

Ontology-based Business Process Chaining in Heterogeneous Systems

Anbumunee Ponniah

BITS Pilani, K K Birla Goa Campus, India
p20170418@goa.bits-pilani.ac.in

Abstract. Large companies deal with complex business processes that span across multiple applications. Process Mining (PM) across the applications is not straight forward due to differing content, format, and context of the event logs. Ontology-based techniques can capture important metadata about applications and processes. The proposed work aims to use the relationship between metadata to identify processes that run across applications. We intend to apply the domain context for extending similarity measures within and across ontologies.

Keywords: Business Knowledge Representation · Domain-Specific Ontology · Event Logs · Knowledge Graph · Natural Language Processing · Ontology Similarity Measures · Process Mining

1 Research problem

Large companies manage many complex business processes that work together to achieve business goals [1]. The advent of IT enables business processes to implement in computers across multiple purpose-built applications [2]. Due to the evolving nature of the business, applications developed at different times by multiple teams follow diverse sets of standards and formats [3]. Standardising the event logs [4] requires expensive and disruptive coding and maintenance effort. This makes the discovery, analysis, and optimisation of such business process a technically challenging problem to solve [5]. Temporal, structural, and context-based techniques are available to analyse event logs to solve ordering, correlation, and scoping entries of the logs [6][7][8]. However, for large processes that run across heterogeneous applications, current techniques can only discover components in each application. There is a need for a Process Mining technique and a toolkit that factors in the number and variety of applications to discover and optimise complex business processes. We performed a literature survey using keyword-based search to extract candidate literature which was filtered for domain-specific application and manually grouped into general process mining, knowledge representation techniques, and intelligent decision systems. Listed below are the gaps identified by us with those addressed by our research project highlighted in boldface font.

1. Use of noise and outliers for context and detecting rarely occurring processes.

2. Modeling process drifts and overlaps of dynamically evolving processes.
3. **Mining from diverse logs that lack structural or transitional relationships**
4. Automating ontology enrichment, validation and downstream rule updates.
5. **Using domain context, Key Performance Indicators and cross-ontology relations for PM and ontology pruning**
6. **Extending distance metrics of ontologies to include domain and implementation-specific factors**

2 Approach

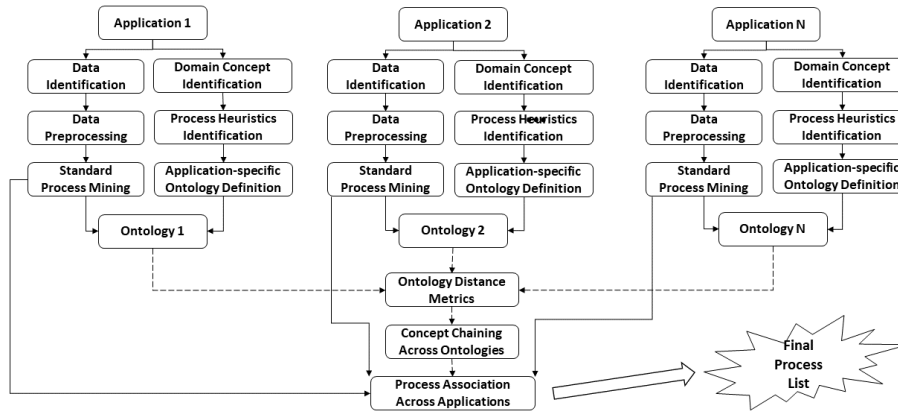


Fig. 1. Overall Research Design Illustrating the Sequence of Activities and Flow of Research.

Ontology-based techniques use metadata in the form of domain concepts or process descriptions [9][10] when identifying and sequencing tasks. Our research is primarily a software application problem that uses semantic web and ontology techniques to help process mining for complex processes. We propose to merge event-log based PM techniques with ontology creation and distance metrics-based concept association techniques to discover processes across applications. We further propose expanding the syntactic, structural and semantic similarity measurement techniques [11] to establish relationships between concepts within and across ontologies.

Figure 1 captures the flow of our research. Our research involves the analysis of event logs from each application separately using available PM techniques. We intend to create granular ontologies for each application combining pieces of discovered processes, domain concepts, application-specific concepts, and heuristics about the process. We plan to apply similarity measures to identify relationships

that span across the ontologies and derive process associations across heterogeneous logs.

3 Methodology

The proposed research will adapt Design Science Research (DSR) [12] for applying ontology-based techniques to mine complex processes. Per the ED Process [13], we propose to follow an iterative approach consisting of design, prototype, and evaluation for making our research available in a toolkit. We will use the mixed-methods research methodology for data collection from event logs (quantitative) and heuristic information from domain experts (qualitative). We will further use PM specific data preprocessing techniques [3] to cleanse, remove noise, and fix errors in data such as incomplete traces, missing events, mashed processes and many more.

4 Intended solution

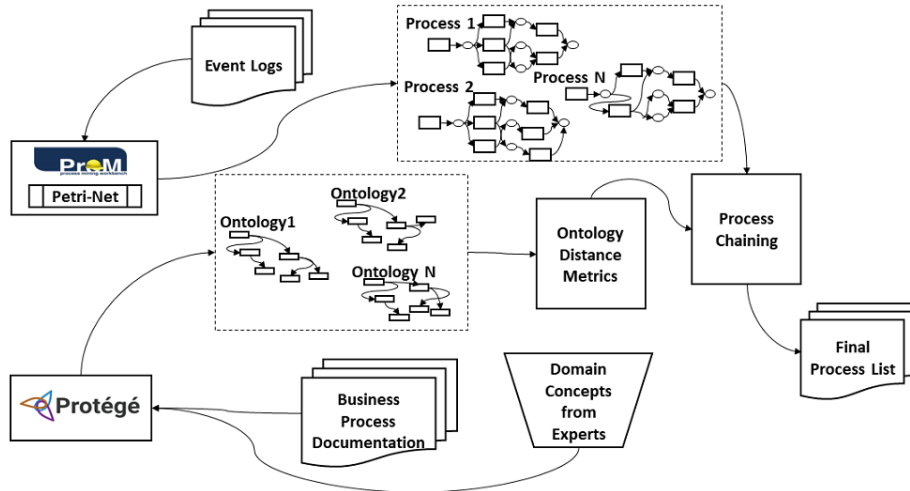


Fig. 2. A High-Level Framework Diagram for Overall Solution. Illustrating the Component Solution Architecture of the Research.

Figure 2 presents our solution as a combination of PM and Ontology techniques. The research project will discover processes from heterogeneous event logs in the chosen domain. We plan to use ProM as the PM tool due to its wide selection of discovery algorithms [14]. We will create ontologies to represent the domain context of the applications. We have shortlisted Protege as the tool for

ontology creation and analysis, due to its domain-friendly support for creating knowledge models [15].

Ontology-based technique have proven useful in PM [16][17]. Similarity measures is a common technique to find the relationship between concepts across ontologies [18]. We propose to devise PM specific distance and similarity measures to establish the relationship between the concepts within and across ontologies. We further plan to use a combination of concept-level and ontology-level relationships to correlate processes in heterogeneous applications.

5 Contributions to BPM research

The issue of PM across heterogeneous applications is not fully explored, which confirms our identified research gap. Our research focuses on arriving at similarity measures that work across multiple domain-specific ontologies. We propose to use those measures to discover and analyse complex processes that flow across heterogeneous applications. The techniques developed will also aid organisations [19] in change management, predicting disruptions and achieving better optimisation across their business. We plan to create a toolkit and sample implementations that would help future researchers in creating multiple domain-specific ontologies.

6 Project status and challenges

Current state of the research project is presented in Table 1.

Table 1. Illustrating the Current State of the Project and Next Phases along with their Tasks.

Phase	Task	Status
Literature survey and Analysis	Survey across PM, ontology techniques and intelligent decision systems.	Completed
Data Preparation	Identify, analyse and preprocess datasets.	In Progress
Process Mining	Baseline using standard PM techniques and tools	Pending
Ontology Creation	Creating and refining domain-specific ontologies with process-specific concepts	Pending
Establishing Similarity Measures	Exploring existing similarity measures and extending them to use PM and domain-specific features	Pending
Evaluation	Evaluating usefulness of ontologies and similarity measures using available techniques	Pending
Toolkit Development	Design and Implement Toolkit	Pending

Our work of applying ontology and distance measures to discover processes can help detect process drifts. It can also aid organisations in identifying overlaps and hidden relationships of processes across various applications. By adding KPI information, techniques developed in our research can be useful in automating downstream rule updates.

References

1. Adisa Delić, Sabina Donlagić Alibegović, and Mersiha Mešanović. The role of the process organizational structure in the development of intrapreneurship in large companies. *Naše gospodarstvo/Our economy*, 62(4):42–51, 2016.
2. Rodney McAdam and Daniel McCormack. Integrating business processes for global alignment and supply chain management. *Business Process Management Journal*, 2001.
3. RP Jagadeesh Chandra Bose, Ronny S Mans, and Wil MP van der Aalst. Wanna improve process mining results? In *2013 IEEE symposium on computational intelligence and data mining (CIDM)*, pages 127–134. IEEE, 2013.
4. XES Working Group et al. IEEE standard for eXtensible Event Stream (XES) for achieving interoperability in event logs and event streams. *IEEE Std 1849*, pages 1–50, 2016.
5. Suriadi Suriadi, Robert Andrews, Arthur HM ter Hofstede, and Moe Thandar Wynn. Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs. *Information Systems*, 64:132–150, 2017.
6. Marcos Rivas Peña and Sussy Bayona-Oré. Process mining and automatic process discovery. In *2018 7th International Conference On Software Process Improvement (CIMPS)*, pages 41–46. IEEE, 2018.
7. Edgar Tello-Leal, Jorge Roa, Mariano Rubiolo, and Ulises M Ramirez-Alcocer. Predicting activities in business processes with lstm recurrent neural networks. In *2018 ITU Kaleidoscope: Machine Learning for a 5G Future (ITU K)*, pages 1–7. IEEE, 2018.
8. Mahmoud AbdEllatif, Marwa Salah Farhan, and Naglaa Saeed Shehata. Overcoming business process reengineering obstacles using ontology-based knowledge map methodology. *Future Computing and Informatics Journal*, 3(1):7–28, 2018.
9. Ji-In Nam, Pawan Nagwani, Sae-Bom Jang, Young-Bin Shin, and Ho Jin. Ontology-based intelligent home assistance system. In *2016 IEEE International Conference on Consumer Electronics (ICCE)*, pages 121–122. IEEE, 2016.
10. TITI Sondes, Hadda BEN ELHADJ, and Lamia CHAARI. An ontology-based healthcare monitoring system in the internet of things. In *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, pages 319–324. IEEE, 2019.
11. Oussama El Hajjamy, Larbi Alaoui, and Mohamed Bahaj. Semantic integration of heterogeneous classical data sources in ontological data warehouse. In *Proceedings of the International Conference on Learning and Optimization Algorithms: Theory and Applications*, pages 1–8, 2018.
12. Roel J Wieringa. *Design science methodology for information systems and software engineering*. Springer, 2014.
13. Andrew Kusiak. *Engineering design: products, processes, and systems*. Academic Press, Inc., 1999.
14. HMW Verbeek, Wil MP van der Aalst, and Jorge Munoz-Gama. Divide and conquer: A tool framework for supporting decomposed discovery in process mining. *The Computer Journal*, 60(11):1649–1674, 2017.
15. Usha Yadav, Gagandeep Singh Narula, Neelam Duhan, Vishal Jain, and BK Murthy. Development and visualization of domain specific ontology using protege. *Indian Journal of Science and Technology*, 9(16):1–7, 2016.
16. AK Alves De Medeiros, Carlos Pedrinaci, Wil MP Van der Aalst, John Domingue, Minseok Song, Anne Rozinat, Barry Norton, and Liliana Cabral. An outlook

- on semantic business process mining and monitoring. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, pages 1244–1255. Springer, 2007.
17. Diego Calvanese, Marco Montali, Alifah Syamsiyah, and Wil MP van der Aalst. Ontology-driven extraction of event logs from relational databases. In *International Conference on Business Process Management*, pages 140–153. Springer, 2016.
 18. Wenlong Lu, Yuchu Qin, Qunfen Qi, Wenhan Zeng, Yanru Zhong, Xiaojun Liu, and Xiangqian Jiang. Selecting a semantic similarity measure for concepts in two different cad model data ontologies. *Advanced Engineering Informatics*, 30(3):449–466, 2016.
 19. Wil Van Der Aalst. Process mining: Overview and opportunities. *ACM Transactions on Management Information Systems (TMIS)*, 3(2):1–17, 2012.
 20. Ciza Thomas. *Ontology in Information Science*. BoD–Books on Demand, 2018.

6.1 Glossary of Terms

1. **Business Process:** A collection of tasks or activities performed by people or equipment to achieve a concrete goal in an organisation.
2. **Business Process Analysis:** Methodology to understand the health of different operations in a business.
3. **Business Process Chaining:** Establishing the relationship between parts of processes across system and application boundaries.
4. **Domain-Specific Ontology:** An ontology or series of ontologies that represent a set of concepts and relationships specific to the domain.
5. **Event Logs:** Digitized data about events executed by IT applications.
6. **Heterogeneous Applications:** Information Technology (IT) applications implemented using different technologies, following different standards and formats for data and logic.
7. **Intelligent Decision System (DSS):** A class of computerised information system that supports business and organisational decision-making activities through rules and learning algorithms.
8. **Intelligent Process Automation:** An emerging set of new technologies that combines fundamental process redesign with robotic process automation and machine learning. It is a suite of business-process improvements and next-generation tools that assists the knowledge worker by removing repetitive, replicable, and routine tasks.
9. **Ontology:** Ontology is the formal representative of concepts and their relationships. This enables computers and humans to interpret semantic relationships among the concepts and infers implicit knowledge [20].
10. **Process Mining (PM):** Approach and techniques to discover, monitor and improve real processes (i.e., not assumed processes) by extracting knowledge from event logs readily available in today’s (information) systems. PM includes (automated) process discovery (i.e., extracting process models from an event log), conformance checking (i.e., monitoring deviations by comparing model and log), social network/organisational mining, automated construction of simulation models, model extension, model repair, case prediction, and history-based recommendations [19].

11. **eXtensible Event Stream (XES)**: An IEEE specification for a tag-based language to capture event logs and event streams. Approved in Nov 2016 [4], the specification achieves interoperability in event logs to enable easier process discovery and analysis.