

Using Semantic-based Approach to Manage Perspectives of Process Mining: Application on Improving Learning Process Domain Data

Okoye Kingsley, Abdel-Rahman H. Tawil, Usman
Naeem, Syed Islam
School of Architecture Computing and Engineering
University of East London
London, United Kingdom
{u0926644, A.R.Tawil, U.Naeem, Syed.Islam}@uel.ac.uk

Elyes Lamine
Université de Toulouse, Mines-Albi, CGI
Campus Jarlard
Albi Cedex 09, France
Elyes.lamine@mines-albi.fr

Abstract— Mining useful knowledge from data readily available in today's information systems has been a common challenge in recent years as more and more events are being recorded, and there is need to improve and support many organisational processes in a competitive and rapidly changing environments. The work in this paper shows using a case study of Learning Process - how data from various process domains can be extracted, semantically prepared, and transformed into mining executable formats to support the discovery, monitoring and enhancement of real-time processes. In so doing, it enables the prediction of individual patterns/behaviour through further semantic analysis of the discovered models. Our aim is to extract streams of event logs from a learning execution environment and describe formats that allows for mining and improved process analysis of the captured data. The approach involves augmenting the informative value of the resulting model derived from mining event data about the process by semantically annotating the process elements with concepts they represent in real time using process descriptions languages, and linking them to an ontology specifically designed for representing learning processes to allow for the analysis of the extracted event logs based on concepts rather than the event tags of the process. The semantic analysis allows the meaning of the learning object properties and model to be enhanced through the use of property characteristics and classification of discoverable entities, to generate inference knowledge which are then used to determine useful learning patterns by means of the proposed Semantic Learning Process Mining (SLPM) formalization - described technically as Semantic-Fuzzy Miner. As a result, the approach provides us with the capability to infer new and discover hidden relationships/attributes the process instances share amongst themselves within the knowledge base, and the ability to identify and address the problem of determining the presence of different learning patterns or behaviour. Inference knowledge discovered due to semantic enrichment of the process model is advantageous especially in solving some didactic issues and answering some questions with regards to different Learners behaviour within the context of process mining and semantic model analysis. To this end, we show that information derived from process mining algorithms can be improved by adding semantic knowledge to the resulting model.

Keywords - process model, process mining, semantic annotation, ontology, learning process, event logs

I. INTRODUCTION

In recent years, handling large datasets extracted from process logs readily available in today's information systems have raised intense discussion within the big data community. Data collection and analysis is proving to be more and more complex in many organisations. These growing complexities are evidenced by the need for richer and more precise description of real-time processes that allows for flexible exploration of different operational data. A common challenge has been on how to create systems capable of providing platforms for data exploration by stemming understandable patterns as well as making the discovered patterns explicable [1][2]. Over the past few decades, advances in data mining techniques has led to significant growths in data analytics and big data research. Researches has shown that a valuable technique that uses data mining techniques to discover meaningful information from event data about any process for further analysis and exploration is the Process Mining technique [3]. Process mining combines techniques from computational intelligence and data mining to process modeling and analysis, as well as several other disciplines to analyze large data sets [2]. The process mining technique 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 any process domain has been made possible, and has proven to be essential for extracting models capable of creating new and meaningful knowledge. A shared challenge with most of the existing process 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 are somewhat ambiguous when confronted with unstructured data. 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 paper, we show that analysis provided by current process mining techniques can be improved by adding semantic information to event logs of the process domain. We use a case study of Research Process to illustrate our

approach. Our focus is 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 model and analysis. We look at how effective semantic reasoning can be used to lift process mining analysis from the syntactic level to a conceptual level through semantic representation of the deployed models. The semantic analysis takes advantage of the rich semantics described in event data log of a learning process, and links them to concepts in an 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 the event logs of the learning process. The method uses the semantics of the sets of activities within a learning process 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. We introduce the approach as means towards discovering and enrichment of the set of recurrent behaviours that can be found within the learning execution environment following the work in [4] to determine attributes the process elements share amongst themselves within the learning knowledge base, or that distinguishes a category of entities (learners) from another. The technique is developed in order to address the problem of determining the presence of different patterns within the learning process model. The standpoint for our described approach 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 learning process domain, to help address the problem of analysing data based on concepts and to answer questions about relationships the learning objects share amongst themselves within the learning knowledge-base”. The semantic viewpoint is captured by annotating 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? [5].

The rest of the paper is structured as follows:

Section 2, explains the search for our primary goal and motivation towards using semantic-based approach to manage perspectives of process mining techniques.

In section 3, we show how we extract the input data necessary to be mapped onto the learning process logs, and also provide an example of a research process execution data which we use to demonstrate our approach throughout this paper.

Section 4 describes the learning problem use case scenario and implementation of the semantic-based approach to show the usefulness of the proposed mining algorithm and formalisation.

In section 5, we discuss and analyse appropriate related works in this area of research.

We finally conclude the paper and point out directions for future works in section 6.

II. APPLICATION OF SEMANTIC-BASED APPROACH FOR PROCESS MINING

One of the benefit provided by semantic process mining and analysis is the ability to describe the semantics behind the tags in an event log of a process considered useful for discovery of new knowledge. The main opportunity is that this analysis is enhanced as a result of it being based on concepts rather than the event tags. The work in this paper is directed towards finding useful aspects of process mining techniques as described in Fig. 1, which are used to discover process models from event logs about a learning process while keeping a clear focus on the use of semantic driven approach for enriching the information value of the derived models through semantic-based analysis.

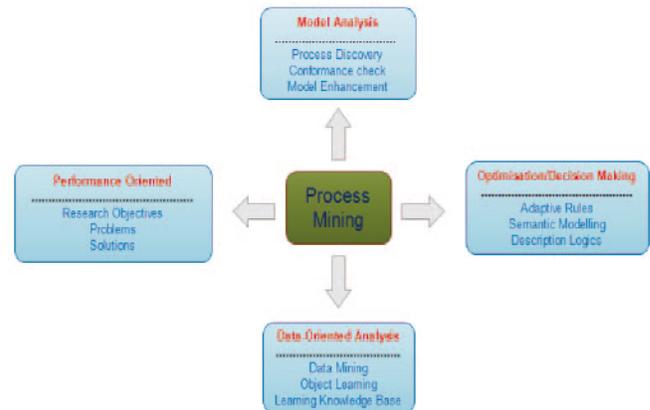


Fig. 1 Conceptual overview of Semantic-based Process Mining and Analysis

Semantic modelling and analysis is an important technique especially to guide our goal which is to “extract streams of event logs from a learning process and then describe formats that allows for mining and improved process analysis of the data”. The search for our primary goal is driven by our aim which is (i) applying process mining techniques to the domain of a learning process and (ii) to provide real time semantic knowledge and understanding about learning process which are useful towards the development of learning process mining algorithms that are more intelligent with high level of effective reasoning capabilities. We address the learning questions LQ1 and LQ2 in section 3, to show in details how the semantic approach is implemented and relevant in the context of process mining to improve formalization and information values of the resulting learning process model. At the core of the approach is a Semantic Learning Process Mining algorithm (Semantic-Fuzzy Miner) and Declarative Process Discovery tactics which we developed to help find answers to our motivational perspectives; Thus

- To show how process mining can be applied to improve the informative value of learning process data.
- Describe how improved process models can be derived from the large volume of event data logs found within the learning process knowledge base.

- Use of semantic representation of the deployed models to enrich the result of the learning process mining through ontological descriptions and inference knowledge discovery.
- Use of ontologies with effective semantic reasoning to lift process mining analysis from the syntactic level to a conceptual level.
- Reveal how references to ontologies and effective raising of process analysis from syntactic to semantic level enables real time viewpoints on the learning process domain; which helps to address the problem of analysing learning data based on concepts and to answer questions about relationships the learning objects (process instances) share amongst themselves within the learning knowledge-base.

Contributions: We delivered as a result of tackling the motivation of this paper, means by which the use of semantic tools can be used to manage perspectives of process mining techniques by enhancing the informative value of the resulting models. This paper makes the following key contributions:

- Semantic motivated synchronization of event log formats for learning process data.
- Ontology driven search for explorative analysis of learning activities and its process executions.
- Techniques for annotating unlabelled learning activity sequences using business process model notations and ontology schema/vocabularies.
- Use of semantics tools to manage perspectives of process mining algorithms and definition of methods towards discovery and enhancement of process model analysis.
- Useful strategies towards development of process mining algorithms that are more intelligent, predictive and robotically adaptive.
- Importance of semantics process mining to augment informative value of data about any domain process: case study of learning process.

III. DOMAIN PROCESS MINING (RESEARCH PROCESS EVENT DATA)

The purpose of designing the semantic-based approach is on revealing information about resources hidden within the learning event data log, the different process instances that makes up the learning model, and identifying useful arrays towards enriching the information values of the resulting process models based on the captured user profiles. Process mining techniques aims to address such problem by establishing a direct connection between the discovered process models and actual low-level event data about the processes, and therefore allows for viewing the same reality from different angles or at different levels of abstraction. As a ground for the research, the work in this paper focuses on using semantic tools to manage insights to application of process mining techniques to provide improved analysis of

the resulting models. Thus, our purpose for designing the Semantic Learning Process Mining algorithm - to extract, semantically prepare, and transform event data about a learning process into mining executable formats that allows us to perform an improved learning process analysis by constructing a semantic model that represents the deployed process model. As shown in Fig. 2, the first step towards achieving this goal is to capture event data about a learning process (research process domain) and generate process model to show in details how the learning activities has been performed and to reveal interesting connections between the different elements (process instances). The viewpoints (process mapping) then allow us to further perform an enhanced analysis of the learning model.

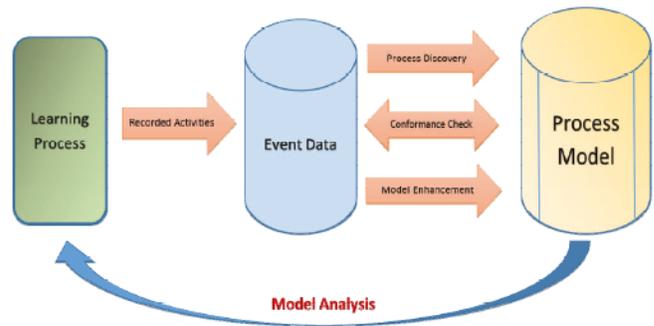


Fig. 2 Application of the Process Mining Techniques

The mining technique described in Fig. 2, involves the extraction of process history data from the learning execution environment (Table 1), which is then followed by submitting the resulting event streams format to the process mining and analytic environment to help in discovery of the process model (Fig. 3) using the Fuzzy Miner algorithm in Disco tool [6].

TABLE 1. EVENT INPUT DATA LOG FOR RESEARCH PROCESS

| Process ID | Activity Type | Start Time | End Time | Event Type | Performer | Role |
|------------|--------------------------|------------------|------------------|------------|-------------|-----------------------|
| L08 | Define Research Topic | 24/05/2014 12:45 | 28/05/2014 12:00 | Completed | John | Research Student |
| L08 | Approve Topic | 30/05/2014 13:15 | 31/05/2014 10:56 | Completed | Richard | Tutor |
| L08 | Check Academic Resources | 01/06/2014 10:45 | 19/07/2014 11:57 | Completed | John | Research Student |
| L08 | Design Research Plan | 20/07/2014 10:57 | 26/07/2014 14:12 | Completed | John | Research Student |
| L08 | Approve Research Plan | 28/07/2014 11:15 | 30/07/2014 13:42 | Completed | Richard | Tutor |
| L05 | Define Research Topic | 31/10/2013 10:57 | 05/12/2013 13:05 | Completed | Paul | Research Student |
| L05 | Approve Topic | 18/12/2013 11:15 | 20/12/2013 10:59 | Completed | David | Tutor |
| L05 | Check Academic Resources | 14/01/2014 11:40 | 31/01/2014 11:50 | Completed | Paul | Research Student |
| L05 | Design Research Plan | 07/02/2014 12:35 | 18/02/2014 15:00 | Completed | Paul | Research Student |
| L08 | Approve Report | 10/08/2014 10:45 | 17/08/2014 13:30 | Completed | Clare | Tutor |
| L03 | Submit Report | 04/09/2014 10:57 | 04/09/2014 13:03 | Completed | Mike | Research Student |
| L03 | Assessment and Viva | 30/10/2014 11:15 | 30/10/2014 16:59 | Completed | Department | Institution Associate |
| L03 | Feedback on Report | 30/11/2014 12:35 | 31/11/2014 15:00 | Completed | Department | Institution Associate |
| L03 | Re-write Report | 04/12/2014 13:30 | 02/01/2015 12:35 | Completed | Mike | Research Student |
| L03 | Re-submit Report | 02/02/2015 12:25 | 02/02/2015 14:57 | Completed | Mike | Research Student |
| L03 | Award Certificate | 02/03/2015 11:45 | 02/03/2015 16:50 | Completed | Institution | Research Institution |
| L07 | Feedback on Report | 28/02/2015 12:49 | 28/02/2015 10:59 | Completed | Department | Institution Associate |
| L07 | Re-write Report | 01/03/2015 09:45 | 22/01/2015 11:50 | Completed | Vera | Research Student |
| L07 | Re-submit Report | 23/03/2015 12:42 | 23/03/2015 15:00 | Completed | Vera | Research Student |
| L07 | Award Certificate | 02/04/2015 10:35 | 02/04/2015 15:35 | Completed | Institution | Research Institution |
| U01 | Confirm Results | 14/12/2014 12:35 | 18/12/2014 15:00 | Completed | Ben | Tutor |
| U01 | Write Report | 19/12/2014 12:49 | 20/02/2015 15:35 | Completed | Danny | Research Student |
| U01 | Approve Report | 21/02/2015 09:45 | 24/02/2015 09:45 | Completed | Mark | Tutor |
| U01 | Submit Report | 25/02/2015 12:42 | 25/02/2015 16:00 | Completed | Danny | Research Student |
| U01 | Assessment and Viva | 04/03/2015 13:30 | 04/03/2015 12:35 | Completed | Department | Institution Associate |
| U01 | Award Certificate | 28/03/2015 10:35 | 28/03/2015 12:56 | Completed | Institution | Research Institution |
| U05 | Confirm Results | 19/08/2014 10:45 | 20/08/2014 13:30 | Completed | Clare | Tutor |

Consequently, suitable learning patterns are determined which enables the automatic creation of the learning process mapping (workflow) as shown in the resulting Fuzzy model in Fig. 3.

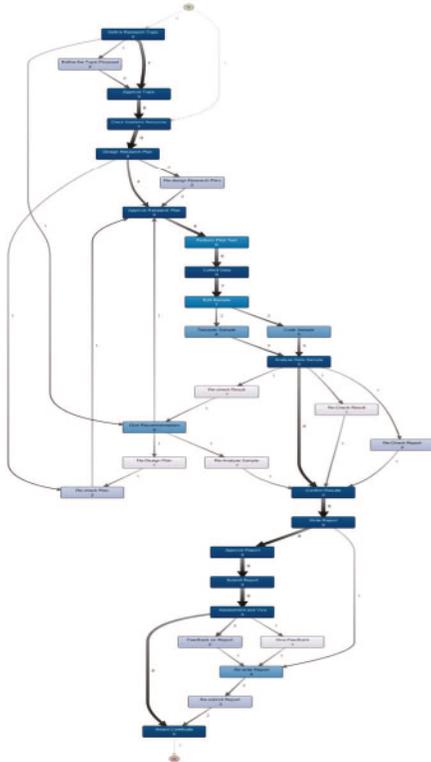


Fig. 3 Resulting Fuzzy Model from Mining the Process Event Data.

The resulting approach and process model shows how we extract the input data necessary to be mapped for the proposed semantic analysis. The benefit of performing such tactics is that the resulting process map allows us to quickly and interactively explore the learning process into multiple directions and to answer concrete questions about the learning activities workflow, and more importantly allows us to further model and hold inference reasoning to generate process improvement ideas along the way. The mapping step was necessary especially when our main purpose is to make the semantics information about the learning data readily available for mining and analysis at conceptual level [7]. The mining approach reveals the process map and provides us with the opportunity to focus on the stream of behaviours, and to see the paths they follow in the process model. The

Process ID tags (Table 1) were used to assign the identifier for process instances and *ActivityType tags* for the set of task that are performed during the learning process. We associate *Timestamp tags* with *Start Time* and *End Time* of event for the activity instances for the purpose of sequencing. The time performance shows how often each task is executed in term of frequency of each activity in the process model.

A. Modelling Process Transition for Learning Concepts and Activities

To perform semantic-based analysis of the resulting model derived by applying the process mining technique, it was necessary to construct process transition information about the mapped processes and semantics of the sequence of the learning activities. We develop a Business Process Model Notation (BPMN) for the learning workflow library to look at the learning activities and grouping of immediate preceding tasks in order to semantically map the process model. To achieve this objective, we construct a BPMN model (Fig. 4) with notational elements capable of describing the nesting of individual activities (*Task*) by using the event-based (AND – XOR – OR) split and join gateways [3]. According to [3] BPMN notations are token based semantics which can be used to replay a particular process instance trace within the discovered process model. We based the BPMN learning model workflow transition on four milestones; *Establish Context* → *Learning Stage* → *Assessment Stage* → *Validation of Learning Outcome* (Fig. 5) with the primary aim to determine and explain the steps taken within the mapped processes, and to provide a consistency check when designing the semantic model to ensure that a process instance enters the model at a particular point in time and not on the whole transition during the lifecycle of the learning process model - from *Defining the Topic Area* to *Review Literature* and *Addressing the Problem* then *Defending the Solution*. The preliminary knowledge of the discovered sequence of activities and learning patterns is what we use in providing new knowledge and a richer set of intelligence towards extending the learning process model to a conceptual level analysis.

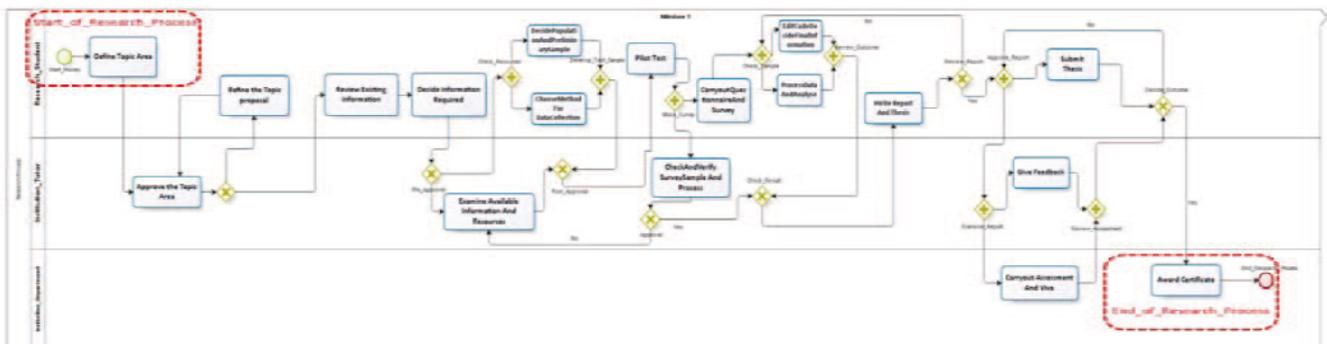


Fig. 4 BPMN Model with notation for Process Sequence-Flow with Bizagi Modeller

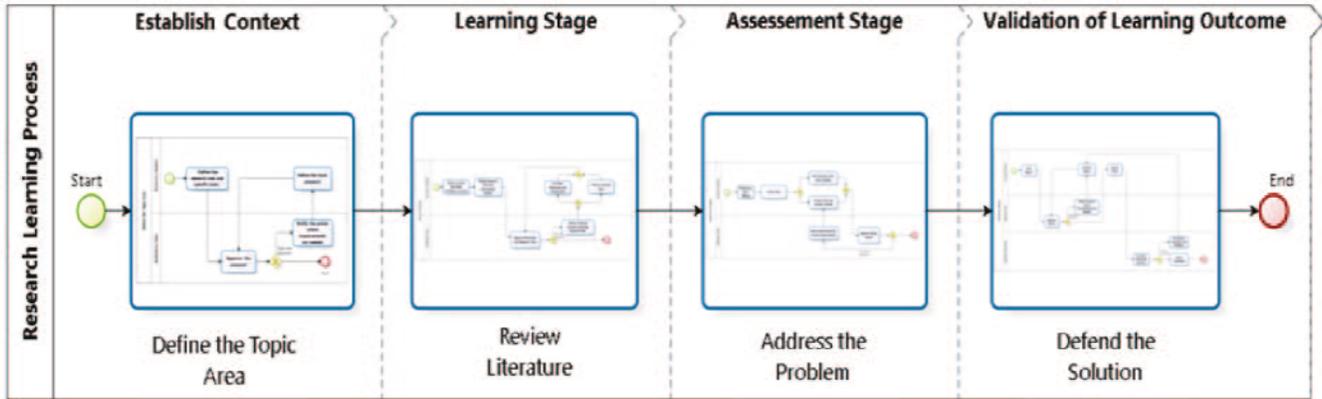


Fig. 5 Process Model with Defined Milestones

IV. USE CASE PROBLEM SCENARIO AND MANAGING PERSPECTIVES OF LEARNING PROCESS MINING OUTCOME

A. Definition of Learning Problem Example

Semantic-based process mining is an advantageous tactic especially in solving some didactic issues and answering some questions with regards to different Learners behaviour. To show the usefulness of this approach, we define resolution towards resolving a real life question and understanding about a learning process. We use the questions in LQ1 and LQ2 to provide learning knowledge representation and specification by revealing interesting connection among the learning object and data types, in order to provide a better understanding of how the different elements within the Learning Process Knowledge base relate and interact with each other. We show the usefulness of the tactics by answering the following question, thus to determine;

LQ1. What attribute or paths do successful learners have in common?

LQ2. What attributes distinguishes such successful learners from the uncompleted ones?

We describe a running example of learning process in execution using the Research Process as use case to prove how our approach can be used to answer the learning question LQ1 and LQ2. The resulting process map discovered using the fuzzy miner algorithm in Disco (Fig. 3) and the BPMN workflow (Fig. 4) - description of the Research Process activity sequence reveals that the first step to conducting a research is to decide on what to investigate (i.e Research Topic), and then go about finding answers to the research questions. At the end of the process, the researcher is expected to be awarded a certificate. These process involves the workflow of the journey from choosing the research topic to being awarded a certificate, and comprises sequence of practical steps or set of activities through which must be performed in order to find answers to the research problems. The workflow for these steps are not static, it changes as a researcher travel along the research process. At each phase or milestone of the process, the researcher is required to complete a variety of learning

activities which will help in achieving the research goal. Despite all of this useful information, the deployed model still does not disclose to us how the individual process instances that makes up the model interact or differ from each other (semantic abstraction levels), which attributes they share amongst themselves within the knowledge base, or the activities they perform together or differently, for instance, who are the individuals that have successfully completed the research process? For this reason, we believe that by adding semantic knowledge to the deployed model, it will be possible for us to determine and address the above mentioned challenges. To explicate this tactics, we define that for a research process to be classified as successful, it is necessary that the researcher must complete a given set of milestones in order to be awarded the degree. However in any case whereby the researcher has not completed a set of milestone(s) which is necessary to ensure the research outcome, such learner can be classified as uncomplete learner. In such way we can ascertain which individuals has successfully completed the research process.

B. Managing Perspectives of Learning Process Mining Results

Semantic descriptions (structured organisation) can serve as a good practice for representation of process mining results by providing formal way of representing the datasets from the data pre-processing stage to the mining results [8]. Many process mining algorithms lack the ability to identify and make use of semantics across different domains. Ontologies are useful tool towards specifying domain process semantics and can reduce the semantic gap by annotating the resulting models from the event data with rich semantics using process description languages such as Ontology Web Language (OWL)[9], Semantic Web Rule Language (SWRL)[10] etc. As a group of concepts and relationships, ontologies has the capability to make consistency inference checks by performing semantic logic reasoning for entities found within the process knowledge base through the object property characteristics and assertions. The work in this paper uses the semantic based inference knowledge discovery approach to find out patterns/behaviour that describes or distinguishes certain entities within the learning knowledge base by recognizing

what paths successful entities (learners) follow or have in common and what attributes distinguishes the successful learners from the uncompleted ones as described in questions LQ1 and LQ2. The purpose is not only to answer the following question using the semantic-based approach, but to show how by referring to the attributes (semantic knowledge) and the application of semantic reasoning, it becomes easy to manage or refer to a particular case/group of learners which in this example we focus on the use case of Successful and Uncomplete learners. We show in the BPMN workflow library (Fig. 4) and the developed semantic model that the flow of processes from the definition of topic area to award certificate; consist of different learning steps which the researcher has to or partly perform in order to complete the research process. We provided Four milestones; Establish Context → Learning Stage → Assessment Stage → Validation of Learning Outcome (Fig. 5) in order to determine and explain the steps taken during the research process; from Defining the Topic Area *-to-* Review Literature *-and-* Addressing the Problem *-then-* Defending the Solution. These milestones consist of sequence of activities, and the order in which these learning activities are carried out has the capability of determining the research outcome.

In Fig. 6, 7, 8 and 9 we show the Learning Activity concepts that are defined in the ontology model and how they are mapped to the various Milestones of the Research Process to ensure sequence of transitions during the learning process.

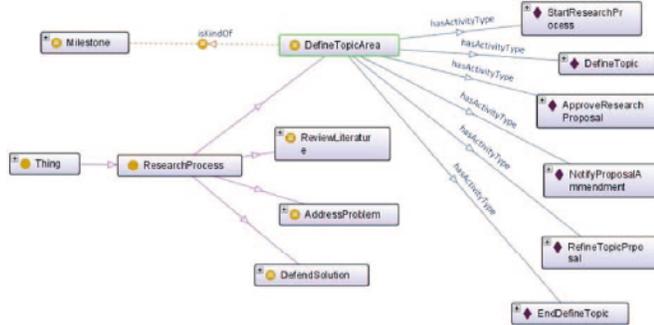


Fig. 6. Ontology Graph and ActivityConcept mapping for the DefineTopicArea Milestone.

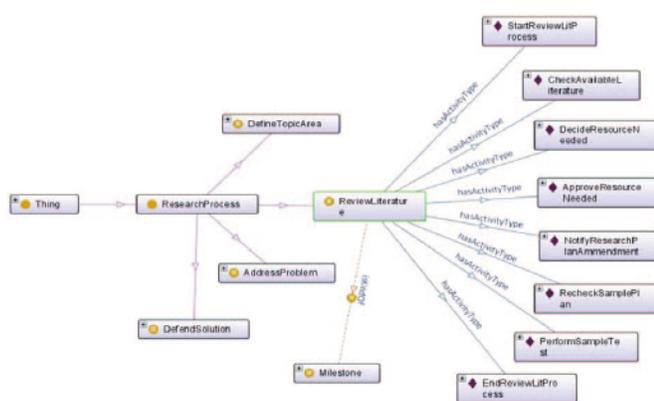


Fig. 7. Ontology Graph and ActivityConcept mapping for the ReviewLiterature Milestone.

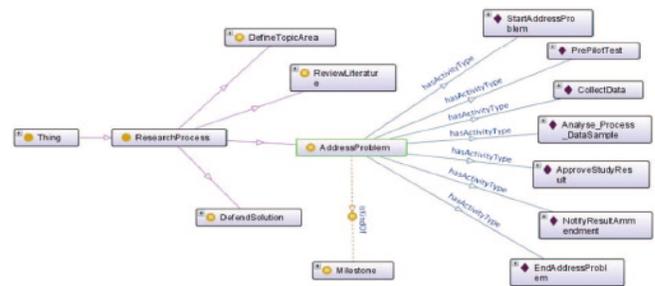


Fig. 8. Ontology Graph and ActivityConcept mapping for the AddressProblem Milestone.

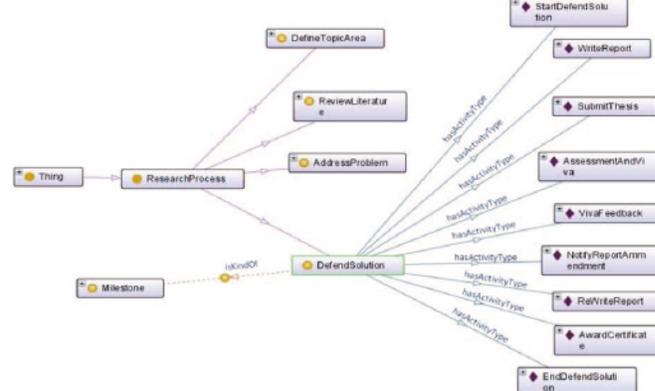


Fig. 9. Ontology Graph and ActivityConcept mapping for the DefendSolution Milestone.

The drive for the semantic mapping of the activity concepts is that the approach allows the meaning of the learning objects and properties to be enhanced through the use of *property characteristics* and *classification of discoverable entities*. For instance, to address the learning problem stated in LQ1, we refer to the deployed model, and to this effect, describe that a Successful Learner is a subclass of, amongst other NamedLearnerCategory, a Person that performs some LearningActivityConcepts, who has a universal object property restriction/relationship with the four milestones of the ResearchProcessClass. As shown in Fig. 10 - the *necessary condition* is: if something is a Successful Learner, it is *necessary* for it to be a participant of the Learning ActivityConcept class and *necessary* for it to have a kind of sufficiently defined condition and relationship with the four class: DefineTopicArea, ReviewLiterature, AddressProblem and DefendSolution.

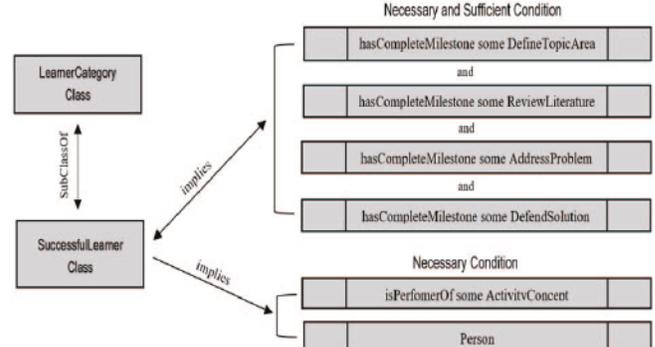


Fig. 10 Object Property Assertion/Attributes for SuccessfulLearner Class.

Accordingly, to address and provide answer to LQ2, it was necessary to establish the object property assertion for the Uncomplete Learners to be able to determine what attributes distinguishes such learners from the Successful ones. We describe in our model that an Uncomplete Learner is a subclass of, amongst other NamedLearnerCategory, a Person that performs some Learning ActivityConcept who has a universal object property restriction/relationship with only some of the milestones of the ResearchProcess Class but not all of the classes. As shown in Fig. 11 - the *necessary condition* is: if something is a Uncomplete Learner, it is *necessary* for it to be a participant of the Learning ActivityConcept class and *necessary* for it to have a kind of sufficiently defined condition and relationship with only some of the Class:- DefineTopicArea, ReviewLiterature, AddressProblem but not all four.

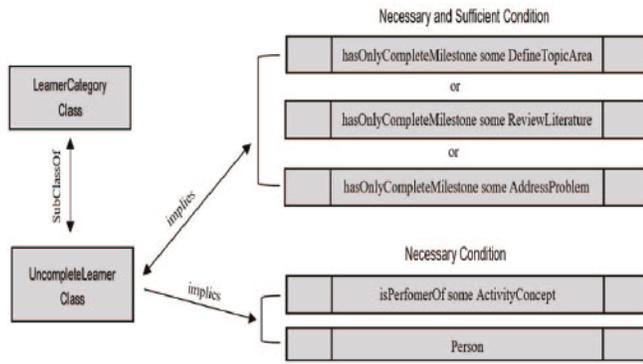


Fig. 11 Object Property Assertion/Attributes for UncompleteLearner Class

Perceptibly, we see that the Object Property Restriction is used to infer anonymous classes that contains all of the individuals that satisfies the restriction, in essence, all of the individuals that have the relationship required to be a member of a particular Class. The consequence is the *necessary and sufficient Condition* which makes it possible to implement and check for consistency in the model which means that it is necessary to fulfil the condition of the universal or existential Restriction - for any individual to become a member of a Class, as we answer using the LQ1 and LQ2 to describe the class SuccessfulLearner (Fig. 10) and UncompleteLearner (Fig. 11).

C. Process Formalization and Ordering of the Semantic Process Algorithm

The following section shows the semantic learning process mining algorithm formalization and ordering for our approach. We show how by constructing transition system and semantic description of the process elements based on the learning event log, it becomes possible for us to determine individual learning patterns and behaviour within the learning knowledge base.

The SLPM algorithm formalization describes the basis for our approach. To explain the strategies for constructing the classification of learning activity concepts and sub sets, we require the following notations, $a, b, c, d \in R$, is a function with domain R and process logs a, b, c, d . The domain R is

a SuperClass of the SubClasses a, b, c, d . The SubClass (also referred to as SubSet) is a set where each of the individual Learning Activity occurs and sometimes may occur multiple times.

SLPM Algorithm Formalization:

Let \mathcal{L} , be a process log for Person, P , over Research process, R , and $a, b, c, d \in R$

where: a = DefineTopicArea Milestone
 b = ReviewLiterature Milestone
 c = AddressProblem Milestone
 d = DefendSolution Milestone

IF $P \dots n$ is a measure of the number of times a, b, c, d occurs in R for Person, P

$$P \dots n = |n \subseteq \mathcal{L} \in R|$$

$$P \dots n = |n \subseteq \mathcal{L}a| \pm |n \subseteq \mathcal{L}b| \pm |n \subseteq \mathcal{L}c| \pm |n \subseteq \mathcal{L}d|$$

THEN

$$\text{SuccessfulLearner, } PSL = |SL \subseteq \mathcal{L} \in R|$$

$$PSL = |SL \subseteq \mathcal{L}a| + |SL \subseteq \mathcal{L}b| + |SL \subseteq \mathcal{L}c| + |SL \subseteq \mathcal{L}d|$$

$$\text{UncompleteLearner, } PUL = |UL \subseteq \mathcal{L} \in R - 1|$$

$$PUL = |UL \subseteq \mathcal{L} \in R - a| \text{ or } |UL \subseteq \mathcal{L} \in R - b| \text{ or } |UL \subseteq \mathcal{L} \in R - c| \text{ or } |UL \subseteq \mathcal{L} \in R - d|$$

For example, $[a1, a2, a3, a4, a2, a5]$ may be the sequence set of learning activity for Person, $P \dots n$ over a , (the DefineTopicArea Milestone Class). i.e. $P \dots n(a) = |n \subseteq \mathcal{L}a|$.

Therefore, IF $a1$ = Define Topic

$a2$ = Approval Activity

$a3$ = Topic decline

$a4$ = Refine Topic

$a5$ = End Topic Proposal

THEN, the sequence set of activities for $P \dots n(a) = \{Define Topic, Approval Activity, Topic Decline, Refine Topic, Approval Activity, End Topic Proposal\}$.

However, for the purpose of the learning question LQ1 and LQ2, our focus is not on the various individual activities that makes up a definitive Class (milestone) but on computing the set of individual process instances that has or not completed a given number of milestones. To complete a given milestone, one must perform the set (or perhaps a subset) of the activities that comprise it. Given the fact for transition purposes, a process instance does not move on to the next milestone without completing a distinctive sequence set of learning activities that makes up the preceding milestone. The sum and difference in process logs for a given number of person, P , is defined in a straightforward way, i.e

$$P \dots n = |n \subseteq \mathcal{L}a| \pm |n \subseteq \mathcal{L}b| \pm |n \subseteq \mathcal{L}c| \pm |n \subseteq \mathcal{L}d|.$$

Therefore, $P \dots n$ is a finite set $|n \subseteq \mathcal{L} \in R|$. For example, we described that Every Person that hasCompleteMilestone a DefineTopicArea and that hasCompleteMilestone a ReviewLiterature and that hasCompleteMilestone an AddressProblem and that hasCompleteMilestone a DefendSolution is a

SuccessfulLearner”. Thus, the Class Successful Learners, **PSL**, will be sum of the set of activities log, \mathcal{L} , that a learner has completed for the milestones **a**, and **b**, and **c**, and **d**.

Hence, IF **PSL** is a Class that consist of the set $|SL \subseteq \mathcal{L}a| + |SL \subseteq \mathcal{L}b| + |SL \subseteq \mathcal{L}c| + |SL \subseteq \mathcal{L}d|$,

THEN **PSL** is the set $|SL \subseteq \mathcal{L} \in R|$.

Similarly, we defined that Every Person that hasOnlyCompleteMilestone a DefineTopicArea or that hasOnlyCompleteMilestone a ReviewLiterature or that hasOnlyCompleteMilestone an AddressProblem is an UncompleteLearner. Therefore, the Uncomplete Learners, **PUL**, is the class of learners where some or set of activities for the milestone **a**, or **b**, or **c**, or **d** is missing over a finite set $|n \subseteq \mathcal{L} \in R|$.

Hence, IF **PUL** is a Class that consist of the set $|UL \subseteq \mathcal{L} \in R-a|$ or $|UL \subseteq \mathcal{L} \in R-b|$ or $|UL \subseteq \mathcal{L} \in R-c|$ or $|UL \subseteq \mathcal{L} \in R-d|$,

THEN **PUL** is the set $|UL \subseteq \mathcal{L} \in R-1|$.

We defined and established in previous section (4(B)) that Successful Learner differs from the Uncompleted Learner through the *Necessary* and *Sufficient Condition* attribute. The assertion implies that for a learner to be regarded as successful they must fulfil the universal restriction by way of the Necessary and Sufficient Condition Object property. Thus, the logic rule expressions in Fig. 12 and Fig. 13;

“Everything that hasCompleteMilestone a DefineTopicArea and that hasCompleteMilestone a ReviewLiterature and that hasCompleteMilestone an AddressProblem and that hasCompleteMilestone a DefendSolution is a SuccessfulLearner”.

On the other hand,

“Everything that hasOnlyCompleteMilestone a DefineTopicArea or that hasOnlyCompleteMilestone a ReviewLiterature or that hasOnlyCompleteMilestone an AddressProblem is an UncompleteLearner.

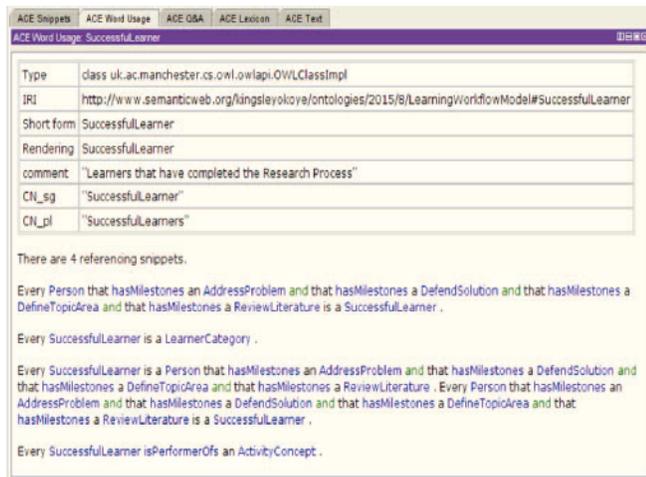


Fig. 12. Referencing Snippet View for the Inferred SuccessfulLearner Class Defined in Protégé.

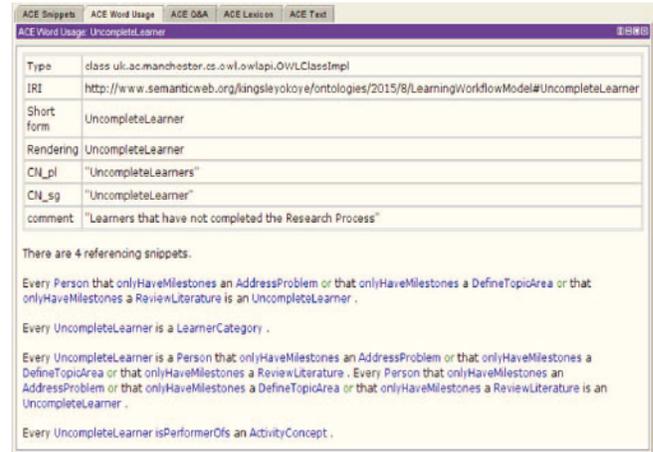


Fig. 13. Referencing Snippet View for the Inferred UncompleteLearner Class Defined in Protégé.

V. RELATED WORKS

Semantic-based Process Mining refers to techniques that systematically integrate domain knowledge especially *formal semantics* into event logs and models about any given process. Research in the area of Semantic Web and Technologies has led to quite mature standards for assembling and modelling process domain knowledge [11]. Currently, Semantic Web Ontologies has become a fundamental tool for information extraction and knowledge processing by providing a structure for distribution of conceptual models about any given process. The OWL [9] has emerged as the standard format for defining Semantic Web ontologies, and is widely accepted and used towards advanced structuring of information and knowledge engineering for enrichment of datasets and depiction of inference rules. According to [8] a well-designed information retrieval or mining system should present results and discovered behaviours in a formal and structured format qua being interpreted as domain knowledge and to further enhance the existing knowledge base. The authors mention that ontology is one of the way to formally represent the mining results as sets of annotated terms and relations towards information extraction and association rule mining especially with Ontology-based Information Extraction (OBIE) systems [12]. [8] also mention that ontology can integrate the use of heterogeneous/unrelated information to guide recommendation systems. According to the authors, ontology-based recommendation system uses ontology for user profiling and personalized search for data resources.

Techniques for semantic process modelling and analysis focuses on information about resources hidden within a process knowledge-base, and how they are related. Semantic-based analysis allows the meaning of the domain object properties to be enhanced through the use of property characteristics and classification of discoverable entities, to permit analysis of the extracted event logs based on concepts rather than the event tags of the process. From learning process application domain perspective, [13] notes that educational process mining is an emerging field in the educational data mining (EDM) discipline, concerned with discovering, analyzing, and improving of educational

processes based on information hidden in datasets and event logs of the process. According to the authors [13], one of the challenges with process discovery and analysis techniques when applied to the educational domain is that they rely exclusively on the syntax of labels in databases, and are very sensitive to data heterogeneity, labelname variation and frequent changes. Due to these challenges, large educational process models are discovered without some kind of hierarchy or structuring. To address these problem, the authors show how by linking labels in event logs to the underlying semantics that describes the discovered models, we can bring educational processes discovery to the conceptual level in order to provide a more accurate and compact mining and analysis of such processes at different levels of abstraction. By extracting educational process models annotated with semantic information, the authors propose a semi-automatic procedure used to associate semantics to training labels. They used the Ontology Abstract Filter plug-in in ProM [14] as input to a semantically annotated event log to produce as output an event log where the names of tasks, i.e. trainings labels, are replaced by the names of a set of chosen concepts. The produced log is then exported as SAMXML [15] file format, and subsequently perform a control-flow mining using the Heuristic Miner algorithm [16] to extract the educational process model based on the concepts that has been derived.

Reference [17] observes that most of the existing techniques for analysing large growing knowledge bases focus on building algorithms to help the knowledge-base automatically or semi-automatically extend. The authors note that the use of an association rule mining algorithm to populate knowledge-base and to improve the relations between the various entities within the knowledge-base is a useful approach considering the fact that most systems constructing large knowledge bases continuously grow, they do not contain all facts for each category, resulting in missing value dataset. To resolve this challenge, the authors developed a new parameter called Modified Support Calculation Measure which generates new and significant rules. They also developed a structure, based on pruning obvious item sets and generalized association rules which decreases the amount of discovered rules in order to help maintain the large growing knowledge-base and rules. In [18] we mention that Association Rule Learning aims at finding rules that can be used to predict the value of some response variables that has been identified as being important but without focusing on a particular response variable. This association aims at creating rules of the form: If X Then Y, where X is often called the antecedent and Y the consequent. Thus, $X \Rightarrow Y$. According to the work in [18] we show that this rule is similar and can be related to the SWRL [10] which is a useful language designed for process description especially to provide an improved learning ontology and enhancement of the process model. The SWRL rule has the form; $\text{atom} \wedge \text{atom} (\text{antecedent}) \dots \rightarrow \text{atom} \wedge \text{atom} (\text{consequent})$. Association rule learning strongly supports the use of such metrics frequently expressed in form of support and confidence. These expressions help in measurement of the strength of the association between learning objects. Support determines how often a rule is applicable to a given

data set which means the fraction of instances for which both antecedent and consequent hold. Hence, a rule with high support is more useful than a rule with low support. A rule that has low support may occur simply by chance and is likely to be irrelevant from a learning perspective because it may not be profitable to monitor, recommend and promote learning activities or patterns.

The authors in [19] notes that pattern discovery algorithms uses statistical and machine-learning techniques to build models that predicts behaviour of captured data. According to the authors, one of the most pattern discovery techniques used to extract knowledge from pre-processed data is Classification. They observe that most of the existing classification algorithms attains good performance for specific problems but are not robust enough for all kinds of discovery problems. [19] propose that combination of multiple classifiers can be considered as a general solution for pattern discovery because they obtain better results compared to a single classifier as long as the components are independent or have diverse outputs. The approach compares the accuracy of ensemble models, which take advantage of groups of learners to yield better results using the Meta Classifier (Staking and Voting) alongside other Base classifiers: Decision Tree algorithm, k-Nearest Neighbour, Naive Bayesian and BayesNet. Explicitly, the works in [20][18][4] shows that the problems of modelling learning processes can be solved by transforming ontology population problem to a classification problem where, for each entity within the ontology, the concepts (classes) to which the entities belongs to have to be determined, hence, classified.

According to the work in [21] Classification is one of the most common data mining technique that aims at finding models or functions that describes or distinguishes data classes or concepts. A useful application of such approach is to annotate the classification labels with the set of relations defined in an ontology especially for use in semantic enrichment of captured data. Semantics encoded in classification tasks has the potential not only to influence the labelled data but also to handle large number of unlabelled data [22][23]. The authors in [23] integrated ontology as consistency constraints into multiple related classification tasks by classifying multiple categories of unlabelled data in parallel to determine labels that violates the ontology. Reference [20] argue that classification is a fundamental task for a lot of intelligent applications, and that classifying through logic reasoning may be both too demanding and frail because of inherent incompleteness and complexity in the knowledge bases. However, they observe that these methods adopt the availability of an initial drawing of ontology that can be automatically enhanced by adding or refining concepts, and have been shown to effectively solve learning modelling problems [4] using Description Logics particularly those based on classification, clustering and ranking of individuals. Reasoning on ontological knowledge plays an important role in the semantic representation of processes such as the learning process. This is possible because semantic reasoning allows the extraction and conversion of explicit information into some implicit information, for instance, the intersection or union of classes, description of relationships and concepts/role assertions.

A large number of developed systems in current literature uses various mining techniques for representation of concepts, knowledge or data which are focused on applying technologies to different aspects of processes [24][25]. Nevertheless, the application of semantic reasoning can help solve the problem of regulating the evolving and static methods for representing such knowledge at theoretical, technological and conceptual levels by making inferences [26][18], retaining and applying what have been learned [27], and discovery and enhancement of new processes [3][4]. In this paper, we apply the semantic-based approach to manage perspective of the learning process mining. The focus is to further enhance this area of research by not only adapting the process mining tools but also present a way to relate semantic-based reasoning for computing various processes within domain knowledge-bases by automatically constructing process models capable of defining, classifying and enhancing observed patterns or behaviours.

VI. CONCLUSION

The work in this paper proves that semantic-based approach can be used to manage perspectives of process mining techniques. We extract streams of event data logs from a learning execution environment, using a case study of research process domain, to describe formats that allows for mining and improved analysis of captured event data logs. The approach makes use of process description languages and semantic annotations to link elements within a learning knowledge-base with concepts that they represent in an ontology specifically designed for representing learning process. By referring to ontologies, the approach provides us with the capability to determine the relationships and attributes the process instances share amongst themselves within the knowledge-base, and then infer and discover hidden patterns automatically by means of semantic reasoning. The semantic analysis allows the meaning of the learning object properties to be enhanced through the use of property characteristics and classification of discoverable entities, and permit analysis of the discovered model based on concepts rather than the event tags of the process.

Future researches could focus on applying the approach described in this paper to analyse streams of events logs that are involved in a different process domain to produce useful knowledge which can be used to load a more enhanced model within the domain area.

REFERENCES

- [1] H.P. Kriegel, K.M. Borgwardt, P. Kröger, A. Pryakhin, M. Schuber and A. Zimek, "Future trends in data mining," *Journal of Data Mining and Knowledge Discovery*. Vol. 15, pp. 87-97, 2007.
- [2] IEEE Task Force on Process Mining, "Process Discovery Contest" In *BPI workshops*, co-located with BPM Conference 2016. Available at: http://www.win.tue.nl/ieeefpm/doku.php?id=shared:process_discovery_contest
- [3] W.M.P. Van der Aalst, "Process Mining: Discovery, Conformance and Enhancement of Business Processes," Springer, 2011.
- [4] K. Okoye, A.R.H. Tawil, U. Naeem and E. Lamine, "Discovery and Enhancement of Learning Model Analysis through Semantic Process Mining*," *Int. Journal of Computer Info. Sysys. and Industrial Mangt. Applications*. ISSN 2150-7988 Volume 8, pp. 093-114, 2016.
- [5] A.K.A de Medeiros, C. Pedrinaci, W.M.P Van der Aalst, J. Domingue, M. Song, A. Rozinat, B. Norton and L. Cabral, "An outlook on semantic business process mining and monitoring," *3rd Int. IFIP Workshop on Sem. Web & Web Sem.*, pp. 1244-1255, 2007.
- [6] Disco Users Guide. Available at: <https://fluxicon.com/disco/files/Disco-User-Guide.pdf>.
- [7] W. Jareevongpiboon and P. Janecek, "Ontological approach to enhance results of business process mining and analysis," *Journal of Business Process Management*, Vol. 19(3) pp. 459-476, 2013.
- [8] D. Dou, H. Wang and H. Liu, "Semantic Data Mining: A Survey of Ontology-based Approaches," In *Proceedings of the IEEE 9th Int. Conference on Semantic Computing*, pp. 244 - 251, 2015.
- [9] OWL Web Ontology Language. <http://www.w3.org/TR/owl-ref/>.
- [10] SWRL: A Semantic Web Rule Lang <http://www.w3.org/Submission/2004/SUBM-SWRL-20040521/>.
- [11] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web". *Scientific American*, Vol 284(5), pp 28-37, 2001.
- [12] D. C. Wimalasuriya and D. Dou. "Ontology-based information extraction: An introduction and a survey of current approaches," *Journal of Information Science*, Vol 36(3), pp 306-323, 2010.
- [13] A. Hicheur-Cairns, J.A Ondo, B. Gueni, M. Fhima, M. Schwarfeld, C. Joubert and N. Khelifa, "Using Semantic Lifting for Improving Educational Process Models Discovery and Analysis," *4th Int. Symp. on Data-driven Process Discv. & Analysis*, Italy, pp. 150-161, 2014.
- [14] ProM Tool. Available at <http://www.processmining.org/prom/start>.
- [15] A.K.A de Medeiros, W.M.P Van der Aalst and C. Pedrinaci, "Semantic Process Mining Tools: Core Building Blocks," In: *ECIS*, June 2008, Galway, Ireland. pp. 1953-1964, 2008.
- [16] A.J.M.M Weijters, W.M.P Van der Aalst and A.K.A de Medeiros "Process Mining with the Heuristics Miner-algorithm," Tech. report, EUT, Eindhoven, BETA working Paper Series, WP 166, 2006.
- [17] R.G.L Miani and E.R.H Junior, "Exploring Association Rules in a Large Growing Knowledge Base," *Int. J. of Comp. Info. Syst. and Ind. Mangt Apps*. ISSN 2150-7988 Vol. 7 (2015) pp. 106-114.
- [18] K. Okoye, A.R.H Tawil, U. Naeem and E. Lamine "A Semantic Reasoning Method Towards Ontological Model for Automated Learning Analysis," In *Advances in Nature and Biologically Inspired Computing*, Springer International Publishing, pp. 49-60, 2016.
- [19] M.H.A Elhebir and A. Abraham, "A Novel Ensemble Approach to Enhance the Performance of Web Server Logs Classification," *Int. J. of Computer Information Systems and Industrial Management Applications* ISSN 2150-7988 Volume 7 (2015) pp. 189-195.
- [20] C. d'Amato, N. Fanizzi and F. Esposito, "Query answering and ontology population: An inductive approach," In: Bechhofer, S., Hauswirth, M., Hoffmann, J., Koubarakis, M., eds.: *Proc. of the 5th Euro. Semantic Web Conf.*, Vol. 5021 of LNCS, 288-302, 2008.
- [21] J. Han and M. Kamber, "Data Mining: Concepts and Techniques". Morgan Kaufmann Publishers Inc., San Francisco, CA, US A, 2005.
- [22] M. Allahyari, K. J. Kochut, and M. Janik, "Ontology-based text classification into dynamically defined topics," In *IEEE International Conference on Semantic Computing (ICSC)*, pages 273-278, 2014.
- [23] N. Balcan, A. Blum, and Y. Mansour, "Exploiting ontology structures and unlabeled data for learning," In *Proceedings of the 30th Int. Conference on Machine Learning*, pages 1112-1120, 2013.
- [24] IEEE Task Force on Process Mining, "Process Mining Manifesto". BPM Workshops LNBP, vol. 99, pp. 169-194. Springer, 2012.
- [25] D. Fahland and W.M.P Van der Aalst, "Repairing process models to reflect reality," In: *Business Process Management*, vol. 7481 of *Lecture Notes in Computer Science*, pp. 229-245. Springer, 2012.
- [26] K. Okoye, A.R.H Tawil, U. Naeem, R. Bashroush and E. Lamine, "A Semantic Rule-Based Approach Supported by Process Mining for Personalised Adaptive Learning," In: *Journal of Procedia Computer Science*, Vol. 37(c), pp. 203-210, 2014.
- [27] E.H.J Nooijen, B.F Van Dongen and D. Fahland, "Automatic Discovery of Data-Centric and Artifact-Centric Processes," *BPM Workshops. Lect N. in Bus. Info. Proc.* Vol. 132, pp. 316-327, 2013