

Graduate Program in Information Systems
PPgSI

The research project summary

Context:

Organizations may want to compare process models to identify operational similarities and differences for different reasons. For example, when organizations merge, they need to determine and resolve the differences between their business processes. An organization may also need to verify whether its business processes conform to the organization's standard or an industry standard. In addition, similarity measures between process models encourage the reuse of model stretches and facilitate the search for similar models in process model repositories or collections. Different characteristics of process models can be considered when measuring the similarity between process models. Similarity techniques can rely on these characteristics, alone or together. Schoknecht, Thaler and Fettke use the term dimensions to refer to these characteristics or perspectives and highlight three main dimensions: natural language, structural and behavioral.

Research problem:

A common technique to measure structural similarity is based on the *edit distance* between the graphs that represent the process models by calculating the number of operations required to transform one graph into another. However, graph matching algorithms may suffer from high computational complexity. Moreover, the vast majority of the structural similarity techniques proposed in the literature is specific to processes modeled through Petri nets, although BPMN is widely used in organizations. BPMN notation has many more elements than the ones commonly shared among most notations. BPMN 2.0 totals over 100 variations of elements, which are responsible for its high semantic expression power. When specific BPMN elements are used, generic techniques cannot be applied. Currently in her work being finalized in PPgSI, Nunes et al. proposed a process comparison technique that addressed subprocesses at multiple levels to give stakeholders a more detailed view and allow them to define which levels are more important for similarity analysis. This technique doesn't support the full set of BPMN 2.0 elements and process models with loops.

Research objective:

This project aims to extend the solution proposed by Nunes et al. to measure the structural similarity between BPMN process models by:

- Incorporating advanced notational elements into the technique to support the full set of BPMN 2.0 elements.
- Addressing loops in the process.

Characteristics of the proposed solution:

Nunes et al. extended to support BPMN the generic similarity technique proposed by Kacimi and Tari, which relies on *vectorial signatures* instead of graph *edit distance*. This technique was selected for its expandability and adaptability to broader contexts and for its potential for better efficiency. Although Kacimi and Tari's technique addresses structural and natural-language based similarity, only the former type was adapted to BPMN as this is the dimension most sensitive to the differences in element types and syntax rules among modeling notations, which is critical for BPMN. The target is to extend the proposed by Nunes et al., implementing a model that supports all elements of BPMN 2.0 and process models with loops without affect negatively the computational performance.

Theoretical foundations:

- **Business Process Model And Notation (BPMN)**: A business process model can be represented textually as a process description or graphically using different notations such as Petri nets, Event-driven Process Chain (EPC), Activity Diagrams of Unified Modeling Language (UML) and Business Process Model and Notation (BPMN). BPMN, currently in version 2.0, is the IT industry's preferred notation for modeling business processes through a highly expressive notation that enables easy communication between IT engineers and business analysts as well as other stakeholders.
- **Similarity between business process models**: The similarity between two business process models is a measure of likeness between both process models. Different characteristics of process models can be considered when measuring the similarity between process models, including the arrangement of elements in the model, which is called graph structure-based similarity (or simply structural similarity). Structural similarity relies on graph theory and has two main strands. The first one measures the similarity by considering characteristics of the overall graph structure, such as the size of the largest subgraph. The second one measures the similarity based on the position and type of the elements in the two process models. In addition to structural similarity, the other two most common similarity dimensions are those based on natural language and behavioral.
- **Kacimi and Tari's similarity technique**: Kacimi and Tari proposed a technique to measure the similarity between two process models by analyzing two dimensions: structural, using positions and type of elements in the process models, and natural language, using the syntactic analysis of activity labels. This technique was designed as notation-independent. The process models represented by graphs are decomposed into paths. All possible paths between the start node and the end node are extracted and stored in data structures called *vectorial signatures*. The extracted paths of each process model are then compared to each other using their *vectorial signatures*, two by two; and the similarity between every two paths is measured. Then, the overall structural similarity value between the two process models are calculated considering the similarity between path-pairs previously calculated.

Correlated works:

- R. Dijkman, M. Dumas, B. van Dongen, R. Käärik, J. Mendling, **Similarity of business process models: Metrics and evaluation**, *Information Systems* 36(2), pp. 498–516, 2011
- S. Ivanov, A. A. Kalenkova, W. M. P. van der Aalst, **BPMNDiffViz: A tool for BPMN models comparison**, *Demo Session of the 13th Int. Conf. on Business Process Management*, pp. 35–39, 2015
- F. Kacimi, A. Tari, **Vectorial signature for matching business process graphs**, *Proc. of the Int. Conf. on Advanced Networking Distributed Systems and Applications*, pp. 93–98, 2014

Validation

The evaluation will be based on real-world model collections and then we will compare against the technique proposed by Dijkman et al..

Limitations, risks, and threats:

The natural language-based and behavioral similarity dimensions are out of the scope of this project.

Scientific contribution:

As scientific contribution, this research proposes a technique that will improve the knowledge about process comparison, more specifically about BPMN structural similarity.

Technical contribution (if pertinent):

As technical contribution this research proposes an algorithm that will enable a better structural comparison of BPMN that will help stakeholders with a more detailed view about the compared flows.

The research method			
Genre (choose ONE)	[ ] Theoretical research	[ ] Practical research	[x] Empirical research [ ] Methodological research
Nature (choose ONE)	[ ] Basic research		[X] Applied research
Approach (choose ONE)	[ ] Quantitative research	[ ] Qualitative research	[X] Quali-quanti research
Literature review* (you can choose more than one)	[ ] Narrative review [ ] Descriptive review [ ] Scoping review	[ ] Meta-analysis [X] Qualitative systematic review [ ] Umbrella review	[ ] Theoretical review [ ] Realistic review [ ] Critical review
Main technical procedure (choose ONE)	[ ] Experimental research [ ] Bibliographic research [ ] Documental research [ ] *Ex-post-facto* research	[ ] *Survey* [ ] Case study [ ] Participatory research [ ] Research-action	[ ] Ethnographic research [ ] Grounded theory [X] Design science [ ] Other Which? _____
Data analysis (you can choose more than one)	[X] Descriptive statistics [ ] Inferential statistics	[X] Statistical test [ ] Content analysis	[ ] Discourse analysis [ ] Others: _____

* Definition of types of literature reviews established by Paré, G., Trudel M-C., Jaana M., Kitsiou, S. Synthesizing Information systems knowledge: A typology of literature reviews. In: Information & Management 52, p. 183-199, 2015. DOI: 10.1016/j.im.2014.08.008

Next steps:

- Literature review
- Development: Support the full set of BPMN 2.0
- Test: Support the full set of BPMN 2.0
- Development: Support process models with loops
- Test: Support process models with loops
- Evaluate the technique
- Dissertation Writing
- Divulgation