

Semantic Data Engineering: Methods, Applications and Practice



Dr. Kingsley Okoye
PhD Software Engineering
University of East London
United Kingdom

(Guest Editor)

Aims and Scope:

Modern tools for collection and analysis of data in all fields of Science, Technology and Business Management are providing more and more data with increasing complexity in their structure. These growing complexities are evidenced by the need for richer and more precise description of real-world objects that allows for flexible exploration of different events and data types. The challenge has been on how to create systems capable of providing platforms for event data exploration which derive understandable patterns as well as making the discovered patterns explicable. A very useful technique for pattern exploration is the Process Mining technique which has been successfully applied for classical mining of processes where each process execution is recorded in terms of events log sequences, and as a result, useful information about how activities depend on each other in a process domain has been made possible, and has proven to be essential for extracting models capable of creating new knowledge. A common challenge with most of the existing mining techniques is that they depend on tags in event logs information about the process, and therefore to a certain extent are limited, because they lack the abstraction level required from real world perspectives. Majority of the process mining techniques in literature are purely syntactic in nature, and to this effect, most of them fail desolately when confronted with unstructured processes. This means that these techniques do not technically gain from the real knowledge (semantics) that describe the tags in event log of the domain process.

In this proposed hot topic issue, the research intend to show that analysis provided by current process mining techniques can be improved by adding semantic information to event logs of the domain processes; using series of reliable experiments and case study of domain processes to illustrate the importance of the semantic rule-based approach. The issue will focus on ascertaining by a series of validation experimentations; how the result of learning process mining algorithms can be enriched through semantic representation of the deployed models and inference knowledge discovery. The issue will look at how ontologies with effective semantic reasoning can be used to lift process mining analysis from the syntactic level to a conceptual level through the semantic knowledge Engineering methods as well as other related techniques. Indeed, the semantic analysis takes advantage of the rich semantics described in event data of any

process domain, and links them to concepts in ontology in order to extract useful patterns by means of semantic reasoning. Semantic Reasoning is supported due to the formal definition of ontological concepts and expression of relationships that exist between event logs of the process. The method uses the semantics of the sets of activities within any of the domain processes to generate rules and events relating to task, to automatically discover and enhance the process model ontology through semantic annotation of the elements within a purposefully developed learning knowledge base. The issue intends to introduce the approach as means towards discovering and enrichment of the set of recurrent behaviors that can be found within the various process execution environments in order to solve some didactic issues and answer some questions with regards to different Learners behavior, the attributes they share amongst themselves within the process knowledge base. The methods will be put forward in order to address the problem of determining the presence of different patterns in process models. The standpoint for instance is based on the probe; "To what extent can references to the ontologies and effective raising of process analysis from the syntactic to semantic level enable real time viewpoints on the process domain, to help address the problem of analyzing data based on concepts and to answer questions about relationships the various objects (process instances) share amongst themselves within the knowledge-base". These semantic viewpoints is captured by controlling the elements in the systems based on two types of analyses (i) how to make use of the semantics that describes the available data? and (ii) how to mine the semantic information?

The proposed issue is expected to be beneficial and be used by professionals, engineers, entrepreneurs and process analysts in understanding their everyday processes and how to improve on them by maintaining a high compliance check and support. On the other hand, the issue will be used for academics purposes especially in Research and Higher Educations purposes, to inform and guide the work of staff and students about the latest state of the art in Semantic Data Engineering – which includes the Methods, Applications and Practice.

In summary, the hot topic issue is intended for use by Software Engineers, Software Application Developers or Designers, Business Project Managements, Academic staffs and research students and community.

At the core of the issue and presented works will be to help find answers to *how*:

1. a) Process mining can be applied to improve the informative value of various process domain data.
2. b) Describe how improved process models can be derived from the large volume of event logs within any process knowledge base.
3. c) Use of semantic representation of the derived models to enrich the results of process mining through semantic knowledge engineering and technologies
4. d) Use of ontologies with effective semantic reasoning to lift process mining analysis from the syntactic level to a much more conceptual level.

By tackling the motivations of this issue, the research intend to deliver means by which the objectives and focus of the semantic data engineering methods and application

contributes to the body of knowledge in current literature. In summary, the main contributions of the issue will be focused on:

1. a) Semantic motivated synchronization of event log formats for multiple process data.
2. b) Ontology driven search for explorative analysis of big data analysis and its process executions.
3. c) Techniques for annotating unlabeled activity sequences using business process model notations and ontology schema/vocabularies.
4. d) Use of semantics tools to manage perspectives of process mining algorithms and definition of methods towards discovery and enhancement of process model analysis.
5. e) Useful strategies towards development of process mining algorithms that are more intelligent, predictive and robotically adaptive.
6. f) Importance of semantics process mining to augment information value of data about a domain process