Journal of Network and Innovative Computing
ISSN 2160-2174 Volume 6 (2018) pp. 025-033


# Process Mining with Semantics: Real-time Application on a Learning Process Domain

**Kingsley Okoye**

School of Architecture Computing and Engineering, College of Arts Technologies and Innovation, University of East London, Docklands Campus, 4-6 University Way, London, United Kingdom. E16 2RD
*k.okoye@uel.ac.uk*

***Abstract*: Automated means for extraction, analysing or harmonization of various kinds of data that are stored in today's information systems - is indispensable to perform an effective process mining (PM). In view of that, this paper introduces a semantic-based process mining approach that is capable of detecting useful patterns or trends within any given data or process base. The work illustrates the method using a case study of the learning process domain. Essentially, the paper takes into account the context of the individual learners activities within a learning knowledge-base in order to find the best possible ways to efficiently realize (meta-analysis) the individual properties or attributes the process instances share amongst themselves within the knowledge-base. The goal is to identify patterns that have an effect on users performance and then respond by making decisions based on individual properties (assertions) and the classification process. Thus, the method of this paper is grounded on the *semantic modelling* and *process mining* techniques. Practically, the method uses the semantics of the captured events logs about the learning process and discovered models to create new knowledge that is applied for enhancement of the existing information knowledge-base. Theoretically, the work focus on augmenting the information values of the resulting process models based on the individual attributes (object and data properties) that are well-defined within an ontology. On one hand, in order to ensure validity, the work looks at the extent to which the individual process elements and harmonization is met. Whereas, reliability refers to the level of consistency in providing a well-suited inference mechanism or knowledge-base management system that is useful towards drawing valuable and/or accurate conclusions as a result of the improved method of process analysis.**

***Keywords*:** process mining, learning process, semantic annotation, ontology, process modelling, event logs.

## I. Introduction

The process mining (PM) [1][2][3] term have gained a significant interest over the past few decades. According to [3] the PM is regarded as a new field within the wider context of business process management (BPM) [2] that uses the data mining (DM) techniques to find out patterns or models from event logs, and predict outcomes through further analysis of the discovered models [1][2]. The PM methods provide useful information about how activities depend on each other within a process execution environment and have also made possible the need for extracting models capable of creating new knowledge from existing ones. Thus, the advanced branch of the PM (Semantic process mining) which helps in determining the semantics behind the tags or labels in an event log considered useful for discovery of new knowledge (or maybe required progressively through time) for improved process analysis and interpretation.

Accordingly, this paper employs the PM techniques to discover sets of recurrent behaviours or patterns that can be found within a learning process base. The method focuses on providing a semantic-based model and analysis framework and its implementation components that is based on the available events log to better support the process analysts and users. Technically, the method involves identification and modelling of data about the different learners (process instances), which in turn, supports the discovery and provision of a new and enhanced model - whilst ensuring the quality of the results and outcomes. Indeed, such metadata analysis (semantic) were necessary to determine suitable learning paths which can be used to address or answer some common question about the different users (process elements).

In fact, the proposed semantic-based PM method applies effective reasoning methods to make inferences over a learning process knowledge-base that leads to the automated discovery of learning patterns/behaviour. Moreover, the method consists of semantic *concepts/schema* that is used to model the events data logs and a *reasoner* that is realistically applied to reason over the underlying ontologies in order to create new abstract knowledge about the process elements. Thus, the term Semantic Learning Process Mining (SLPM). Basically, the conceptual means of analysis and the resultant learning models is what we use in providing new knowledge or information about the process in reality (or as performed). In other words, the concept of *semantic modelling* makes it possible to match same ideas as well as use the coherence and structure itself to inform and answer questions about relationships the process instances share amongst themselves within the learning knowledge-base. In turn, suitable learning patterns were determined through the automatic means of analysis or creation of workflows.

The rest of the paper is structured as follows: in Section 2, the work introduces the case study of the research process which we use to illustrate the semantic-based method throughout this paper. Section 3 presents the resulting learning model and how we extract the input data (i.e. events log) into minable executable formats for an improved process mining and analysis. In addition, the section describes how the work enables more effective reasoning and tactical strategies that are used for adaptation and decision-making purposes. Section 4, shows the various steps (procedures) we used for ample implementation of the SLPM approach, and how we semantically apply the representations for a semantic-based model analysis to draw conclusions and make predictions based on the analysis of the available dataset. Finally, Section 6 concludes the paper and draws a roadmap for future work.

## II. Use Case Study of the Learning Process

The learning process is very similar to undertaking a journey. The prerequisite to starting a journey is deciding where to go and which route or how to get there. A typical example of a learning process as used for the purpose of the work in this paper is the *Research Process*. The work demonstrates that a research process can technically be described as a *workflow* [2]. This workflow(s) constitute of the journey from choosing the research topic to completing the research (Figure 1) and comprises of practical steps or set(s) of activities through which must be performed (most often in a sequential manner) [3] in order to find answers to the research problems. Actually, the route/path to finding answers to the research questions is what makes up the entire process. Moreover, the workflow of those routes are not static, it changes as a learner travels along the research process. For instance, at each phase of the process, the researcher is required to choose from a variety of methods, procedures, or models which will help in achieving the research goals.

Therefore, the aim of this work in view of the aforementioned definition - is to adopt suitable tools and sets out to provide a method that is capable of creating useful knowledge and/or enhancement of the information values of the discovered learning models while ensuring validity and reliability of the outcomes.

Moreover, the work in [4] notes that a typical research process involves systematic, controlled, valid and rigorous exploration and description of what is not known as well as the establishment of associations and causations that permit the accurate prediction of outcomes under a given set of conditions e.g. as described in reference [5].

Perhaps, one of such tools that have the capability of enabling the non-trivial understanding of the various processes involved in modelling the learning process is the process mining [3]. According to [4], the main purpose is to decide, describe, justify and explain how the users (e.g. learners) go about finding answers to the research (learning) questions.

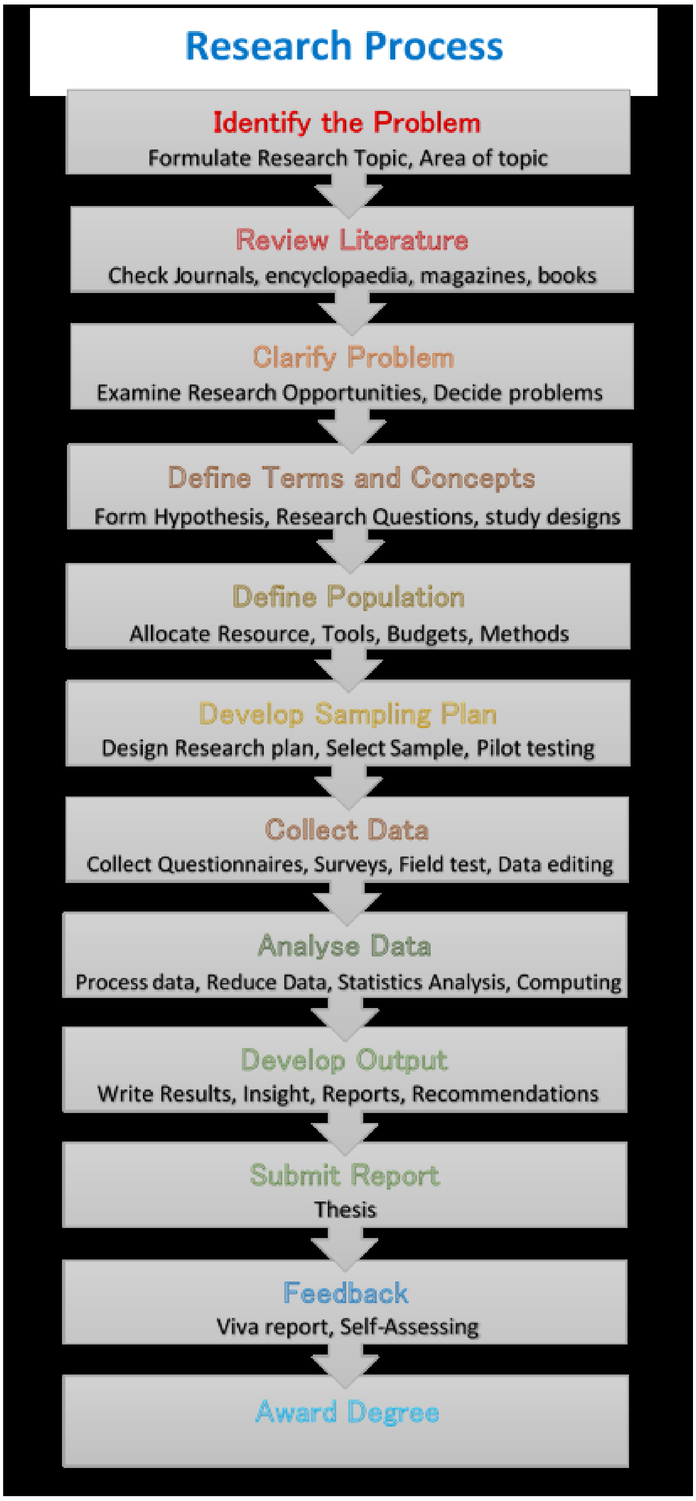


**Figure 1.** Control-Flow (workflow) of a Research Process [4]

As gathered in the figure (Figure 1) - the flow of the research process (i.e. learning activities) from the identification of the research problem to the award of the degree, consist of different learning steps which the learner has to or partly perform in order to complete the research process. The order in which the learning steps (referred to as *learning process units* [6]) are carried out has the capability of determining the time of completing the research or reliability of the research outcomes. Thus, the motivation of the method of this paper is focused on constructing accurate models that meet the researcher's goals or expected outcomes [7] based on the individual learning properties or behaviours. The work claims that a method which is capable of semantically providing guidance or integration of the different components will go a long way to improve the learning process models.

Accordingly, with reference to the learning process attributes as listed in the figure (Figure 1) - this work target to minimize the need and problems of analysing large volumes of data sets (quantity) with improved performance (quality) without affecting the learning outcome or information that are contained in the events logs.

Thus, to construct process transition and information about the learning activities concepts, we develop in this work a business process model notation (BPMN) as noted in [5] and explained in detail in Figure 2 in order to visualise the learning activities and the immediately preceding events that represents (map) the learning transitions.

Essentially, the learning transition (or sequence of activities) is based on four milestones, namely: Establish Context → Learning Stage → Assessment Stage → Validation of Learning Outcome as shown in Figure 3.

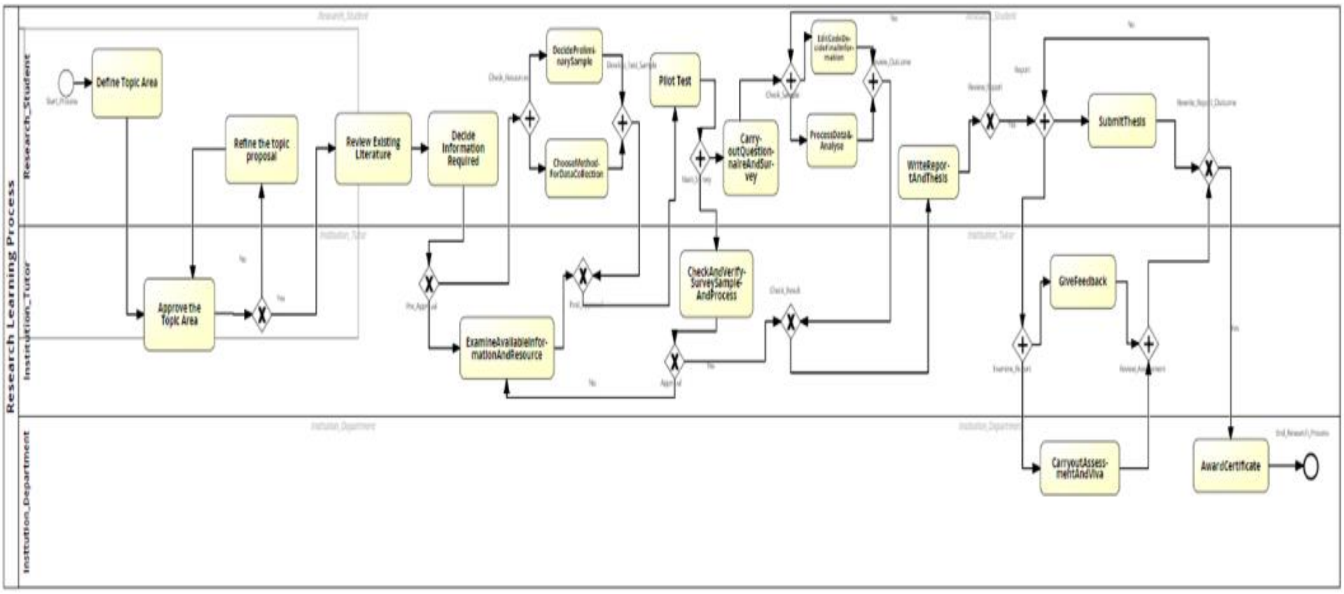


**Figure 2.** Example of Workflow of the ResearchProcessModel with BPMN in Bizagi Modeller

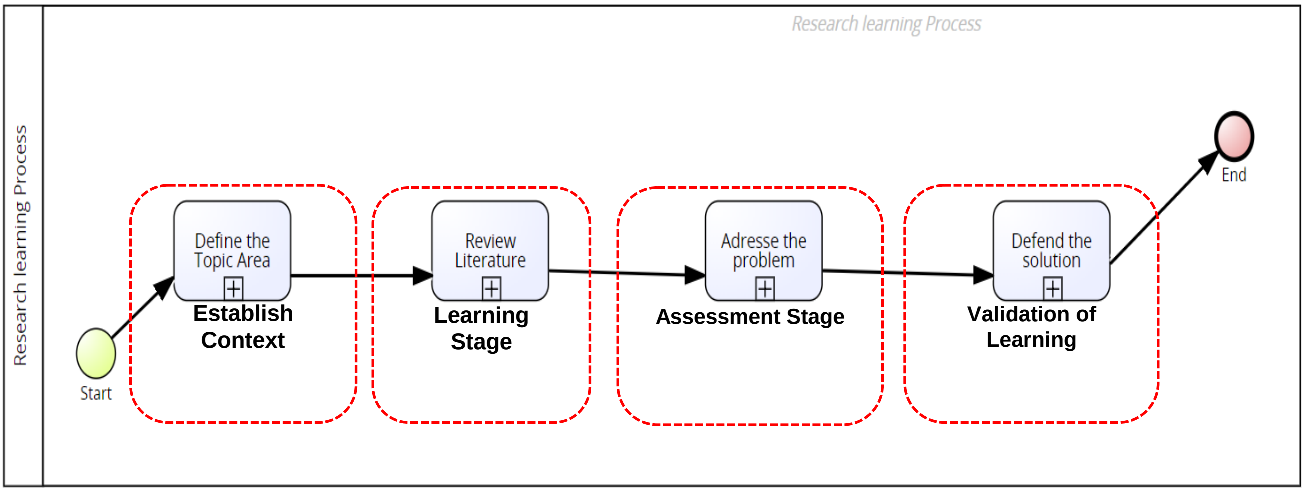


**Figure 3.** The Four Milestone of the Learning Process with BPMN

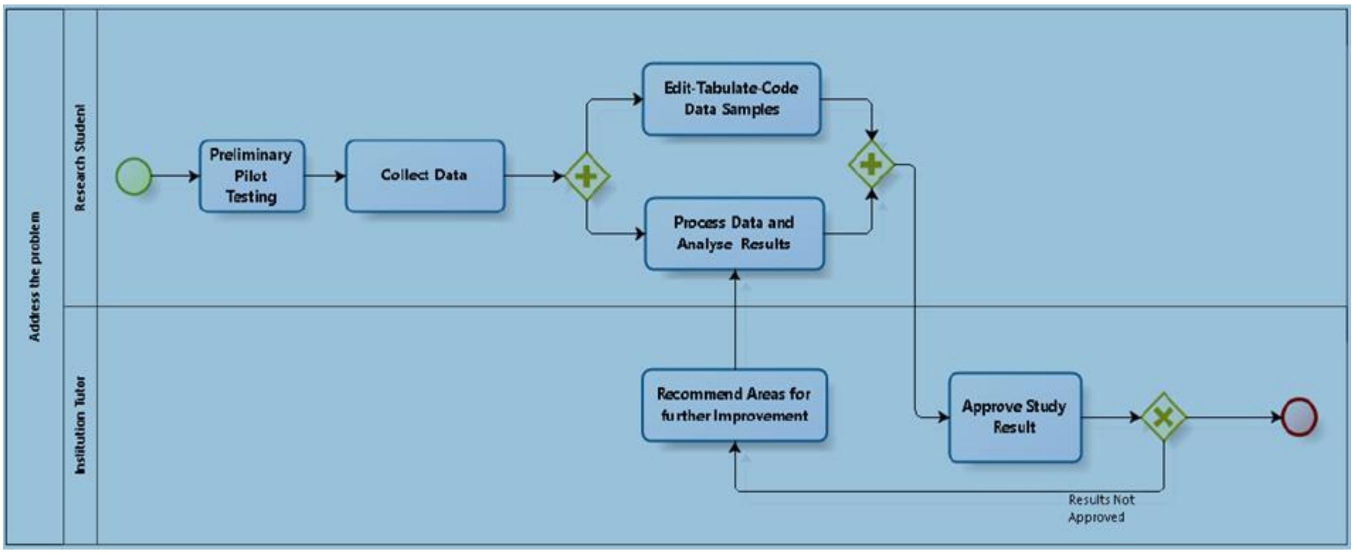


**Figure 4**. Address the Problem (*Assessment Stage*) Milestone or Sub-process with BPMN Notation.

Technically, the work provides the sets of milestones as shown in Figure 3 (and an example of Milestone *Address the Problem* in Figure 4) in order to determine and explain the steps taken within the learning process model. For instance, to ensure that a process instance enters (start) the model at a particular point in time, and not on the whole transition during the lifecycle of the model. Thus, from Defining the Topic Area to Review Literature, and Addressing the Problem, then Defending the Solution.

## III. The Research Learning Process Mining and Modelling Approach

The research learning process mining and modelling process is primarily focused on establishing information about resources hidden within the learning knowledge-base and how they are related. The work shows that the *conceptualisation* method is not only relevant during the requirement and design stage, but also for effective monitoring and enhancement of the whole process. The work employs the PM technique as shown in Figure 5 to put the captured volumes of data within a learning process (specifically the research process) into a process context or visualization (process mapping).

The PM method described in Figure 5 focus on transforming the existing raw data into meaningful and useful information, which are used to provide a system with a much more effective conceptual (semantic) reasoning capabilities.

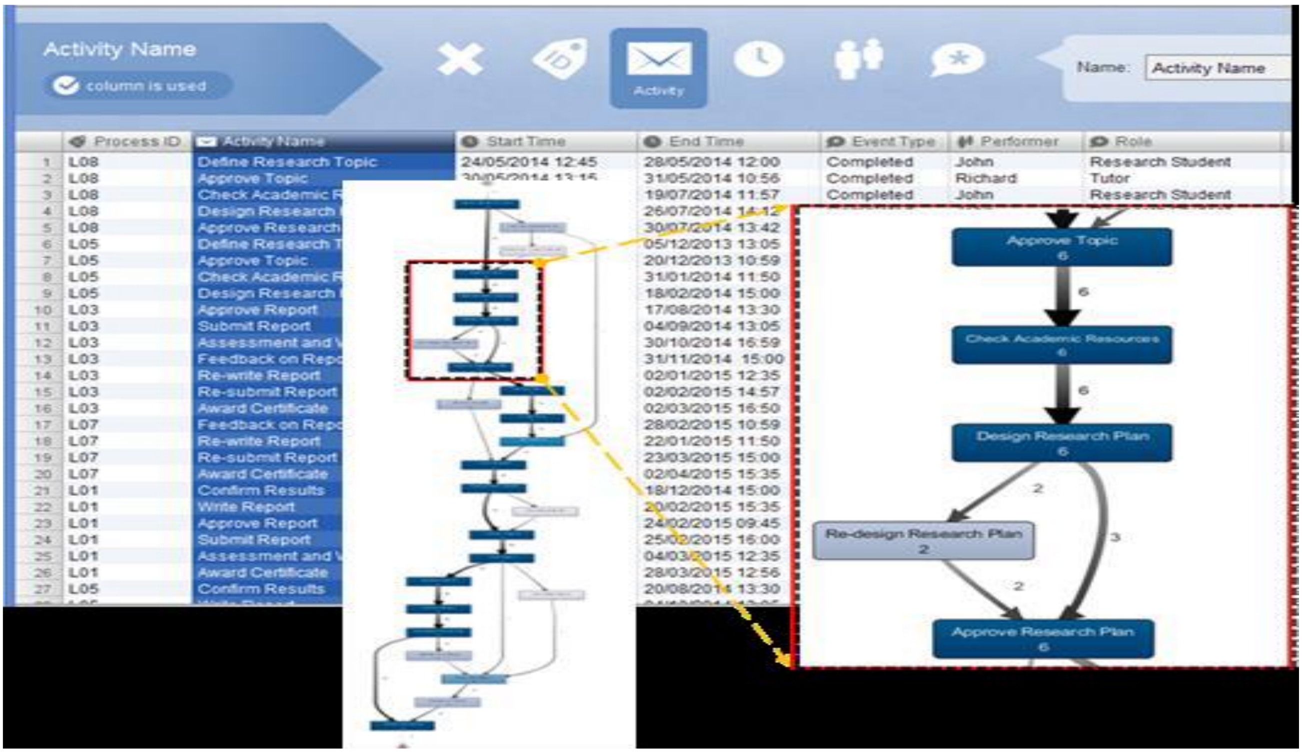


**Figure 5**. Input Events Data log with Mapped Processes in real-time.

As shown in the figure (Figure 5) and noted in the work of [3] - the input data for any process mining task is most often given as a table and the resulting data sets may be patterns, equations, graphs, tree structures, clusters or rules, e.g. as described in Figure 5. Indeed, Figure 5 represents the input data (about the learning process) and the resultant process model (process map) used for the purpose of this work. In fact, Figure 5 shows how the work extracts the input data that were necessary to be mapped (visualized) as a process model. In turn, the process mappings provide us with reliable and trustworthy results for the available datasets (learning process logs) based on the proven framework of the Fuzzy Miner [8] for further improvement and analysis.

*A. Design Framework and Implementation of the Semantic-based Process Mining Method*

Semantic-based process mining [1][5] is useful in addressing the problem of analysing item-sets (e.g. the column labels or variables in an event log) based on concepts rather than events tag/labels about the process in question. The *conceptual* means of analysis is beneficial in the formulation of robust and sharable descriptions of processes for an enhanced reasoning capability, as well as, increase in knowledge awareness and data management cycle as illustrated in Figure 6.


**Figure 6.** Implementation cycle of the Semantic-based Learning PM and Analysis Framework.

In Figure 6 the work shows that to perform Semantic-based PM: the data (e.g. learning process) and categories need to be captured. By category, we refer to every entity within the knowledge-base (e.g. the activity logs that makes up the learning process). The work shows in Figure 6 that first of all, the identification and modelling of the learning objects and/or data about the different users is necessary. Further, those data is selected from the knowledge-base, in order to carry out analysis of the captured data (and subsequently the discovered models) at a more abstraction level. In short, we describe in the figure (Figure 6) the learning process implementation cycle and how the learning data is being extracted, prepared and transformed into a minable format that allows for the Semantic Learning Process Mining (SLPM) to follow. For this approach, we focus on enhancing the informative value of the resulting process models as well as their usefulness by enriching the attributes of the various process elements or instances.

Figure 7 is the prototype (design) framework for the proposed semantic learning process mining and analysis approach. Clearly, the framework shows that the development of the semantic PM tools/methods entails three building blocks - Annotated Event Logs, Ontologies, and Semantic Reasoning [5][1] that are collectively targeted towards discovering, conformance and extension of processes as performed in real-time settings.

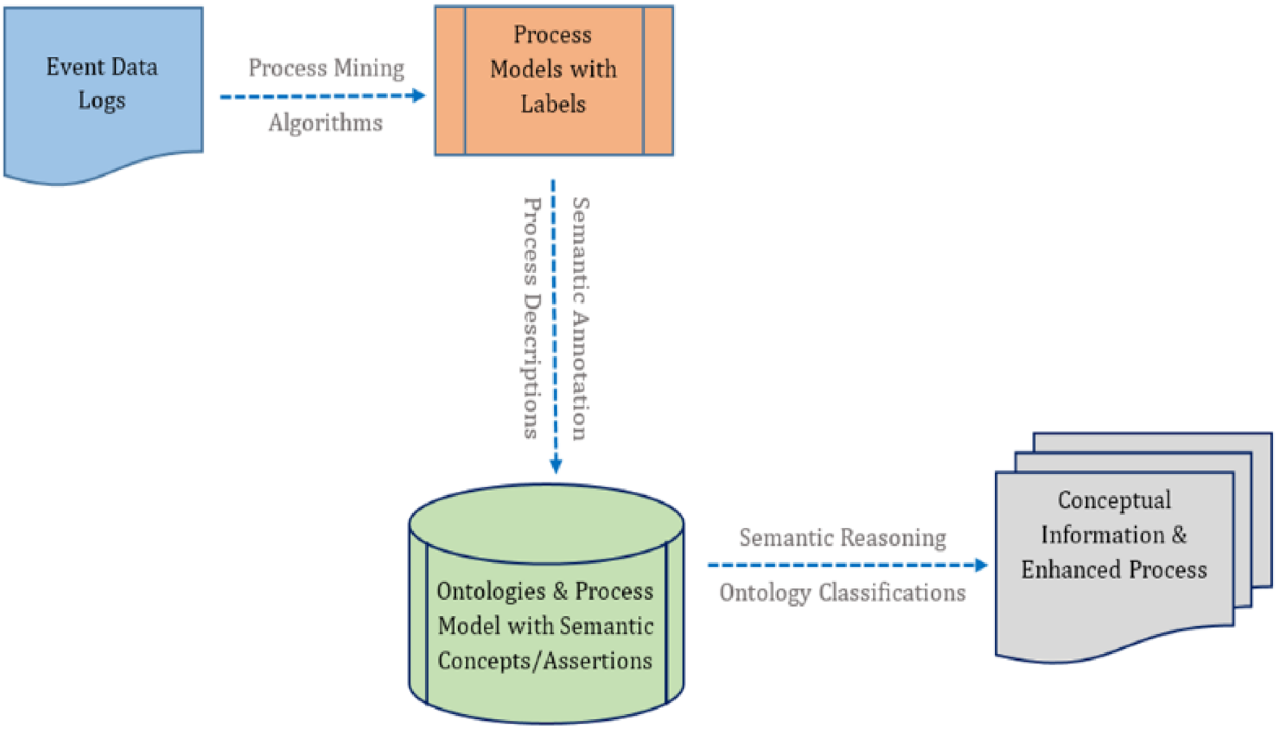


**Figure 7.** Conceptual Framework for implementing the Semantic-based Learning PM approach.

In short, the main idea of the framework (Figure 7) is to allow for semantical annotation of the elements in the events logs/model with concepts that they represent in real-time and then links them to concepts within an ontology specifically defined for representing the learning process.

To end with, reasoning over the ontologies with reference to elements in the log provides us with a robust way to answer questions about relationships the process elements share amongst themselves and to perform a more conceptual analysis capable of providing real-world answers that are closer to human understanding.

Essentially, the implementation of the method allows the meaning of the learning objects to be enhanced through the automated reasoning or classification of discoverable entities, especially by using the main function offered by the *Reasoner* to help in checking for consistency and/or relationships that exist in the model.

## IV. Experimental Settings and Results Discussion

The design framework (i.e. semantic-based PM and analysis framework) of this paper is focused on analysing the extracted streams of event logs about any given process domain based on *concepts* rather than the event tags of the process.

To illustrate the usefulness and real-time application of the semantic-based framework, the work makes use of the Semantic LTL Checker algorithm [1] in ProM [9], to provide a more conceptual description and enhancement of the informative value of the learning process data and the resultant process models. As shown in Figure 8 the algorithm applies concepts (activities attributes) in an ontology as input to parameters of a Linear Temporal Logic (LTL) formulae to formulate and answer questions about relationships the process elements share within the knowledge-base. Moreover, the algorithm uses the WSML2Reasoner to infer all the necessary associations.

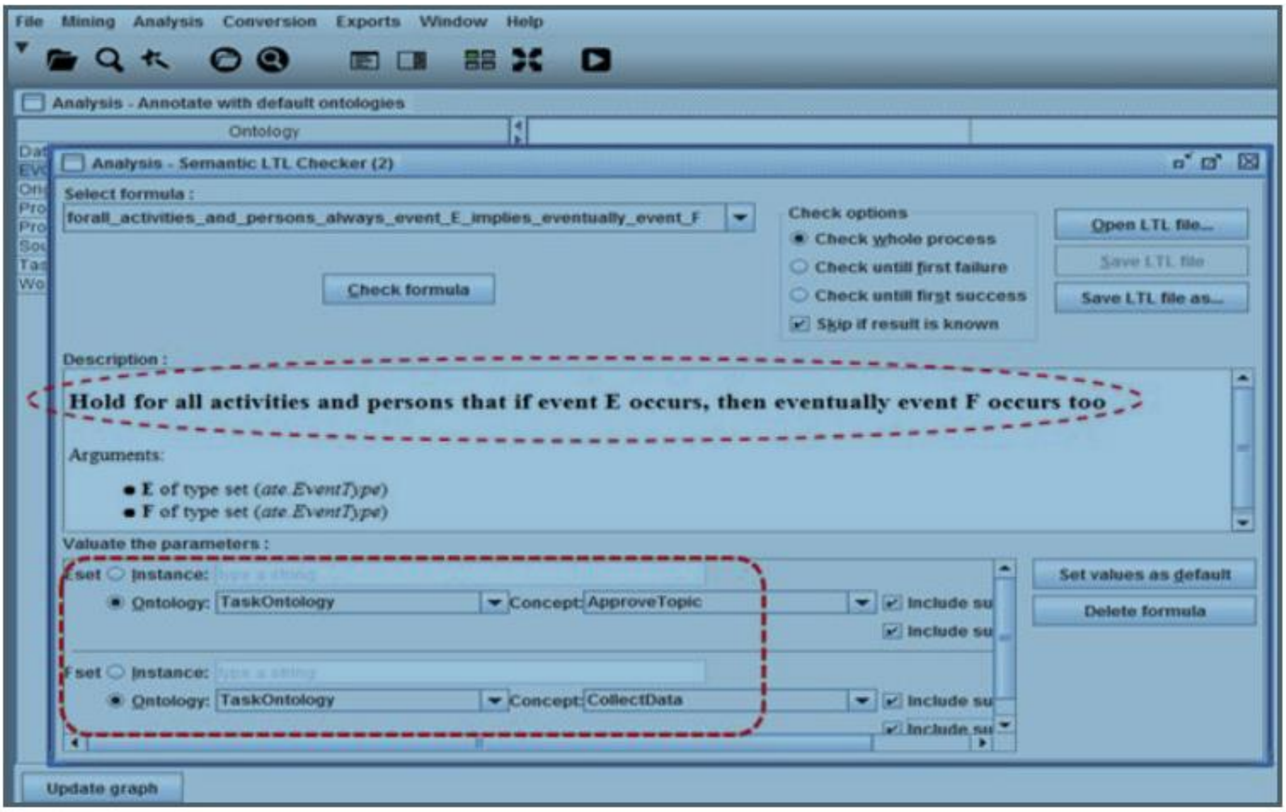


**Figure 8.** Experimental example of the Semantic-based Learning PM and Analysis Method.

First, the work uses the Semantic LTL Checker to carry out the annotation and construction of ontologies from the events log about the learning process based on attributes in the logs. The following are examples of the defined ontologies within the model: TaskOntology, OriginatorOntology, ProcessInstanceOntology, ProcessOntology, EventOntology, SourceOntology, DataFieldOntology, WorkFlowLogOntology etc. Consequently, based on the underlying ontologies (concepts/class taxonomies) and the reasoner, we perform a more conceptual analysis of the events logs as shown in Figure 8. Thus, by checking the formulae in Figure 8:

"*forall_activities_and_persons_always_event_E_implies_eventually_event_F*"

where:

Parameters: **E** set points to **Ontology = TaskOntology**, & **Concept = ApproveTopic**.

Parameters: **F** set points to the **Ontology = TaskOntology**, & **Concept = CollectData**

The result of executing the formula is an association that holds for all Activities and Persons that **IF** event **E** (ApproveTopic) occurs **THEN** eventually event **F** (CollectData) occurs too. Truly, such method is essentially important (e.g. in measuring the progress of a learner within the learning execution environment) and the purpose is not only to match the questions one would like to answer as seen in Figure 8 but also, the ability to identify and monitor deviations/bottlenecks or distinguish and/or establish the attributes the process elements (instances) share within the learning knowledge-base.

## V. Related Works

The application of semantic technologies (e.g semantic annotation, ontologies, and semantic reasoning) can help solve the problem of regulating the evolving and static methods for representing knowledge both at technological and theoretical levels [10]. The technology has proved to increase (enhance) the capability of understanding the concepts within the models by making inferences [5], retaining and applying what has been learned [11] as well as the discovery of new and/or enhancement of existing processes [3]. Even more, a greater number of developed systems that are used for process analysis makes use of various mining (data extraction and information retrieval) techniques for representation of the concepts, knowledge or data which are all directed towards the application of PM technologies to various aspects of processes [12]. Accordingly, the proposed method in this work uses the PM techniques and the semantic modelling method to represent the learning process. The aim is to further enhance this area of study by not only adopting the PM tools but also to present a method that leverages the semantic reasoning capabilities for computing the various individual elements that can be found within a process knowledge-base. This is done by automatically constructing worthwhile models capable of defining (through classification) and enhancing the observed learning patterns or behaviours. A number of works have been directed towards the use of the semantic-based PM for models discovery and analysis [1][13]. These works have shown that the ontological modelling and semantic reasoning methods are the essential building blocks for any semantic-based process mining and analysis approach.

On the other hand, amongst the existing methods used for traditional PM is the *Alpha algorithm* (α-algorithm) introduced in [3]. The authors in [3] used the algorithm to extract Petri net models from the event logs and have also been proven to support to an extent - both semantic and non-semantic process analysis of data. Moreover, the authors in [14] presented the Decision Miner used for decision point analysis in discovered models to detect data dependencies that can impact the mapping of events. Even though their approach does not support semantic process analysis, we show in this paper how the decision-making process can be improved by performing conceptual inferences over a learning knowledge-base to discover and establish valuable new knowledge or information by means of the semantic reasoning aptitude.

The authors in [15] argue that MXML is not all that is needed or prerequisite for any PM task. According to the authors, an MXML log is only able to refer to an identity tag for a particular entity within the log. The actual semantics which describes the object properties and the relationship the entities share within the process model is not readily available. In essence, MXML suffers from a lack of machine-processable semantics, even though it may be possible to create means of retrieving abstract knowledge or information from the discovered models as described in this paper. Practically, the work in this paper uses its semantic-based annotation and ontological modelling technique to link elements in an event log, with concepts that they represent in an ontology specifically designed for representing a learning process.

More so, the eXtensible Events Stream (XES) has also been introduced to address the problem of semantically adding attributes and definition of different concepts. Although the authors in [13] mention that most of the supporting algorithms are still under development, XES has of late been accepted as the standard file format for process mining [12][9].

Explicitly, in context of the work in this paper, the problem of modelling any given process (e.g learning process) can be solved by transforming the ontology population problem to a classification problem [16] where for each entity within the ontology, the concepts (classes) to which the entities belong to have to be determined (classified). According to the works of [16][17][5] such method of data analysis assumes that there already exists a probabilistic or fuzzy knowledge-base upon which this method is able to predict the patterns (classification) of new but not previously observed object/data types within the process domain in question. Indeed, such method (i.e. conceptual means of data analysis) have proved to be effective towards PM [5][17]. Moreover, the authors in [16] observes that those methods adopt the availability of an initial drawing of ontology [18] which can be automatically enhanced by adding or refining concepts [18][19][20] and have been proved to effectively solve process modelling problems [5] using description logics queries [21] particularly those based on classification, clustering and ranking of individuals.

## VI. Conclusion

The method of this paper proves that the PM techniques can be used to address the problem of determining the presence of different patterns within any given process (e.g the learning process). The method focuses on the extraction and modelling of events data about the learning process to enrich the informative values of the resulting models through the proposed *semantic-based process mining and analysis* technique. As a result, suitable process related information was determined which were then used to address the problem of analysing events logs and models based on concepts rather than the events tags or labels about the process. In short, the work demonstrates that the development of semantic-based PM approaches entails three main building blocks (Annotated Event Logs/Models, Ontologies and Semantic Reasoning) that are all aimed at discovering, conformance and extension of any given process domain (such as the learning process).

Future work will focus on analysing sets of activities within a learning knowledge-base to produce the behaviour of a particular group of learners which can then be used to load a more enhanced model for learning to improve the user's performance.

## Acknowledgment

The author is most grateful to the tutors and colleagues at the School of Architecture Computing and Engineering, College of Arts Technologies and Innovation, University of East London, UK for their wonderful support throughout the time of this research work. The author would also like to thank the reviewers and editors for their valuable input and comments.

## References

[1] A.K.A. de Medeiros, W.M.P. Van der Aalst, C. Pedrinaci. “Semantic Process Mining Tools: Core Building Blocks”. *In ECIS*, 2008, Galway, Ireland, pp. 1953-1964, 2008.

[2] W.M.P. Van der Aalst, A.J.M.M Weijters, L. Maruster. “Workflow mining: discovering process models from event logs”. *IEEE Transactions on Knowledge and Data Engineering*. vol. 16(9), pp. 1128-1142, 2004.

[3] W.M.P. Van der Aalst. “*Process Mining: Data Science in Action*”. 2nd Ed. Springer, 2016.

[4] R. Kumar. “*Research Methodology: A Step-by-Step Guide for Beginners*” In SAGE, London, pp. 33-42, 2014.

[5] K. Okoye, U. Naeem, S. Islam. “Semantic Fuzzy Mining: Enhancement of process models and event logs analysis from Syntactic to Conceptual Level”. *International Journal of Hybrid Intelligent System*, (IJHIS) IOS Press, vol. 14 (1-2), pp. 67–98, 2017.

[6] R. Koper, C. Tattersal. “Learning Design: “*A Handbook on Modelling and Delivering Networked Education and Training*” Springer, Berlin, pp. 25-40, 2005.

[7] P. Newby. “*Research Methods for Education*”. In Pearson Education Ltd, Essex, England, pp. 221-259, 2010.

[8] C.W. Gunther, W.M.P. Van der Aalst. “Fuzzy Mining: Adaptive process Simplification Based on Multi-Perspective metrics”. *International conference on Business Process Management* (BPM 2007), Springer, Berlin, vol. 4714. pp. 328-343, 2007.

[9] H.M.W. Verbeek, J.C.A.M. Buijs, B.F Dongen, W.M.P. Van der Aalst. "XES, XESame, and ProM 6". In Soffer, P. and Proper, E. (Eds), *Information Systems Evolution* – CAiSE Forum 2010, Hammamet, Tunisia, Selected Extended Papers, Springer, Berlin, pp. 60-75, 2011.

[10] D. Barbieri, D. Braga, S. Ceri, E.D. Valle, Yi Huang, V. Tresp, A. Rettinger, H. Wermser. "Deductive and Inductive Stream Reasoning for Semantic Social Media Analytics". *Journal of Intelligent Systems*, IEEE, vol. 25(6), pp. 32-41, 2010.

[11] E.H.J. Nooijen, B.F. van Dongen, D. Fahland. "Automatic Discovery of Data-Centric and Artifact-Centric Processes". *BPM Workshops.* Lecture Notes in Business Information Processing, vol. 132, pp. 316-327, 2013.

[12] IEEE Task Force on Process Mining. Process Mining Manifesto. *BPM Workshops* LNBIP, Springer, vol. 99, pp. 169–194, 2012.

[13] W. Jareevongpiboon, P. Janecek. "Ontological approach to enhance results of business process mining and analysis". *Business Process Management Journal*, vol. 19(3), pp. 459 – 476, 2013.

[14] A. Rozinat, W.M.P. Van der Aalst. "Decision mining in ProM". *In 4th Business Process Management Conference*, Vienna, Austria, pp. 420-425, 2006.

[15] C. Pedrinaci, J. Domingue. "Towards an ontology for process monitoring and mining". *Workshop on Semantic Business Process and Product Lifecycle Management*, Innsbruck, Austria, pp. 76–87, 2007.

[16] C. d'Amato, N. Fanizzi, F. Esposito. "Query answering and ontology population: An inductive approach". In S. Bechhofer et al., editors, *5th European Semantic Web Conferenc*e, vol. 5021, pp 288–302, 2008.

[17] K. Okoye, S. Islam, U. Naeem. "Ontology: Core Process Mining and Querying Enabling Tool". In: C. Thomas, Editor. *Ontology in Information Science*, InTechOpen Publishers, Chap. 7, pp. 145-168, 2018.

[18] K. Okoye, A.R.H. Tawil, U. Naeem, E. Lamine. "A Semantic Reasoning Method Towards Ontological Model for Automated Learning Analysis". In: Pillay, N., Engelbrecht, A., Abraham, A., du Plessis, M., Snášel, V., Muda, A., editors. *Advances in Intelligent Systems and Computing.* NaBIC Conference 2015. Springer, Switzerland, pp. 49-60, 2016.

[19] A. Bogarín, R. Cerezo, C. Romero. "A Survey on Educational Process Mining". *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* (WIRES), Wiley Periodicals Inc, 1 Jan/Feb, vol. 8(1), pp. e1230, 2018.

[20] K. Okoye, A.R.H. Tawil, U. Naeem, S. Islam, E. Lamine. "Semantic-based Model Analysis towards Enhancing Information Values of Process Mining: Case Study of Learning Process Domain". In: A. Abraham, A. K. Cherukuri, A. M. Madureira & A. K. Muda, editors. *Advances in Intelligent Systems and Computing book series* (AISC, volume 614). Proceedings of SoCPaR Conference 2016. Springer International Publishing. pp. 622-633, 2018.

[21] F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, P.F. Patel-Schneider. "*Description Logic Handbook: theory, implementation, and applications*". 1st [ed]. New York, USA, Cambridge University Press, 2003.

## Author Biography

**Kingsley Okoye** received his Bachelors degree in Computer Science in 2007, a Masters degree in Technology Management in 2011, and a Doctorate Degree in Software Engineering from the School of Architecture Computing and Engineering, College of Arts Technologies and Innovation, University of East London. He is an MIET member at the Institution of Engineering and Technology, UK. Graduate Member at the Institute of Electrical and Electronics Engineers, IEEE. He is a devoted Researcher, Industry, and Academia in operational, hardware and software fields of computing in general such as the Data Science, Machine Learning, Artificial Intelligence, Big Data and Advanced Analytics, Software Development and Programming, and Business Process Management. His Research interests includes Process Mining, Business Process Modelling and Automation, Learning Analytics and Systems, Semantic Web Technologies, Knowledge Management, Big Data Analysis and Process Querying, Internet Applications and Ontology.