International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 8 (2016) pp. 093-114

Discovery and Enhancement of Learning Model Analysis through Semantic Process Mining*

Kingsley Okoye[1], Abdel-Rahman H. Tawil[1], Usman Naeem[1] and Elyes Lamine[2]

[1] School of Architecture Computing & Engineering, University of East London, UK
{u0926644, A.R.Tawil, U.Naeem}@uel.ac.uk

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

Abstract*: Semantic concepts can be layered on top of existing learner information asset to provide a more conceptual analysis of real time processes capable of providing real world answers that are closer to human understanding. Challenges from current research shows that even though learning data are captured and modelled with acceptable performance to accurately reflect process executions, they are still limited for many process mining analysis because they lack the abstraction level required from real world perspectives. The work in this paper describes a Semantic Process Mining approach directed towards enriching streams of event data logs from a learning process using semantic descriptions that references concepts in an Ontology specifically designed for representing learning processes. The proposed approach involves the extraction of process history data from learning execution environments unfolding how we extract the input data necessary to be mapped unto the learning process logs, which is then followed by submitting the resulting eXtensible Event Streams - XES and Mining eXtensible Markup Language - MXML format to the process analytics environment for mining and further analysis. The consequence is a learning process model which we semantically annotate with concepts they represent in real time using semantic descriptions, and then linking them to an ontology to allow for analysis of the extracted event logs streams based on concepts rather than the event tags of the process. The aim is to provide real time knowledge about the learning process which are more intuitive and closer to human understanding. By referring to ontologies and piloting series of validation experiments, the approach provides us with the capability to infer new and discover relationships the process instances share amongst themselves and to address the problem of determining the presence of different learning patterns within the learning knowledge base. To this end, we demonstrate how data from learning process can be extracted, semantically prepared, and transformed into mining executable formats to enable prediction of individual learning patterns and outcomes through further semantic analysis of the discovered models. Therefore, our approach is grounded on Process Mining and Semantic Modelling Techniques.

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

I. Introduction

Analysis provided by current process mining techniques can be improved by adding *semantic* information to the event logs of the domain processes. Semantic technologies and its application have gained a significant interest within the field of process mining [1]. The advance is Semantic Process Mining which is currently being adopted and technically applied as a tool towards enhancement and improvement of processes derived from logs created through traditional process mining. Several mining algorithms has also been developed which practically have proven to be useful towards process analysis [2][3][4][1][12]. As a result, useful information about how activities depend on each other (workflow) in a process execution environment has been made possible, and essential for extracting models capable of creating new knowledge. Process mining technique is successfully applied for classical mining of processes where each process execution is recorded in terms of events log sequences [4]. Most of the existing process mining techniques depend on tags in event logs information, and therefore, to a certain extent is limited; because they lack the abstraction level required from real world perspectives. Majority of the 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 these tags. Modern tools for collection and analysis of data in all fields of science are providing more and more data with increasing complexity in their structure [5]. This growing complexity is proved by the need for richer and more precise description of real-world objects that allows for flexible exploration of the objects/data type. According to [5] future development will be to find richer patterns by developing systems which derive understandable patterns as well as making the discovered patterns explicable. Thus, the need for systems capable of providing platform for pattern exploration where users can browse for knowledge they might consider as interesting. These challenges have paved way for Semantic Process

Mining which takes the advantage of the rich semantics described in event data of a process, and links them to concepts in an ontology in order to extract useful models 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 a 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 in the information knowledge base.

One of the benefit provided by semantic process mining is the ability to describe the semantics behind the tags in an event log considered useful for discovery of new knowledge. The main opportunity is that this analysis are enhanced because it is based on concepts rather than the event tags. These semantic viewpoint is captured by annotating the elements in the systems based on two probes (i) *how to make use of the semantic data*, and (ii) *how to mine the semantic information* [6]. Semantic process mining is a new area in the field of process mining and there are few existing applications that demonstrates the capabilities of the technique [7][1].

In this paper, we introduce a Semantic Process Mining approach directed towards discovering and improvement of the set of recurrent behaviours that can be found within a learning executing environment following the work in [8]. The proposed algorithm is developed in order to address the problem of determining the presence of different learning patterns in process models. The standpoint for our approach is based on the following objectives to;

- Show how data from learning processes can be extracted, semantically prepared and transformed into mining executable formats for improved learning model analysis.
- Load a more enhanced model for learning which is useful towards the development of learning process mining algorithms that are more intelligent, predictive and robotically adaptive.
- Prove how semantic process mining can be utilized to address the problem of analyzing concepts and relationships amongst learning objects, which also aid in discovering of new and enhancement of existing learning processes.

The technique involves the extraction of process history data from the learning execution environment, which is then followed by submitting the resulting event streams format to the process analytics environment for mining and further analysis. The focus is on identifying data about different process instances within a learning model, using a case study of research process, and enriching the information values of the resulting model based on the captured user profiles. The learning activity logs is enriched using semantic annotations that references concepts in an ontology specifically designed for learners.

The rest of the paper is structured as follows; in Section 2, we provide a conceptual overview of our approach to Semantic Learning Process Mining. We also provide an example of a research learning process execution environment which we use to illustrate our approach throughout this paper in section 3. Section 4 presents learning process mining in action to show information about resources hidden within the learning process, and how they are connected to our model to enable a more effective reasoning and tactical strategies for adaptation and decision making. In addition, we describe how we extract the input data into mining executable formats necessary to be mapped unto the event logs for improved process analysis. Section 5, presents the semantic preparation and annotation of the learning data for Semantic Learning Process Mining (SLPM) and how we semantically apply the representations of the learning process towards provision of an ontological model for learning. Further, we show the ontology model and how semantic concepts and schema can be layered on top of the extracted learner information asset to provide more enhancements to the learning process model; through concept matching and semantic reasoning. This step is then then followed by application of the SLPM algorithm in section 6 to draw conclusions and make predictions based on the analysis of available event streams and describing in details its implementation and mining outcomes. In Section 7, appropriate related work is analyzed and discussed. Finally, Section 8 concludes the paper and points out direction for future research.

II. General System Overview

The main focus for designing the Semantic Learning Process Mining algorithm is to extract, semantically prepare and transform event data of learning process into mining executable formats that allows us to perform an improved learning process analysis, and then build a semantic model to represent the deployed model. The primary aim is to provide platform that allows us to carry out effective reasoning on the resulting ontologies to infer and answer questions about relationships the process instances share amongst themselves within the learning knowledge-base. We load a more enhanced model for learning which is useful towards the development of learning process mining algorithms that are more intelligent, predictive and robotically adaptive, and then prove how semantic process mining can be utilized to address the problem of analyzing concepts and relationships amongst learning objects.

Figure 1 shows that the development of Semantic Learning Process Mining algorithms entails three building blocks; Annotated Event Logs, Ontologies and Semantic Reasoning that aim at discovering, conformance and extension of Learning processes.

Accordingly, the semantic learning process mining algorithm makes use of semantic annotations to link elements in the event logs 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 the process instances share within the knowledge-base and then infer and discover learning patterns automatically by means of semantic reasoning. The purpose is to perform a more conceptual analysis capable of providing real world answers that are closer to human understanding.

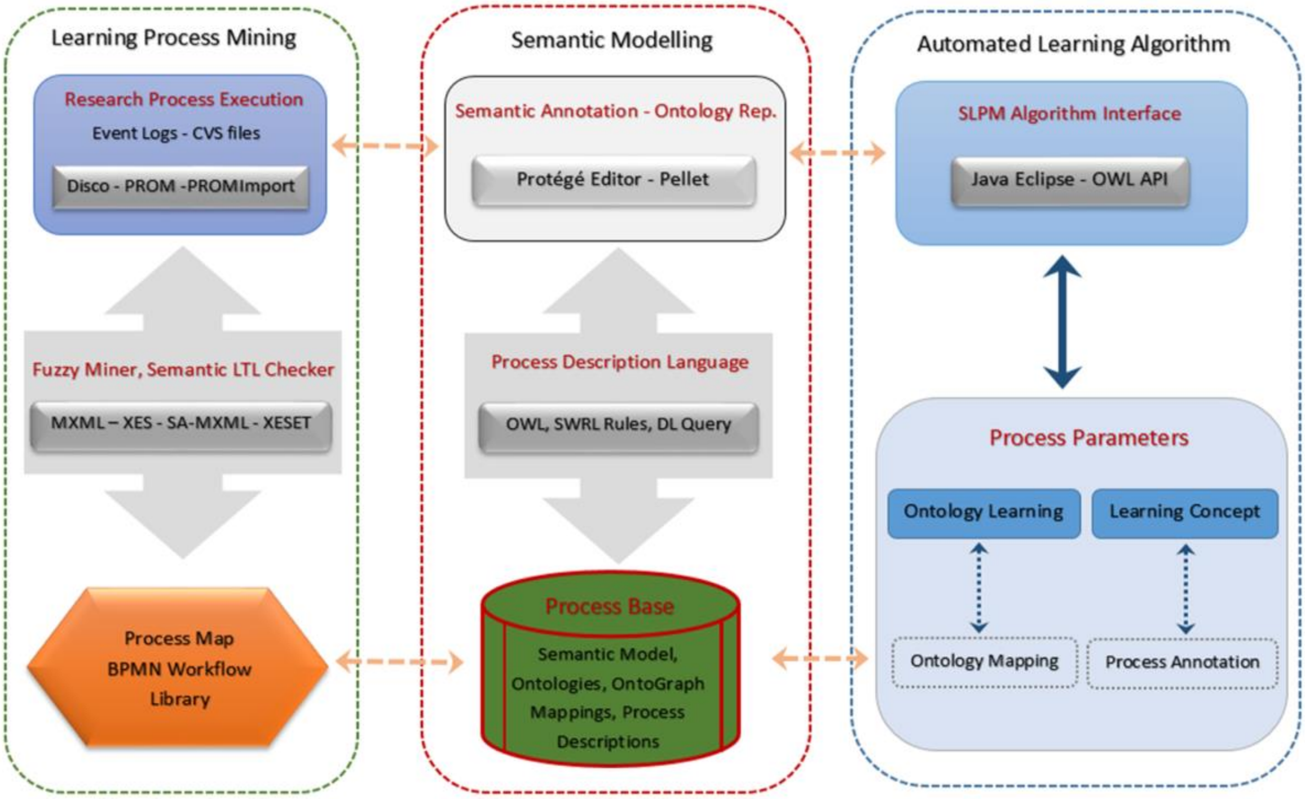


Figure 1. Conceptual overview of Semantic Learning Process Mining Algorithm.

III. Running Example of Learning Process

In this section, we describe a running example of learning process in execution which we use to demonstrate our approach throughout this paper. A distinctive example of a learning process is a Research Process. First step to conducting a research is to decide on what to investigate i.e., the research topic, and then go about finding answers to the research questions. The process constitute of workflow of the journey from choosing the research topic to completing the research, and comprises a 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 learner travel along the research process. At each phase of the process, the learner is required to choose from a variety of method or procedures which will help in achieving the research goal.

Our aim is to adopt tools and techniques that helps create understanding and enhancement of information values extracted from the learning execution environment, and finding the best possible outcome while ensuring validity and reliability. The focus is to make knowledge available for real-time use in a form which is adapted to the context of use and to the needs and cognitive profile of the users [9] as we show and apply to the domain of learning process in this paper. The authors in [9] referred to this practice as Knowledge Mobilization. According to the authors the approach contributes to research on the Semantic Web, as it develops new methods for knowledge activation and ontology building.

We focus on providing an automated learning approach that is capable of providing real world answers about a research learning process that are closer to human understanding through the use of conventional process mining and semantic modelling technique. The main purpose is to decide, describe, justify and explain how learners go about finding answers to the research questions, determine the sequence set of activities that takes place, and what further improvement is needed or may be required progressively through time [10][11].

In Figure 2, we show that the flow of processes from the definition of topic area to award of degree; consist of different learning steps which a learner has to or partly perform in order to complete the research process. The order in which these learning steps are carried out has the capability of determining time of completing the research, as well as reliability of the research outcome. To construct process transition and information about learning activities within the learning model, it is necessary to look at learning events and the immediate preceding process instances that maps the learning transitions.

We provide Four milestones; Establish Context → Learning Stage → Assessment Stage → Validation of Learning Outcome, in order to determine and explain the steps taken within the learning process model. The four stages are based on the rational that a process instance enters the model at a particular point in time and not on the whole transition during the lifecycle of the model; from Defining the Topic Area –to- Review Literature –and- Addressing the Problem –then- Defending the Solution.

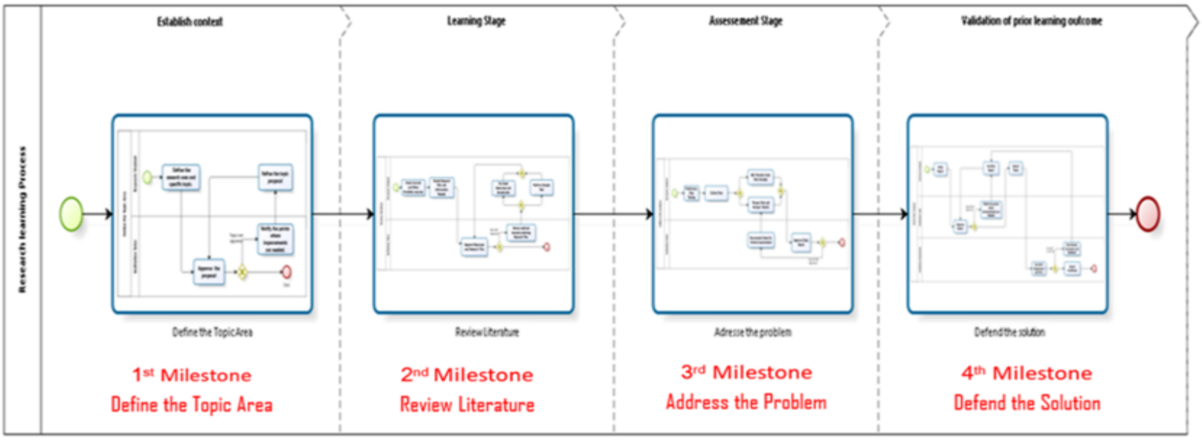


Figure 2. The Four Milestones of the Research Process Model with Bizagi BPMN Modeller

IV. Learning Process Mining and Semantic Preparation of Data

Process mining techniques are not only relevant during the identification and design stage of a learning process but also for the monitoring and enhancement of the whole process. Process mining builds on Data Mining and Process Modelling techniques. The technique focuses on information about resources hidden within a process knowledge-base, and how they are related. In this paper, we use some of the mining techniques to put the captured volumes of data within a learning knowledge-base into a process context, since, in our approach, we represent data at two levels; *process level* and *data level*.

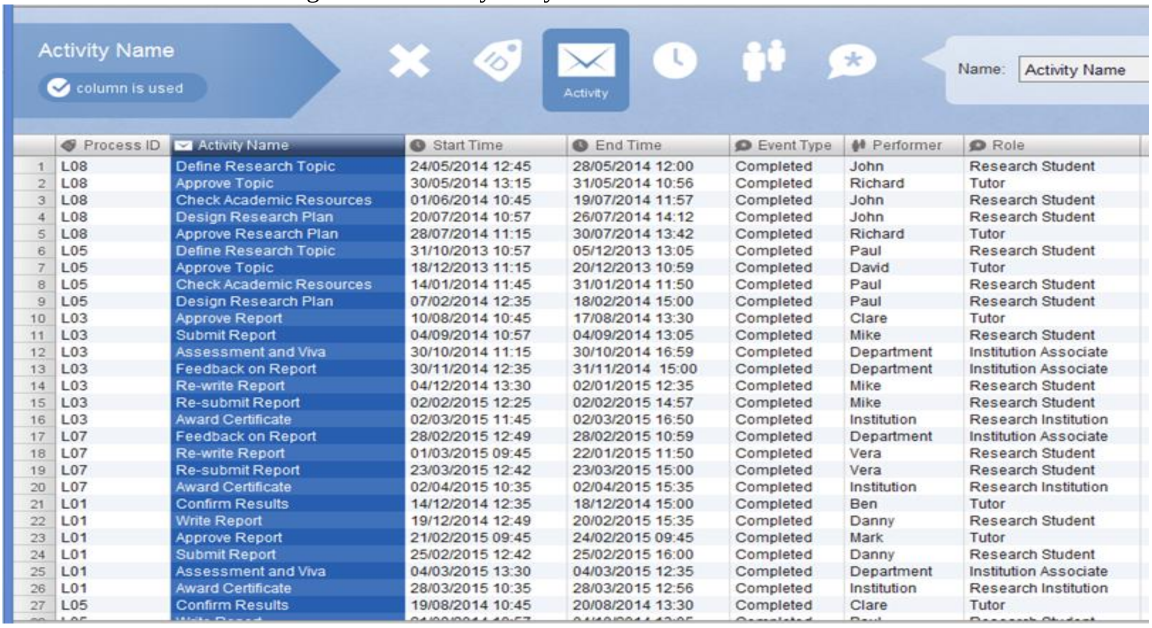
This makes it necessary for us to describe processes that are useful and how they are connected to our model. We focus on transforming the existing raw data extracted from the learning execution environment into meaningful and useful information that can be used to enable more effective reasoning and tactical strategies for adaptation and decision making that are directed towards the data resources/process instances. According to Van de Aalst et al [4] the input data is most often given as a table, as we utilized in Table 1 and Figure 3, and the resulting data sets are often patterns, equations, graphs, tree structures, clusters or rules.

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 11:45	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
L03	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 13:05	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
L01	Confirm Results	14/12/2014 12:35	18/12/2014 15:00	Completed	Ben	Tutor
L01	Write Report	19/12/2014 12:49	20/02/2015 15:35	Completed	Danny	Research Student
L01	Approve Report	21/02/2015 09:45	24/02/2015 09:45	Completed	Mark	Tutor
L01	Submit Report	25/02/2015 12:42	25/02/2015 16:00	Completed	Danny	Research Student
L01	Assessment and Viva	04/03/2015 13:30	04/03/2015 12:35	Completed	Department	Institution Associate
L01	Award Certificate	28/03/2015 10:35	28/03/2015 12:56	Completed	Institution	Research Institution
L05	Confirm Results	19/08/2014 10:45	20/08/2014 13:30	Completed	Clare	Tutor

Table 1. Example of event data for learning processes

A. *Learning Process Mapping*

In this section of the paper, we describe how we extract the input data necessary to be mapped unto the learning process logs. Mapping step is indispensable especially when our aim is to make the semantics about the learning data available to the process mining and analysis tool [2][12]. The works in [2] and [12] shows that semantic annotation of process execution data of-ten requires the mapping of concepts to instances stored within the learning knowledge base. The standard format for storing event logs as shown in Table 1, is by using XML-based formats such as Mining eXtensible Markup Language (MXML) [7] and eXtensible Event Streams (XES) [13]. MXML and XES are standard event log formats used by many process mining algorithms, where process or activity names are normally assumed to be unique by assigning a case identifier and/or using both start and end times to obtain activity durations. According to [4] XES is the successor of MXML. Figure 3 and 4 shows the first mining approach that provides us with reliable and trustworthy results for data sets of arbitrary complexity based on the proven framework of the Fuzzy Miner [14] which we used to create and map the MXML and XES file formats for further extension and analysis.



	Process ID	Activity Name	Start Time	End Time	Event Type	Performer	Role
1	L08	Define Research Topic	24/05/2014 12:45	28/05/2014 12:00	Completed	John	Research Student
2	L08	Approve Topic	30/05/2014 13:15	31/05/2014 10:56	Completed	Richard	Tutor
3	L08	Check Academic Resources	01/06/2014 10:45	19/07/2014 11:57	Completed	John	Research Student
4	L08	Design Research Plan	20/07/2014 10:57	26/07/2014 14:12	Completed	John	Research Student
5	L08	Approve Research Plan	28/07/2014 11:15	30/07/2014 13:42	Completed	Richard	Tutor
6	L05	Define Research Topic	31/10/2013 10:57	05/12/2013 13:05	Completed	Paul	Research Student
7	L05	Approve Topic	18/12/2013 11:15	20/12/2013 10:59	Completed	David	Tutor
8	L05	Check Academic Resources	14/01/2014 11:45	31/01/2014 11:50	Completed	Paul	Research Student
9	L05	Design Research Plan	07/02/2014 12:35	18/02/2014 15:00	Completed	Paul	Research Student
10	L03	Approve Report	10/08/2014 10:45	17/08/2014 13:30	Completed	Clare	Tutor
11	L03	Submit Report	04/09/2014 10:57	04/09/2014 13:05	Completed	Mike	Research Student
12	L03	Assessment and Viva	30/10/2014 11:15	30/10/2014 16:59	Completed	Department	Institution Associate
13	L03	Feedback on Report	30/11/2014 12:35	31/11/2014 15:00	Completed	Department	Institution Associate
14	L03	Re-write Report	04/12/2014 13:30	02/01/2015 12:35	Completed	Mike	Research Student
15	L03	Re-submit Report	02/02/2015 12:25	02/02/2015 14:57	Completed	Mike	Research Student
16	L03	Award Certificate	02/03/2015 11:45	02/03/2015 16:50	Completed	Institution	Research Institution
17	L07	Feedback on Report	28/02/2015 12:49	28/02/2015 10:59	Completed	Department	Institution Associate
18	L07	Re-write Report	01/03/2015 09:45	22/01/2015 11:50	Completed	Vera	Research Student
19	L07	Re-submit Report	23/03/2015 12:42	23/03/2015 15:00	Completed	Vera	Research Student
20	L07	Award Certificate	02/04/2015 10:35	02/04/2015 15:35	Completed	Institution	Research Institution
21	L01	Confirm Results	14/12/2014 12:35	18/12/2014 15:00	Completed	Ben	Tutor
22	L01	Write Report	19/12/2014 12:49	20/02/2015 15:35	Completed	Danny	Research Student
23	L01	Approve Report	21/02/2015 09:45	24/02/2015 09:45	Completed	Mark	Tutor
24	L01	Submit Report	25/02/2015 12:42	25/02/2015 16:00	Completed	Danny	Research Student
25	L01	Assessment and Viva	04/03/2015 13:30	04/03/2015 12:35	Completed	Department	Institution Associate
26	L01	Award Certificate	28/03/2015 10:35	28/03/2015 12:56	Completed	Institution	Research Institution
27	L05	Confirm Results	19/08/2014 10:45	20/08/2014 13:30	Completed	Clare	Tutor

Figure 3. Running example of Event Data in Discovery Miner.

In Figure 3, we imported the set of event data of our learning process into Disco [15] to show in details how the processes have been performed. The approach reveals the process mapping and provides us with the opportunity to focus on the stream of behaviours, and to see the paths they follow in the process model. *Case id tags* are used to assign the identifier for process instances and *Activity tags* for the set of task that are performed during the learning process. We associate *Timestamp tags* with 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. This is shown using the *Frequency Analysis* to determine how often a given process is performed. The variants show the process in a more detailed manner by revealing all the cases that has been created during process execution. Accordingly, the most frequent variants are also determined in the model.

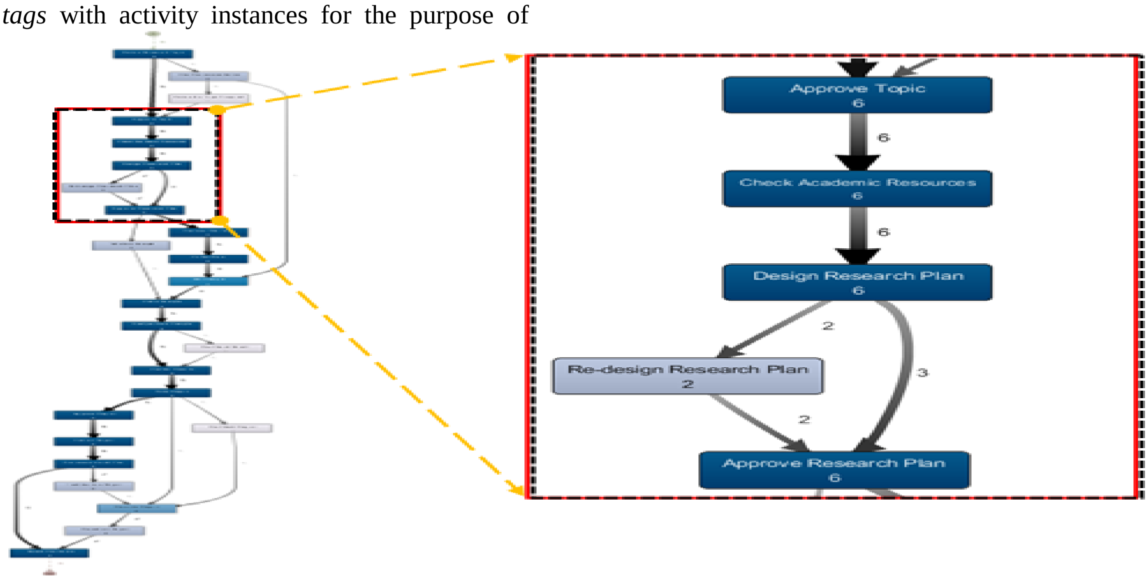


Figure 4. Control-flow of Mapped Processes in execution.

While the Map view in Figure 4 gives us an understanding about the process flows, and the Statistics view provides a detailed performance metrics about a process; the Cases view actually goes down to the individual case level and shows the raw data. In order to inspect individual cases, it is important to verify the findings and see concrete examples particularly for strange behaviour that will most likely occur during the process analysis. In total, there are six powerful filter types available in Disco, and they can be combined and stacked in any order. However, we focus on the *Attribute filter* [15] which describes as well as exclude certain activities, resources, or process categories based on data attributes. In addition to the analysis views, the filtering capabilities allows us to quickly and interactively explore processes into multiple directions and to answer concrete questions about the learning process and more importantly allows us to further model and hold inference reasoning to generate process improvement ideas along the way. The resulting format is XES schema (Figure 5) which is less restrictive and truly extendible based on the different attribute types: *Concept*, *Life-Cycle*, *Organizational*, *Time* and *Semantics*.

Figure 5 shows the XES file format which provides us with the capability of further extending the Learning model. The format reveals the number of activity clusters and elements which are executed from the start of *Define Research Topic* event to the completion of *Award Certificate*. One of the usefulness of XES syntax is that; it does not only provide semantics for commonly used attributes but also provides semantic extension which are capable of using the semantics of the data captured in event logs to create new knowledge or enhance existing ones [4].

```
<global scope="event">
    <string key="concept:name" value="name"/>
    <string key="lifecycle:transition" value="transition"/>
    <string key="org:resource" value="resource"/>
    <date key="time:timestamp" value="2015-04-05T14:30:13.876+01:00"/>
    <string key="Activity_Name" value="string"/>
    <string key="Event_Type" value="string"/>
    <string key="Performer" value="string"/>
    <string key="Role" value="string"/>
</global>
<classifier name="Activity" keys="Activity_Name"/>
<classifier name="Resource" keys="Performer"/>
<string key="lifecycle:model" value="standard"/>
<string key="creator" value="Fluxicon Disco"/>
<string key="library" value="Fluxicon Octane"/>
<trace>
    <string key="concept:name" value="L08"/>
    <string key="creator" value="Fluxicon Disco"/>
    <event>
        <string key="concept:name" value="Define Research Topic"/>
        <string key="lifecycle:transition" value="start"/>
        <string key="org:resource" value="John"/>
        <date key="time:timestamp" value="2014-05-24T12:45:00.000+01:00"/>
        <string key="Activity_Name" value="Define Research Topic"/>
        <string key="Event_Type" value="Completed"/>
        <string key="Performer" value="John"/>
        <string key="Role" value="Research Student"/>
    </event>
```

Figure 5. Fragment of the XES file format.

At present, XES is supported by process mining tools such as ProM, Disco, Open XES and XESame [13]. According to [4] there are five standard extensions of XES defined in terms of the *Concept*, *Life-Cycle*, *Organizational*, *Time* and *Semantics*.

The semantic extension (Figure 6) is inspired by the Semantic Annotated version of the MXML [7] through definition of the *model Reference* attribute for all elements in the logs.

```
<ProcessInstance id="L08" description="" modelReference="file:/C:/Users/KINGSLEY/Desktop/Process%20mining%20files/ProcessInstanceOntology.
wsml#ProcessInstance " >
  <AuditTrailEntry>
    <Data>
      <Attribute name="Role" modelReference="file:/C:/Users/KINGSLEY/Desktop/Process%20mining%20files/DataFieldOntology.wsml#Role " >Research
      Student</Attribute>
      <Attribute name="Event Type" modelReference="file:/C:/Users/KINGSLEY/Desktop/Process%20mining%20files/DataFieldOntology.wsml#EventType " >
      Completed</Attribute>
    </Data>
    <WorkflowModelElement modelReference="file:/C:/Users/KINGSLEY/Desktop/Process%20mining%20files/TaskOntology.wsml#DefineResearchTopic " >Define
    Research Topic</WorkflowModelElement>
    <EventType modelReference="file:/C:/Users/KINGSLEY/Desktop/Process%20mining%20files/EVO.wsml#Start " >start</EventType>
    <Timestamp>2014-05-24T12:45:00.000+01:00</Timestamp>
    <Originator modelReference="file:/C:/Users/KINGSLEY/Desktop/Process%20mining%20files/OriginatorOntology.wsml#John " >John</Originator>
  </AuditTrailEntry>
```

Figure 6. Fragment of SA-MXML file format for our Learning Process Logs.

The Semantic Annotated Mining eXtensible Markup Language (SA-MXML) format - Figure 6, is a semantic annotated version of the MXML format. The SA-MXML file incorporates an additional attribute called the *model Reference* for all elements in the log except for *Audit Trail Entry* and *Timestamp*. The *model Reference* attribute points and links between elements in the learning logs and a list of concepts within the learning ontology, which is an important structure towards implementing our Semantic Learning Process Mining technique. The SA-MXML format is supported by tools such as ProM 5.2 [16] and ProMImport [17] open source frameworks for process mining algorithms. This extension incorporates references between elements in logs and concepts in ontology; which is a great way to define processes. The reference associates meaning to tags in event logs by pointing to concepts defined in an ontology. For instance, there may be an ontology describing different kinds of Learners. By means of the *model Reference* attribute, a trace can point to this ontology to classify the Learners. Thus, we describe the class Learner to be a subclass of Learning Process - the *necessary condition* is: if something is a Learner, it is *necessary* for it to be a participant of the class Learning Process and *necessary* for it to have a kind of sufficiently defined condition and relationship with other classes say LearningActivity, LearningInstitution, Department, etc.

The approach allows the meaning of Learning objects/properties to be enhanced through the use of *property characteristics* and *classification of discoverable entities* and then utilize the main function offered by the Reasoner to help in checking for consistency in the resulting domain, to test whether or not a class is a subclass of another class, or checking whether or not it is possible for a class to have any instances. This means a class is said tobeinconsistent if it does not have any instances. By performing such test (i.e. *Classification*) it becomes possible for the Reasoner to compute the inferred Learners or activity hierarchy.

B. *Semantic Preparation and Annotation of Learning Workflow Library*

In this section of the paper, we model the mapped processes using Business Process Mining Notations (BPMN) and apply semantics concepts to describe the process map workflow of the learning activities as they happen in reality. The purpose is to allow us to semantically represent the Workflow Activity Patterns (WAPS) of the learning concepts based on the sequence flow of learning tasks to show in detail how the various individual processes has been performed. Semantically enriching process execution data has the capability to successfully raise the process analysis from the syntactic to the semantic level, and enable multiple perspectives of analysis on the domain process [2][6]. According to the authors in [2], when data are gathered and transformed into a unified format, they become a primary target for semantic enrichment which are useful towards obtaining multiple levels of abstraction for semantic process mining and analysis.

Figure 7, 8, 9 and 10 is a semantic representation of the learning concepts and association of entities within the deployed learning model based on sequence flow of the learning activities.

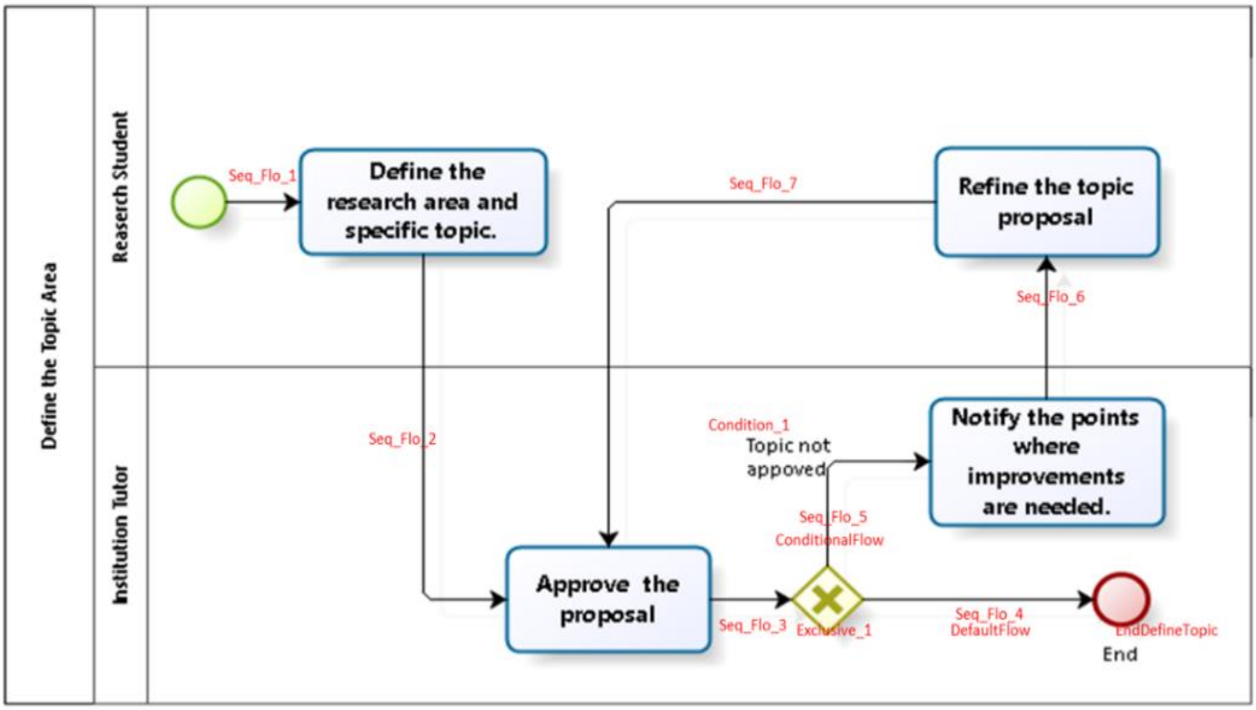


Figure 7. 1st Milestone - Define Topic Area: Seq_Flo_1 to Seq_Flo_7.

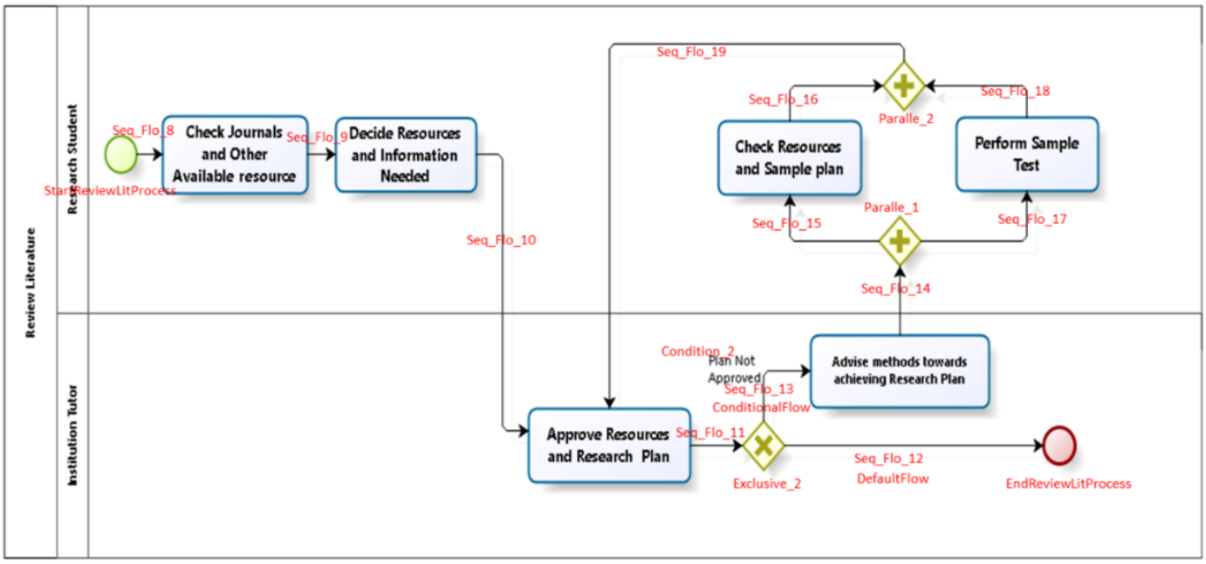


Figure 8. The 2nd Milestone – Review Literature: Seq_Flo_8 to Seq_Flo_19.

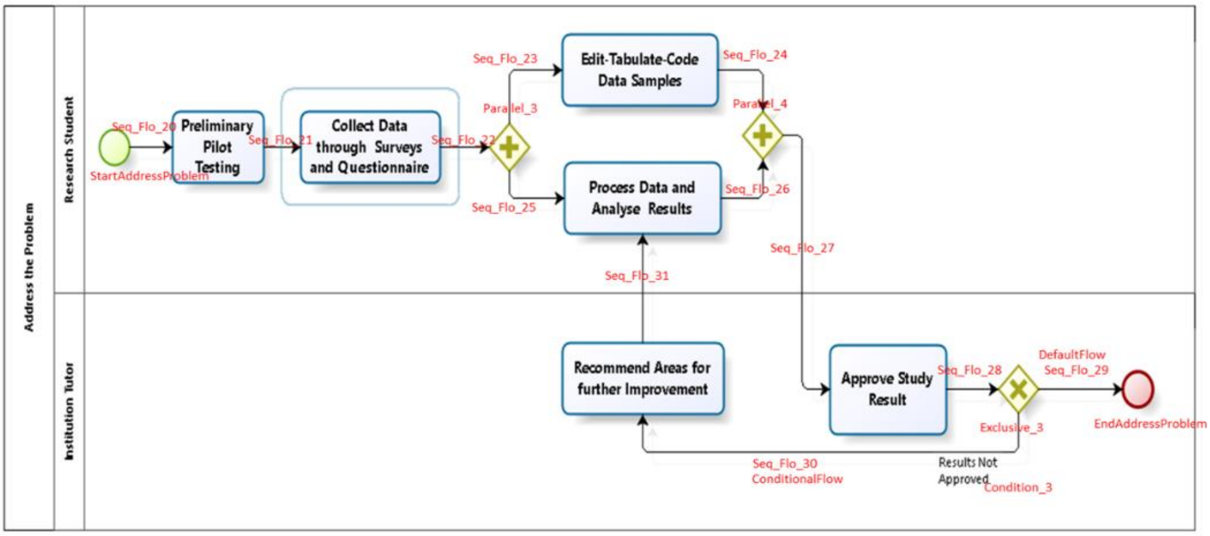


Figure 9. The 3rd Milestone – Address the Problem: Seq_Flo_20 to Seq_Flo_31.

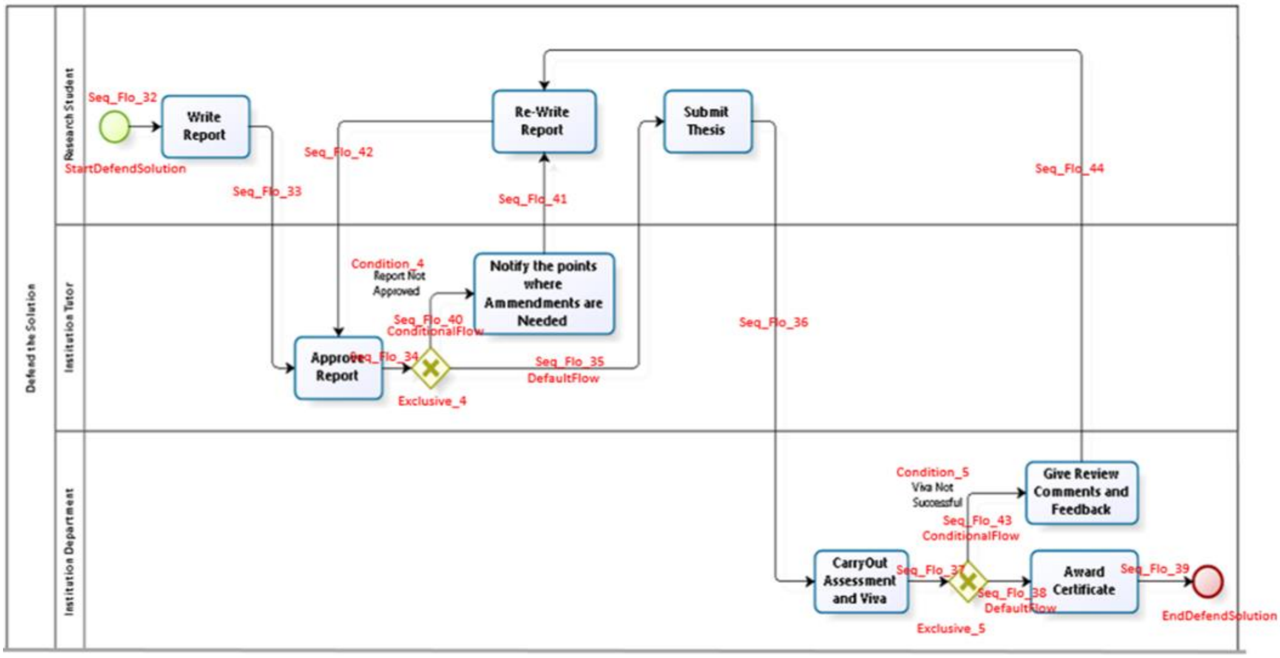


Figure 10. The 4th Milestone – Defend Solution: Seq_Flo_32 to Seq_Flo_44.

V. Semantic Learning Process Mining and Analysis

Semantic process mining is useful in addressing the problem of analyzing concepts and relationships amongst item-sets. The approach is suitable in formulation of robust and sharable descriptions of processes for an enhanced reasoning capability as well as increase in knowledge awareness and performance. According to [7] the development of semantic process mining tools entails three building blocks *Annotated Event Logs, Ontologies* and *Semantic Reasoning* that aim at discovering, conformance and extension of processes.

In previous sections, we show how data is being extracted, prepared and transformed into extensible formats that allows for Semantic Learning Process Mining (SLPM) to be implemented. In the next sections, we reveal how semantic concepts and process descriptions can be layered on top of the extracted learner information asset to provide more enhancements to the learning model through concept matching (ontology classification) and semantic reasoning as shown in Figure 11. Semantic process mining aims at analyzing the extracted event logs streams based on *concepts* rather than the event *tags* of the process with the primary aim of providing real time knowledge about the learning process which are intuitive and closer to human understanding. The key step is to provide a semantic model for the event logs to semantically represent elements in the learning model with concepts that they represent in real time, by linking them to an ontology. By referring to ontologies, the approach provides us with the capability to infer and discover relationships among the concepts and process instances.

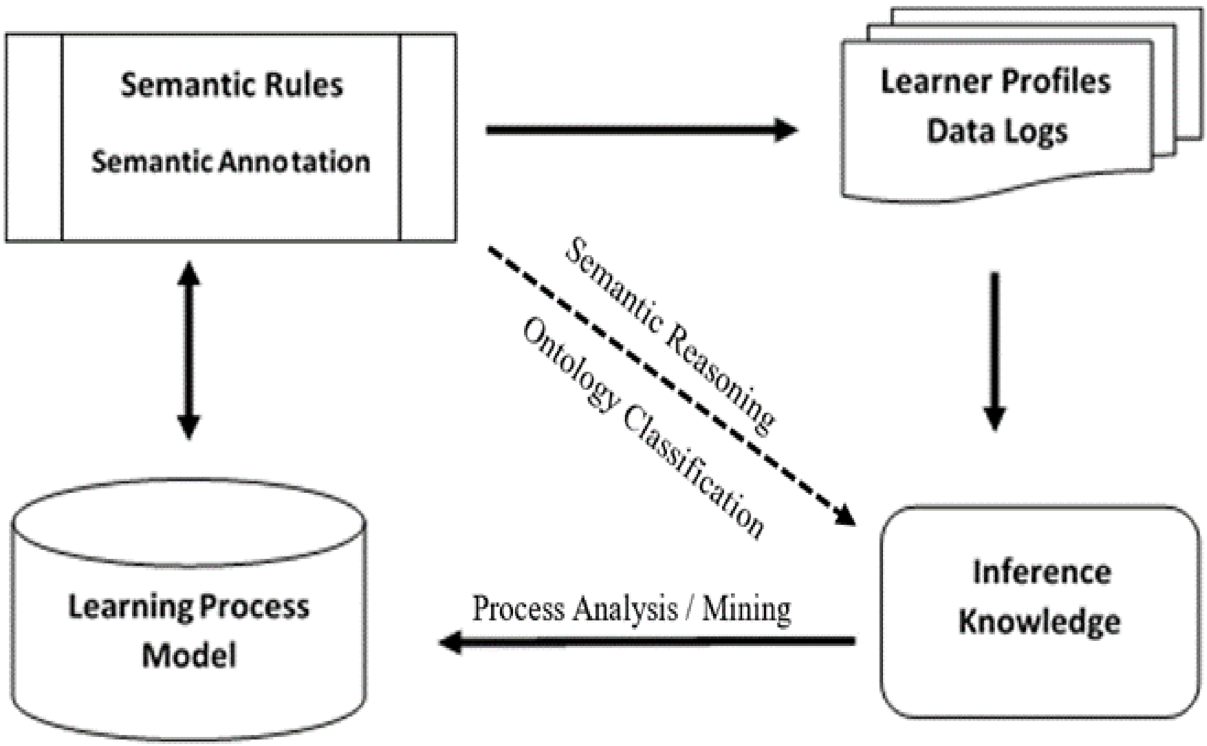


Figure 11. Framework for implementation of Semantic Learning Process mining and Analysis.

Reasoning over the ontologies with reference to elements within the learning model provides us with a robust way to answer questions about relationships which the elements (process instances) share amongst themselves, and to perform a more conceptual analysis capable of providing real world answers that are closer to human understanding. Few algorithms has been developed in literature which has the capability of performing this kind of semantic analysis such as the Semantic LTL Checker proposed by [7] which applies concepts in an ontology as input to parameters of a Linear Temporal Logic Formulae to formulate and answer questions about process instances and their relationships, using the WSML2 Reasoner to infer all the necessary associations. The association of patterns reveals interesting connection among domain entities, the individual cases and object/data types to provide a better under-standing of how the different elements within the Learning Process Knowledge base relate and interact with each other.

To demonstrate this approach, in the next sections we present how automation of learning process and semantic representation of the flow of activities within the learning knowledge-base (technically described as Workflow) can be used to provide inference knowledge and allow the meaning of properties to be enhanced through the use of *property characteristics* and *classification of discoverable entities* as shown in section **V**.*(A)*, V.*(B)* and **V**.*(C)*

A. Semantic Modelling: Ontological Representation of Learning Workflow Sequence with BPMN and Learning Activity Concepts

Being able to use the Reasoner to automatically compute the class hierarchy and control flow of activities within a learning knowledge-base is one of the major benefits of building semantic mod-el using ontology. Ontology provides us with benefits in discovery, flexible access and information integration due to the inherent connectedness (inference), concept matching and reasoning capability. This characteristic is the ability to match same idea as well as use the coherence and structure itself to inform and answer questions about relationships the learning objects (process in-stances) share amongst themselves within the learning knowledge-base.

An important aspect of maintaining semantic models and the processes they support is the capability to analyse them. This analysis can be performed in real-time and most often lead to some new knowledge being discovered, and then used to inform about the individual entities involved in the process due to the semantic perspective captured by annotating the elements in the model. The ontology association process tends to help solve the semantic problem of using tags and meaning by individuating a collection of similar entities which may belong to different ontologies, and then enabling a full understanding among different actors involved in exchanging information within the knowledge-base [18].

The main opportunity is that this analysis can be enhanced because they are based on concepts rather than tags. Process mining techniques has been used to discover and enhance ontologies as well as automatically infer concepts with elements that are not semantically annotated but that belong to partially annotated event logs or models [6]. According to the authors in [6] the use of ontologies provides two opportunities for process mining techniques. The first opportunity is to make use of the ontological annotations in the event logs or model to develop more robust process mining techniques that analyses the event logs or model at the concept level by linking them to ontologies. The second opportunity is to use process mining techniques to discover and enhance ontologies based on the data in the event logs.

For our approach, we show through ontological representation of the learning model workflow sequence and the association with BPMN activity concepts - Figure 12 & 13, the capability of using semantics to classify instances to explain the dependent variables in terms of independent ones; which is a great way to compliment the way we look at processes. We prove that the various units/activities within a learning process model can be related to exactly one case and assigned a case identifier [2] which results in automatic creation of workflow processes [3] and can help to maintain the resulting hierarchy correctly. This technique is made possible by using semantic an-notation scheme and vocabularies to represent the sets of various entities, properties and classes within the learning knowledge base and then create inferences capable of providing new knowledge and a richer set of intelligence within the resulting model. The approach associates new content with prior knowledge which leads to unrelated data being discovered, examined and further grouped and labelled in order to draw conclusions as well as make predictions based on the analysis of the data.

In Figure 12 & 13, we use Protégé Editor [19] to construct an ontology that expresses the functionality of our learning model in terms of individual learning characteristics. Annotation properties were used to add information (Metadata – data about data) to the classes, individuals and object/data properties in the ontology which allows the meaning of properties to be enhanced through the use of property characteristics and classification of discoverable entities.

Semantic representation of data about the learning process is an important tool to-wards unlocking the information value of the event data within the model, by way of finding useful and previously unknown links between the activity concepts. The motivational perspective is the search for explanatory and predictive patterns within the learning domain especially with regards to the large volume of data that are involved. The authors in [20] refers to this tactics as Creative Knowledge Discovery which is concerned with the creation of new and effective patterns either by generalization of existing patterns or by analogy to patterns embedded in other domains. According to the authors in [20] an important prerequisite for the approach is that we understand the relations within the data, thereby allowing us to find structures that are hidden in the event data and to extract novel concepts that can be utilized for subsequent processing/analysis.

The ability to analyse information and create concepts is fundamental to ontological representation/modelling of event data, and in use of Reasoner to infer process instances. This techniques can be applied towards automation of learning processes and the extraction of useful models, as we describe below:

- Create the Learning Domain, BPMN concepts, Classes and Individuals that will be inferred.
- Provide Process Descriptions for all Object and Data Types that allows for Query and Reasoning (*Class_Assertions; Object_Property_Assertions; Data_Property_Assertions*).
- Create SWRL rules to map the existing classes with BPMN concepts.
- Check for Consistency for all Defined Classes within the Model using DL queries.

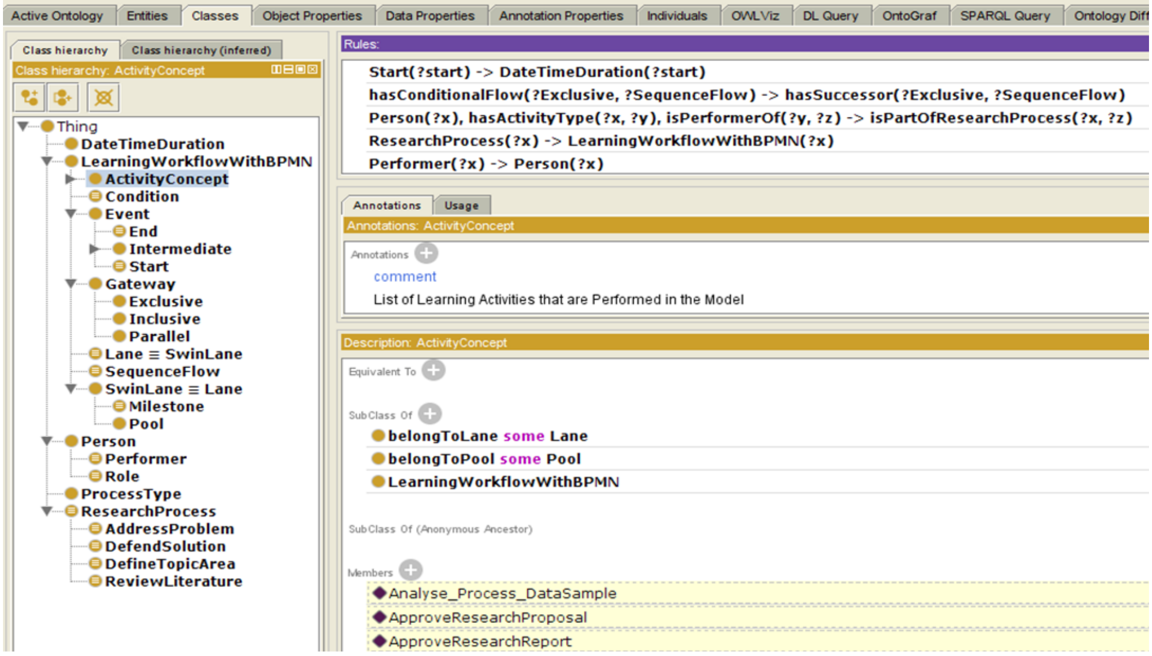


Figure 12. Learning Workflow Ontology with BPMN Concepts, Domain Classes and SWRL Rules

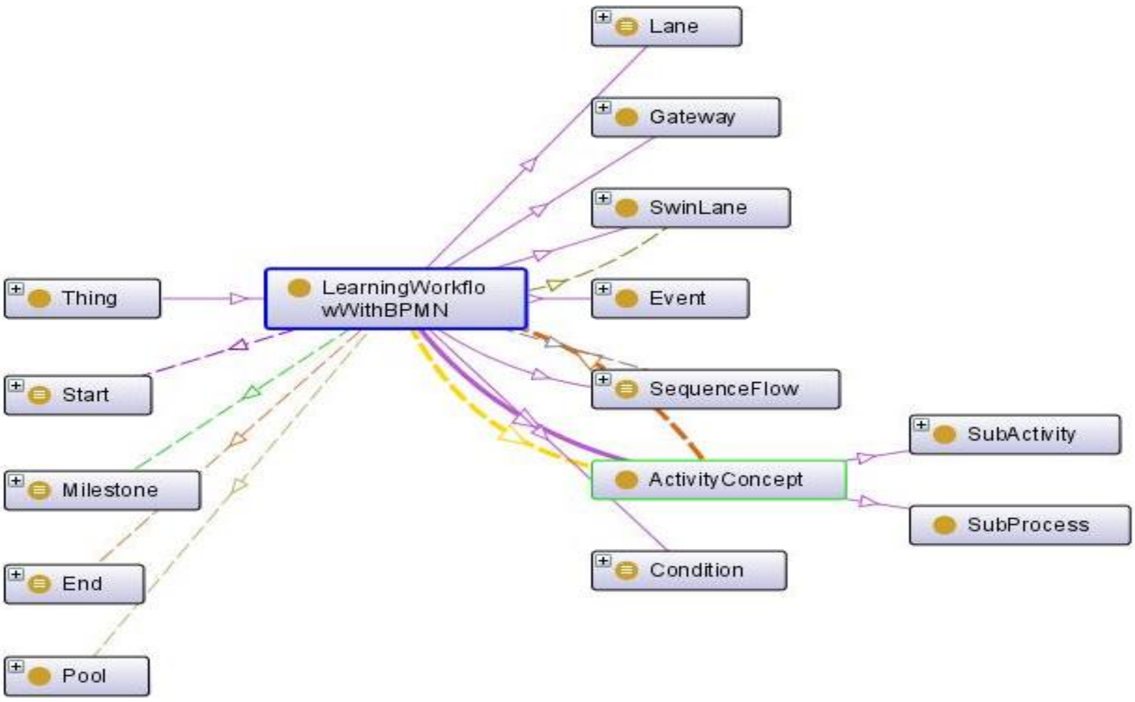


Figure 13. Learning Workflow Ontology Graph.

In Figure 14, we show how we map the sequence flow of activities from the Start of the Research Process to the End of the Research Process using the *hasPredecessor* and *hasSuccessor* object property assertion. The purpose is to make connections between the different BPMN concepts using the Object Property Assertion.

In Figure 15, we use the *EndDefineTopic* to show how we map and represent the different Milestones within the Learning model. It describes that the EndDefineTopic is an End activity type for the *DefineTopicArea* (1st Milestone) but also hasSuccessor *StartReviewLitProcess* which is the Start of the 2nd Milestone *ReviewLiterature.*

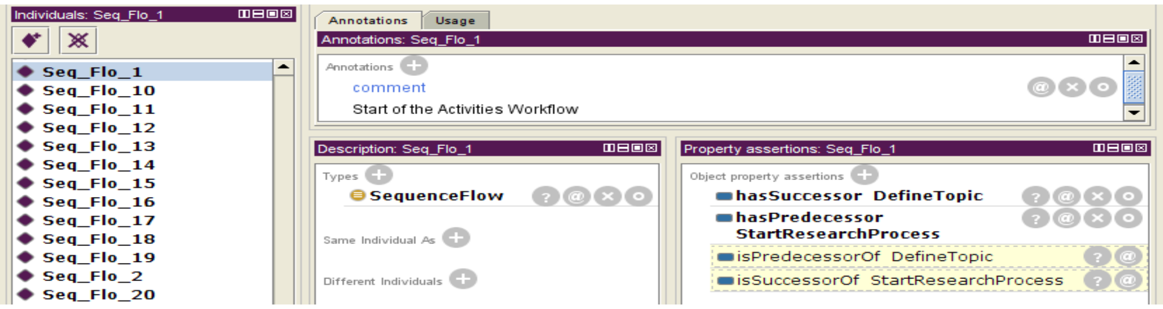


Figure 14. Sequence Workflow Mapping.

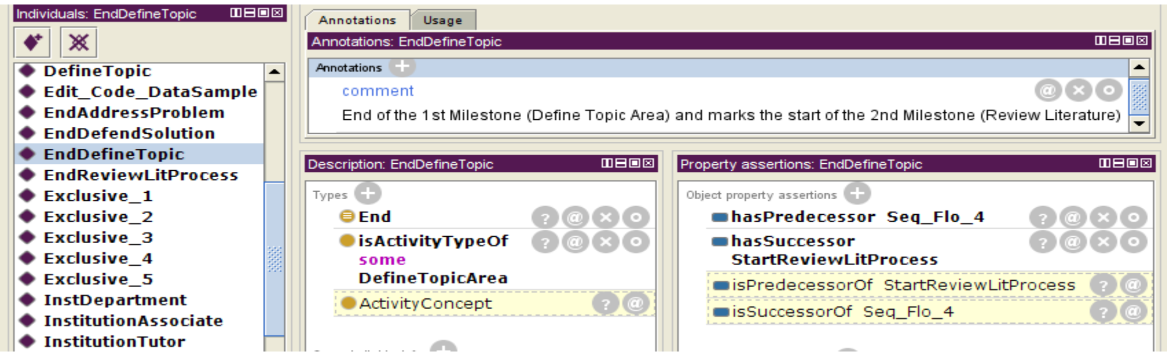


Figure 15. Milestone description and Mapping

B. Discovery and Enhancement of Patterns within the Learning Model Using Semantic Rules and Description Logic Queries

This section of the paper shows how we utilize ontology schema, process descriptions and web rule languages such as Semantic Web Rule Language (SWRL) [21] to discover sets of relationships that can be found within the learning process knowledge base. This results in suitable learning patterns being determined by means of semantic reasoning, which is then used to address the problem of extraction and association of useful patterns from captured learning data source to provision of knowledge. We further show the process model and automated discovery of learning patterns using Description Logic (DL) [22] queries and then utilize the main function offered by the Pellet Reasoner to help in checking for consistency in the model; to test whether or not a class is a subclass of another class, or checking whether or not it is possible for a class to have any instances. This means a class is said to be inconsistent if it does not have any instances. To this end, our approach reveals that ontology concepts and semantic rules can be layered on top of existing information asset to provide a more conceptual analysis of real time processes capable of providing real world answers that are closer to human understanding [1][12].

In Figure 16, we create SWRL Rules to associate existing domain classes with the right concepts in order to automatically infer the whole Ontology.

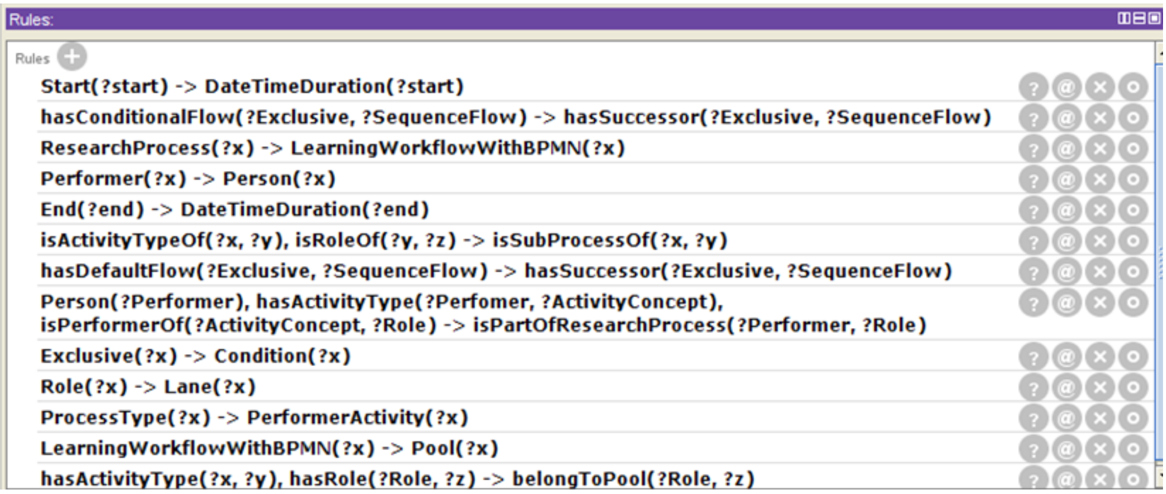


Figure 16. Semantic Web Rule Language (SWRL) in Prot ég éEditor

Furthermore, we provide some of the definition and functionality of the Rules that are implemented in our Semantic Model:

1. Person (?Performer), hasActivityType (?Perfomer, ?ActivityConcept), isPerformerOf (?ActivityConcept, ?Role) -> isPartOfResearch Process

(?Performer, ?Role)

= this Rule describes that any person that performs a Learning Activity classified as a Role is then automatically part of the Research Process.

2. hasDefaultFlow (?Exclusive, ?SequenceFlow) -> hasSuccessor (?Exclusive, ?Sequence Flow)

= describes that if we have a Default flow for an Exclusive gateway then this flow is also a Successor i.e., If X hasDefaultFlows Y then Y is DefaultFlowOf s X as shown in Figure 17.

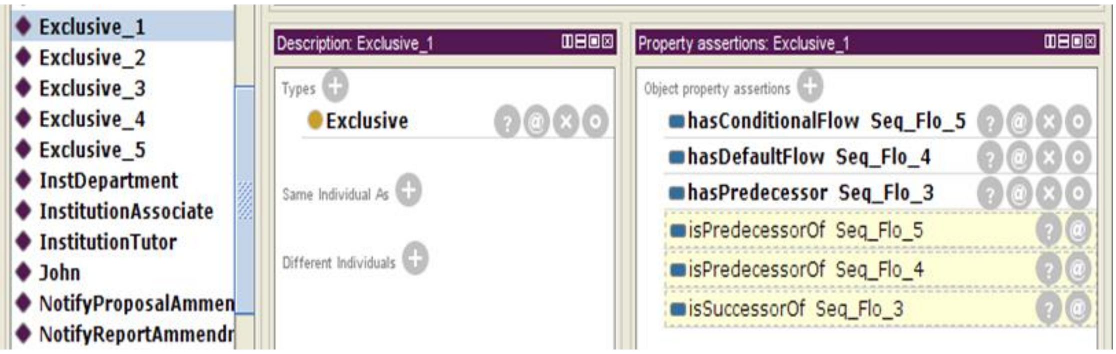


Figure 17. Running example of hasDefaultFlow exclusive gateway.

3. Research Process (?x) -> Learning Workflow With BPMN (?x)

= describes that if we have a research process then it is automatically a Learning Workflow

4. LearningWorkflowWithBPMN (?x) -> Pool (?x)

= describes that if we have a Learning Workflow then it is also a Pool

5. Role (?x) -> Lane (?x)

= describes that if we have a Role then it is also a Lane

6. hasActivityType (?x, ?ActivityConcept), hasRole (?ActivityConcept, ?Role) -> belongToPool (?x, ?Role)

= describes that if we have a learning activity which is performed under a particular Role, then this activity belong to the pool of that Role. Role has also been described as a Lane.

C. Querying of Learning Parameters and Real time Reasoning of Process Using Description Logic (DL) Queries

Description Logic (DL) Queries is a process description syntax which can be used to check for consistency for all defined entities within the learning model. We use the query to compute the inferred classes and individual assertions in order to check that all parameters/entities within the discovered classes are true and at least falls within the universal restriction of validity by definition, and that there are no inconsistency of data or repeatable contradicting discovery.

We provide the following queries to explain how we utilize the process description format.

1. Is DefineTopic an Activity of the first Milestone (DefineTopicArea)?

DL Query: ActivityConcept and is ActivityType Of some DefineTopicArea

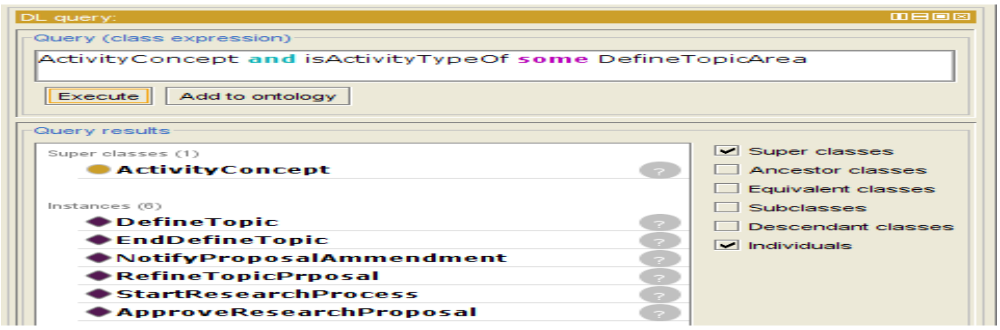


Figure 18. DL Query for Example 1.

= the query executed in Figure 18 checks if the activity of the first Milestone equal to Define Topic, thus Compare the activity of the first Milestone DefineTopicArea with Activity Concept (DefineTopic)

2. Is the Last Activity of the Research Process Award Certificate?

DL Query: i. ResearchProcess and hasEnd value AwardCertificate

ii. ActivityConcept and isEndOf some ResearchProcess

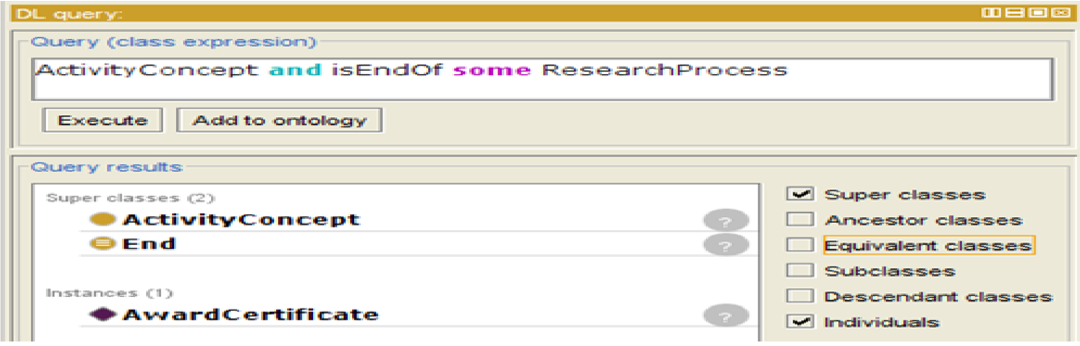


Figure 19. DL Query for Example 2.

= the query executed in Figure 19 checks the last Milestone in the research process and compares if the last activity is equal to Award Certificate. Hence compares the activity of the last Milestone DefendSolution with AwardCertificate

3. What is the Start Activity of the second Milestone Review Literature?

DL Query: ActivityConcept and isStartProcessOf some ReviewLiterature

= computes and checks the start event of the second Milestone ReviewLiterature, thus compare the activity of the second milestone with the result StartReviewLitProcess. Hence, Every Review Literature hasStartOfProcess StartReviewLitProcess.

4. Is CollectData an Activity of the Third Milestone Address Problem?

DL Query: ActivityConcept and isActivityTypeOf some AddressProblem

= checks and computes the activities of the Third Milestone AddressProblem, thus compare if the result is equal to the Activity Concept CollectData

5. Does Person P Activity A?

Example: *Does Person (Richard) Activity Approve Research Proposal?*

DL Query: Person and hasActivityType value ApproveResearchProposal

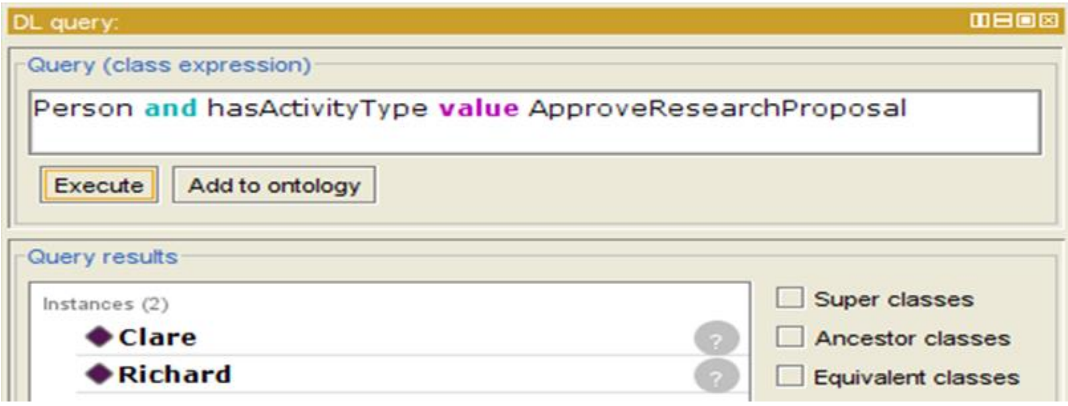


Figure 20. DL Query for Example 5

= the query in Figure 20 computes and check persons associated with the Approve Research Proposal and then compares if person (Richard) does the activity ApproveResearchProposal.

6. Does person P activity of activity A and B?

Example: *Which Persons does Activity RecheckSamplePlan and ReWriteReport?*

DL Query: Person and hasActivityType some {RecheckSamplePlan, ReWriteReport}

= computes and check which persons in the model does activity RecheckSamplePlan and ReWriteReport.

7. Does Person P activity A and then B and then C?

Example: *Does person Paul activity of type CollectData and then Edit_Code_Data Sample and then Analyse_Process_Data Sample?*

DL Query: Person and hasActivityType some {CollectData, Edit_Code_Data Sample, Analyse_Process_Data Sample}

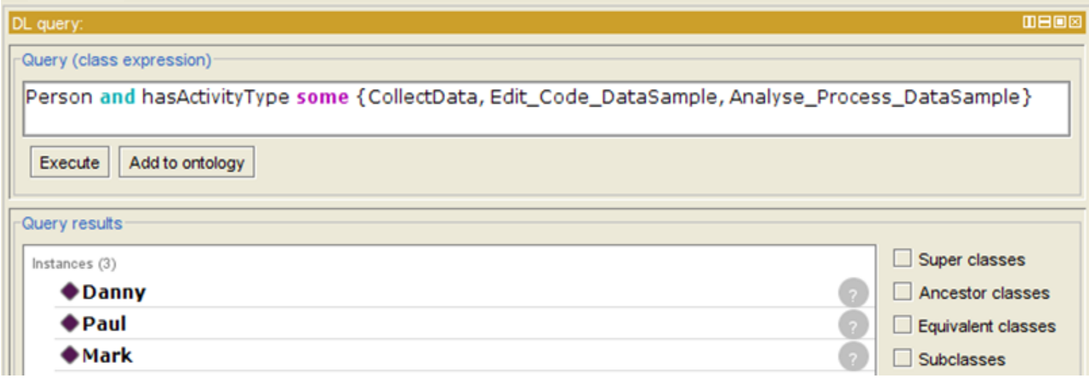


Figure 21. DL Query for Example 7.

= the query executed in Figure 21 computes and check if person Paul does the activity {Collect Data, Edit_Code_Data Sample, Analyse_Process_Data Sample}

8. Does Person P have Activity at least value of 3?

Example: *Does Person (Danny) Activity at least three?*

DL Query: Person and hasActivity Type min 3

= computes the Persons in the Model with a minimum of three Activities and compare if the result is equal to Person Danny

9. Who performs Learning Task T?

Example: *What are the different category of performers of a Learning Task in the Model?*

DL Query: Performer and isPerformerOf some ActivityConcept

Or simply execute *Role* because of the SWRL description *Role (?x) -> Lane (?x)* which describes that if we have a Role then it is also a Lane.

= the execution of the query in Figure 22 computes and checks for various category of Persons in the Model that performs a Learning task. Performer has been described also as a Person by the SWRL Rule: *Performer (?x) -> Person (?x)*

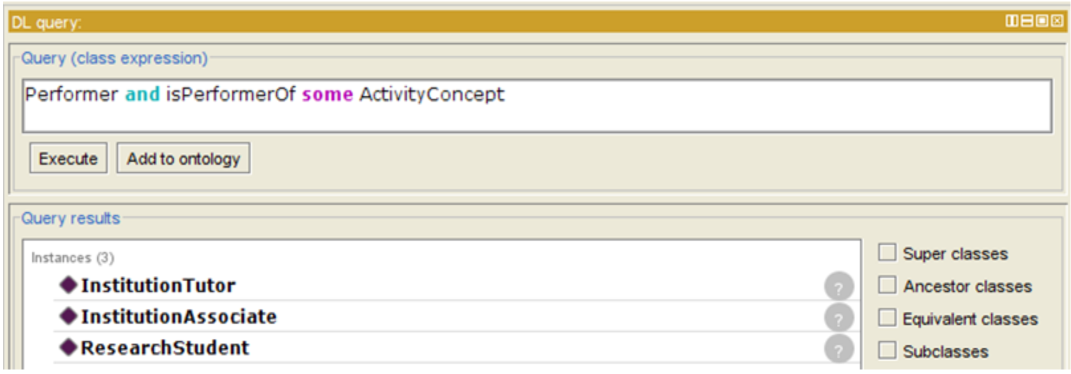


Figure 22. DL Query for Example 9

10. Does Person P perform Learning Task T?

Example: *Which Persons Performs a role as Institution Tutor?*

DL Query: Person and hasRole value InstitutionTutor

= computes the persons in the model that has role as an Institution Tutor

11. Does person P the first Milestone?

Example: *Does person Clare the first Milestone (Define Topic Area)?*

DL Query: Person and hasActivityType some DefineTopicArea
= compares the Persons of the First Milestone DefineTopicArea with Clare i.e., checks if the persons of the first Milestone equals Clare thus, if an activity of the first Milestone is done by person Clare.

12. Does Person P the second Milestone?

Example: *Does person Ben the Second Milestone (Review Literature)?*

DL Query: Person and hasActivityType some ReviewLiterature

= checks if the persons of the second Milestone equals Ben, i.e., compares the Persons of the second Milestone ReviewLiterature with Ben, thus if an activity of the second Milestone (Review Literature) is done by person Ben.
13. Does person P the Third Milestone?

Example: *Does Paul the Third Milestone (Address Problem)?*

DL Query: Person and hasActivityType some AddressProblem

= compares the Persons of the Third Milestone with Paul i.e., Checks if the persons of the Third Milestone equals Paul, thus if an activity of the Third Milestone DefendSolution is done by person Paul?

14. Does person P the Last Milestone?

Example: *Does person Danny the Last Milestone (Defend Solution)?*

DL Query: Person and hasActivityType some DefendSolution

= computes and check if the result of Persons in the Last Milestone DefendSolution is equal to person Danny i.e., compares the Persons of the Last Milestone with Danny, thus if an activity of the Last Milestone (Defend Solution) done by person Danny?

15. For all Activities always Event E implies eventually Event F?

Example: *For all Activities always Event (End) implies eventually Event (Start)*

DL Query: Event and hasSuccessor some Start

= describes and computes that - Hold for all activities that if event End occurs, then eventually event Start occurs too. We use this to define the *Start* and *End* of each Milestone in the Model e.g., we define that = Every DefineTopicArea is a ResearchProcess that isKindOf a Milestone and hasEndOfProcess EndDefineTopic and hasStartOfProcess StartResearchProcess.

We can then ask: Does the *End* of a Milestone eventually means the *Start* of the next Milestone too?

Question 16 below answers this…

16. Eventually Event E and then F?

Example: *Eventually EndDefineTopic and then StartReviewLitProcess?*

DL Query: Event and hasSuccessor value StartReviewLitProcess

= checks and compares that the End of the DefineTopicArea during the research process means the Start of the next milestone ReviewLiterature.

17. Finally Person P?

Example: *List all the Persons that performs an Activity in the Research Process?*
DL Query: Person and hasActivityType some ResearchProcess

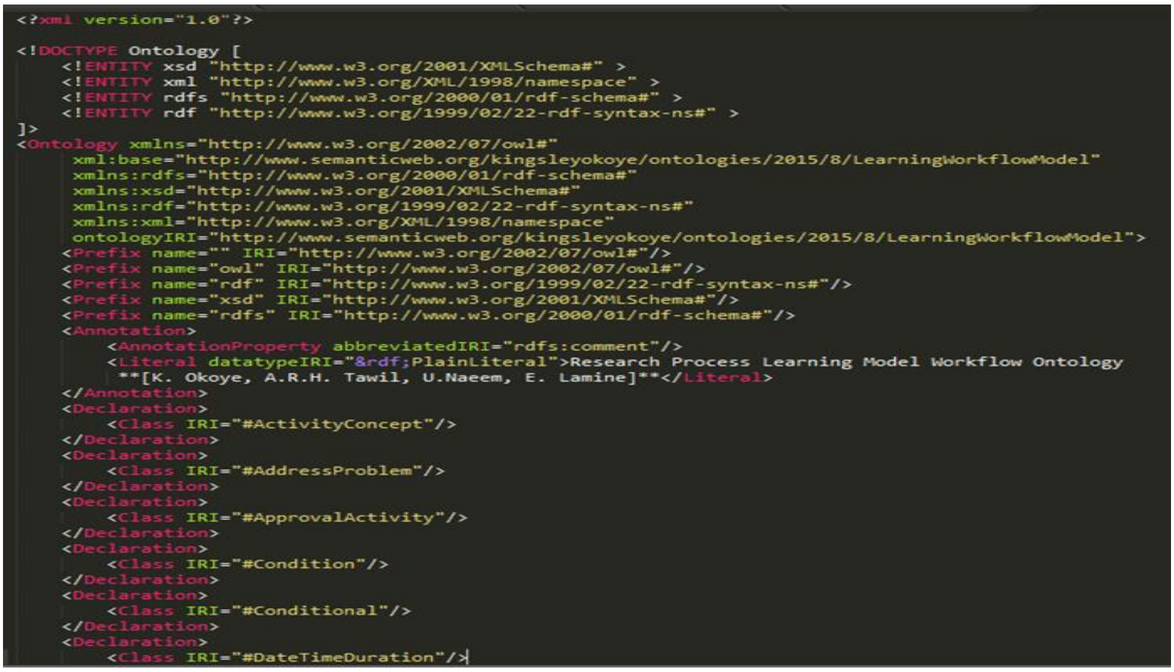
= computes any Person P that is a performer of a Learning Task in the Model. This has been described via the SWRL rule:

Person (?Performer), hasActivityType (?Perfomer, ?ActivityConcept), isPerformerOf (?ActivityConcept, ?Role) -> isPartOfResearch Process (?Performer, ?Role)

= the SWRL rule describes that any person that performs a Learning Activity is then automatically part of the Research Process.

From all definitions in the Learning Model and the extracted OWL XML format - Figure 23, we reveal that the application of semantic reasoning and process descriptions allows the extraction and conversion of explicit information into some implicit information by defining relationships/role assertions and deducing inferences based on such rule-base design. In Learning Process models, these metrics can be used to dramatically reduce the exploration or drilling down space when constructing the set of learning activities.

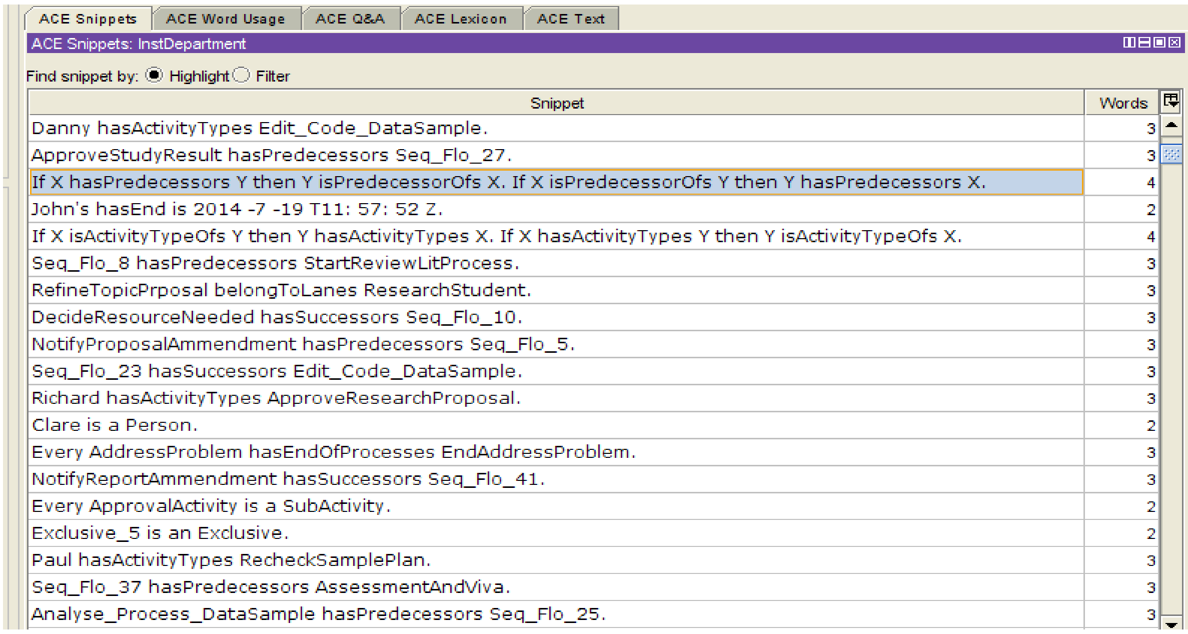
Figure 24 - shows a descriptive declaration of some of the relationships that has been computed and inferred in our semantic learning model. This form of expression as shown in Figure 24 has been used to provide process specification and expressive language formats that are logical and fundamental to knowledge representation such as the Knowledge Interchange Format (KIF) [23] which makes it possible to understand the meaning of logic expressions through Declarative Semantics. Designers of knowledge base systems can use this type of rule expressions to help identify new opportunities especially for enhancement of process models. The association strategies reveals interesting connection among domain entities, the individual classes and object/data types to provide a better understanding of how the different elements within the Learning Process Knowledge base relate and interact with each other.



```xml
<?xml version="1.0"?>

<!DOCTYPE Ontology [
    <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
    <!ENTITY xml "http://www.w3.org/XML/1998/namespace" >
    <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
    <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
]>
<Ontology xmlns="http://www.w3.org/2002/07/owl#"
     xml:base="http://www.semanticweb.org/kingsleyokoye/ontologies/2015/8/LearningWorkflowModel"
     xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
     xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
     xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
     xmlns:xml="http://www.w3.org/XML/1998/namespace"
     ontologyIRI="http://www.semanticweb.org/kingsleyokoye/ontologies/2015/8/LearningWorkflowModel">
    <Prefix name="" IRI="http://www.w3.org/2002/07/owl#"/>
    <Prefix name="owl" IRI="http://www.w3.org/2002/07/owl#"/>
    <Prefix name="rdf" IRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#"/>
    <Prefix name="xsd" IRI="http://www.w3.org/2001/XMLSchema#"/>
    <Prefix name="rdfs" IRI="http://www.w3.org/2000/01/rdf-schema#"/>
    <Annotation>
        <AnnotationProperty abbreviatedIRI="rdfs:comment"/>
        <Literal datatypeIRI="&rdf;PlainLiteral">Research Process Learning Model Workflow Ontology
        **[K. Okoye, A.R.H. Tawil, U.Naeem, E. Lamine]**</Literal>
    </Annotation>
    <Declaration>
        <Class IRI="#ActivityConcept"/>
    </Declaration>
    <Declaration>
        <Class IRI="#AddressProblem"/>
    </Declaration>
    <Declaration>
        <Class IRI="#ApprovalActivity"/>
    </Declaration>
    <Declaration>
        <Class IRI="#Condition"/>
    </Declaration>
    <Declaration>
        <Class IRI="#Conditional"/>
    </Declaration>
    <Declaration>
        <Class IRI="#DateTimeDuration"/>
```

Figure 23. Fragment of the OWL XML file format of our Semantic Learning Model.



Snippet	Words
Danny hasActivityTypes Edit_Code_DataSample.	3
ApproveStudyResult hasPredecessors Seq_Flo_27.	3
If X hasPredecessors Y then Y isPredecessorOfs X. If X isPredecessorOfs Y then Y hasPredecessors X.	4
John's hasEnd is 2014 -7 -19 T11: 57: 52 Z.	2
If X isActivityTypeOfs Y then Y hasActivityTypes X. If X hasActivityTypes Y then Y isActivityTypeOfs X.	4
Seq_Flo_8 hasPredecessors StartReviewLitProcess.	3
RefineTopicPrposal belongToLanes ResearchStudent.	3
DecideResourceNeeded hasSuccessors Seq_Flo_10.	3
NotifyProposalAmmendment hasPredecessors Seq_Flo_5.	3
Seq_Flo_23 hasSuccessors Edit_Code_DataSample.	3
Richard hasActivityTypes ApproveResearchProposal.	3
Clare is a Person.	2
Every AddressProblem hasEndOfProcesses EndAddressProblem.	3
NotifyReportAmmendment hasSuccessors Seq_Flo_41.	3
Every ApprovalActivity is a SubActivity.	2
Exclusive_5 is an Exclusive.	2
Paul hasActivityTypes RecheckSamplePlan.	3
Seq_Flo_37 hasPredecessors AssessmentAndViva.	3
Analyse_Process_DataSample hasPredecessors Seq_Flo_25.	3

Figure 24. ACE Snippet view - Description of Entities Relationships in the Learning Model.

VI. SLPM – Automated Learning Algorithm.

In this section, we present the final approach for the semantic learning process mining algorithm. We implement the automated learning application (SLPM) using Eclipse Java Runtime Environment [24] to create the methods and build the interface for loading the Ontology and Process Parameters. We use Ontology Web Language Application Programming Interface (OWL API) [25] to extract and load the inferred ontology of the learning model. The purpose is to match the questions one would like to answer about relationships the process instances share amongst themselves by linking to Concepts within the Learning Ontology. The purpose of the application in Figure 25 is to allow us to perform a more conceptual analysis capable of providing real world answers that are closer to human understanding by querying the learning model based on concepts that they represent in real time environment by linking them to our deployed ontology. By pointing to references in the ontology and process parameters, it becomes easy to refer to a particular case or events within the Learning Model.

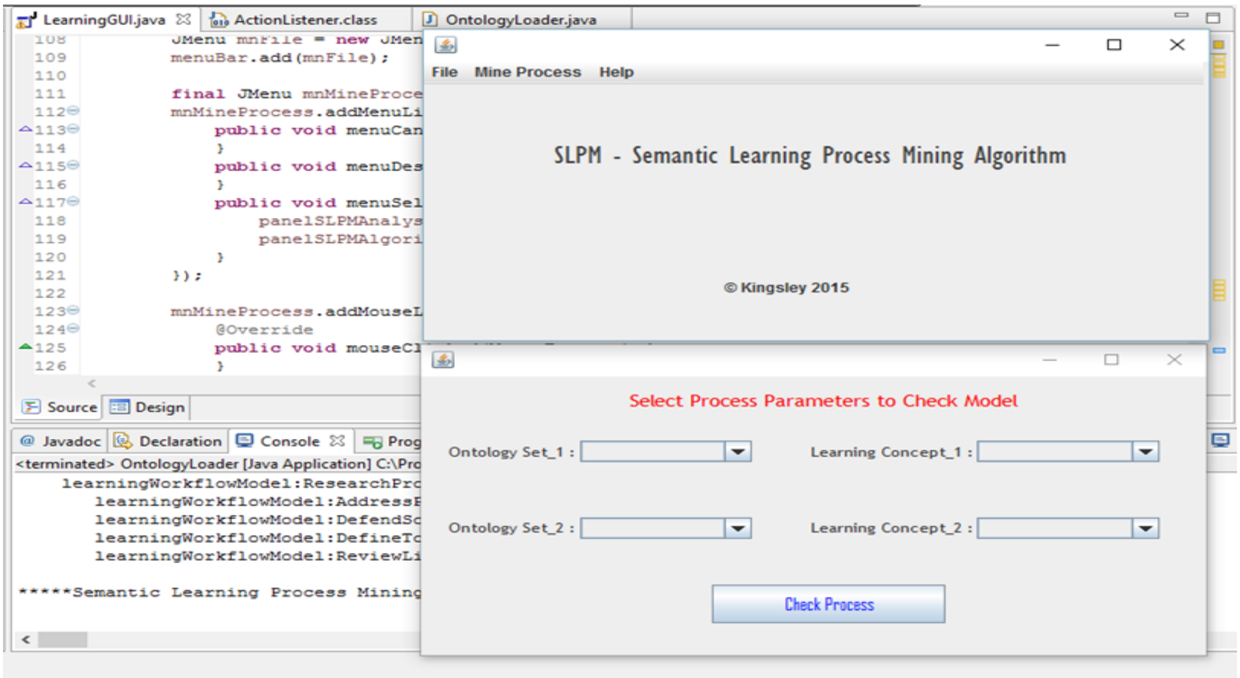


Figure 25. SLPM Interface in Java runtime environment.

The drive for such our approach is that by pointing to references in the deployed ontology and application of the programming interface and Semantic Reasoning, it becomes easy to refer to a particular case or event. This is a useful technique especially in solving some didactic issues and answering some questions with regards to different Learners behavior. Questions like “What attribute or paths do successful learners have in common” or “What attributes distinguishes such successful learners from the unsuccessful ones” can be established. By answering such questions, it can be seen that the design of automated learning approaches must not only focus on the paths/attributes for learners who completes the learning process, but should also anticipate and predict the problems encountered by other unsuccessful ones. The purpose is to match the questions one would like to answer as we revealed in this paper and the ability to identify bottlenecks and monitor deviations within a learning execution environment.

VII. Related Works

Process workflow description and management have been applied in many enterprise information systems [26][27][28] such as Staffware, IBM, MQSeries, and COSA, which offer generic modelling and enactment capabilities for structured processes. Many other software systems have adopted workflow technology, for example ERP (Enterprise Resource Planning) systems such as SAP, PeopleSoft, Baan and Oracle, CRM (Customer Relationship Management) software. However, despite its advantages, many problems are still being encountered when applying workflow technology. One of the problems is that these systems require a workflow design [26]. This means that a designer has to construct a detailed model accurately describing the routing of work, which most often requires deep knowledge of the workflow system and management involved. Secondly, creating a workflow is a complicated time-consuming process and typically there are discrepancies between the actual workflow process and the actual processes as perceived by the management. Nooijen et al. [29] notes that to fill in these gaps, event type specification needs to be utilized in order to construct database queries that extracts attributes from all event logs, groups them into cases, orders them by time stamps, and then writes the result into a classical logs in separate database columns.

According to Verbeek et al. [13] the most recent generic approach to event log extraction is XESame which manually defines mapping between source data and event logs, sorts them into traces and then translates their mappings to SQL queries which are subsequently stored in a database. Stored data can be queried to retrieve the events of the logs from central process data. Many approaches have been tested to extract event logs from ERP (Enterprise Resource Planning) systems such as SAP [30] and People Soft [31]. Consequently, as ERP systems in general provide multiple case identifiers, the majority of these approaches failed. The authors in [29] argue that success could only be reported when database tables are carefully selected by hand or a better view of data is provided using means like Semantic Annotation.

A number of works has been directed towards using the techniques of process mining for Semantic Process Analysis [1][2][7]. These works shows that ontological modelling and reasoning are the essential building tools for semantic process mining. Amongst the existing methods used is the *Alpha Algorithm* introduced by van der Aalst et al. in [4]. The authors used the algorithm to extract process model from event logs and has been proven to supports both semantic and non-semantic process data. In this paper, we mine activities logs within a learning knowledge-base to determine the association of elements in the logs by referencing concepts in an ontology specifically designed for representing learning process.

Rozinat and van der Aalst [3] presented the *Decision Miner* used for decision point analysis in discovered process 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 decision making can be improved by performing inferences over a knowledge-base of learning process to discover and establish new knowledge by means of semantic reasoning.

Pedrinaci and Domingue [32] argue that MXML is not all that is needed or prerequisite for semantic process mining. According to the authors, MXML log is only able to refer to an identity tag for a particular entity. The actual semantics which describes the object types and the relationships they share within the process model are not readily available. In essence, MXML suffers from a lack of machine processable semantic, even though it may be possible to create means of retrieving knowledge or information. This means that modelling and analyzing of levels of data and concepts needed for process mining, requires technologies capable of recognizing relationships and relating across the knowledge-base.

Many other semantic log file formats has been established which supports MXML. The SA-MXML [7] supports semantic annotation of elements in event logs by linking terms to concepts in ontologies [7]. According to [2] the supporting tool to generate SA-MXML files is in ProMImport [17] which serves log files from Process-Aware Information Systems (PAIS) under SUPER. XES [13] has also been introduced to address the problem of semantically adding attributes and definition of different concepts, even though the authors in [2] mentions that most of the supporting algorithms are still under development.

Most developed systems use various mining techniques for representation of concepts, knowledge or data which are focused on applying technologies to different aspects of processes [33][34]. Nevertheless, the application of Semantic Reasoning can help solve the problem of regulating evolving and static methods for representing knowledge at theoretical and technological levels by making inferences [8][35], retaining and applying what have been learned [29], and discovery and enhancement of new processes [4]. In this paper, we adopt Semantic Process Mining to represent Learning processes. Our 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 a learning knowledge-base by automatically constructing process models capable of defining, classifying and enhancing observed learning behaviours. In general, these method assume that there already exist a probabilistic or fuzzy knowledge-base for learning, upon which this methods are able to predict patterns/behaviour of new but not previously observed event/data types within the process.

d'Amato et al. [36] notes that various methods have been proposed in literature which are directed towards obtaining a more expressive model from knowledge bases [37]. The authors [36] 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 using Description Logics particularly those based on classification, clustering and ranking of individuals. Learning Process modelling has been tackled over the years by customising Machine Learning methods such as Instance Based Learning [38] and Support Vector Machine (SVM) [39] to Description Logics (DLs) [22] queries; which is the standard theoretical foundation upon which semantic web languages such as OWL and SWRL are based.

Reasoning on ontological knowledge plays an important role in the semantic representation of processes such as 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. Thom et al. [40] describes Workflow Activity Patterns (WAPS) as common structures involving the interaction between individual entities and the control-flow constructs used to model the semantics of activities as they are being performed. Workflow systems assume that a process can be divided into small, unitary actions, called Activities [41]. To perform a given process, one must per-form the set (or perhaps a subset) of the activities that comprise it. Hence, an Activity is an Action that is a semantic unit at some level, which can be thought of as a function that modifies the state of the process in terms of the semantics of the patterns and can be discovered automatically by means of semantic reasoning [8][12].

According to [42] and [43] Bayesian models have paved way for new machine-learning algorithms with more powerful and more human-like capabilities. Semantic web ontology and its application cannot be explained without mentioning the Bayesian theory of probability [44][45]. The Bayesian probabilistic theory have been proven to be one of the few mathematical interpretation of predictive concepts for representing a state of knowledge, thus, an extension of logic proposals that enables reasoning with hypothesis whose true or false values is uncertain. Bayesian model is based on 3 vital probes: What are the content of probabilistic theories? How can they be used to support reasoning? And how can they themselves be reasoned upon? The hypotheses are measured by computing the Bayes' rule, where: Probability, $P(x\backslash h, T)$ measures how well each argument predicts the data and the initial marking or likelihood. $P(h\backslash T)$ expresses the plausibility of the hypotheses given the users background knowledge. The posterior probability, $P(h\backslash x, T)$, is proportional to the result of the two expressions representing the level of certainty in each of the hypotheses given both the constraints of the background theory T, and observed data x. According to Tenenbaum et al [46], the challenge comes in specifying hypotheses and probability distributions that support Bayesian inference for a given task/domain. The authors argue that both structured knowledge and statistical inference are necessary to explain the nature, use and acquisition of such human knowledge and further introduced a theory-based Bayesian framework for modelling inductive learning and reasoning.

According to [47] 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 [8] 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 [8] we show that this rule is similar and can be related to the Semantic Web Rule Language, SWRL [21] which is a useful language designed for process description especially to provide an improved learning ontology and enhancement of the learning process model. The SWRL rule has the form; *atom* ^ *atom* (antecedent)... → *atom* ^ *atom* (consequent). Association rule learning strongly supports the use of such metrics frequently expressed in the 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.

Elhebir and Abraham [48] 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. The authors [48] 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 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 i.e, classified. [36][8][12].

VIII. Conclusion and Future Work

The work in this paper focus on identifying and modelling of different event data about Learning Process. The goal is to enrich the information values of the resulting model based on semantic process mining and analysis. As a result, suitable process statistics were determined which is then used to address the problem of analyzing concepts and relationships amongst learning objects, to discover new and enhance existing learning processes. We use semantic process mining to perform the discovery of learning patterns based on semantically annotated event logs extracted from the learning execution environment, and submitting the resulting extensible event stream file formats for further process analysis. We show that the development of Semantic Learning Process Mining approach entails three building blocks; Annotated Event Logs, Ontologies and Semantic Reasoning that aim at discovering, conformance and extension of Learning processes. Our approach makes use of Semantic Annotations to link elements in event log of a learning process with concepts that they represent in an Ontology. By referring to ontologies, the approach provides us with the capability to determine the relationships the process instances share within the knowledge-base and then infer and discover learning patterns automatically by means of Semantic Reasoning. We prove that learning patterns or behaviours can be discovered as a consequence or condition of a Rule. Rather than displacement of prior learning knowledge, the approach provides us with benefits in discovery, flexible access and information integration due to the inherent connectedness (inference), concept matching and reasoning.

Future work could focus on extending the approach described in this paper by applying the technique to a different process domain. This will help in analyzing the streams of events logs that are involved in the process in order to produce inference knowledge, which can then be used to load a more enhanced model within the process domain area.

Acknowledgment

The authors are most grateful to the anonymous reviewers for their valuable comments.

References

[1] A. K. A. de Medeiros and W. M. P. Van der Aaslt, “Process Mining towards Semantics.” Advances in Web

Semantics. Lecture Notes in Computer Science Volume 4891, pp. 35-80, 2009.

[2] W. Jareevongpiboon and P. Janecek, "Ontological approach to enhance results of business process mining and analysis." Journal of Business Process Managt, Vol. 19(3) pp. 459-476, 2013.

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

[4] W. M. P. Van der Aalst,"Process Mining: Discovery, Conformance and Enhancement of Business Processes." Springer, 2011.

[5] H.P. Kriegel, K.M. Borgwardt, P. Kröger, A. Pryakhin, M. Schuber and A. Zimek "Future trends in data mining". Journal of Data Min. Knowl. Discov. Vol. 15, pp. 87-97, 2007.

[6] 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" In: 3rd International IFIP Workshop on Semantic Web & Web Semantics (SWWS '07), 29-30 Nov 2007, Vilamoura, Algarve, Portugal, pp. 1244–1255, 2007.

[7] 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. pg. 1953-1964, 2008.

[8] 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.

[9] C. Carlsson, J. Mezei and M. Brunelli "Fuzzy Ontology Used for Knowledge Mobilization" Int. J. Intell. Syst., 28: 52–71. doi:10.1002/int.21574, 2013.

[10] W. M. P. Van der Aalst, M.H. Schonenberg and M. Song, "Time Prediction Based on Process Mining." Journal of Information Systems Vol. 36(2), pp. 450-475, 2011

[11] R. Kumar "Research Methodology: A Step-by-Step Guide for Beginners." In: SAGE, London, pp. 33-42, 2014.

[12] 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.

[13] H. M. W. Verbeek, J. C. A. M., Buijs, B. F. Dongen and 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.

[14] C. W. Gunther and W. M. P. Van der Aalst, "Fuzzy Mining: Adaptive process Simplification Based on Multi-Perspective metrics." International conference on Business Process Management (BPM 2007). Vol. 4714. pp. 328-343, 2007.

[15] Disco Users Guide. https://fluxicon.com/disco/files/Disco-User-Guide.pdf.

[16] ProM Tool http://www.processmining.org/prom/start.

[17] C. Gu¨nther and W. M. P. Van der Aalst, "A generic import framework for process event logs," In: Eder, J. and Dustdar, S. (Eds.), Business Process Management Workshops, Springer, Berlin, pp. 81-92, 2006.

[18] G. Acampora, V. Loia, S. Salerno and A. Vitiello, "A hybrid evolutionary approach for solving the ontology alignment problem". Int. J. Intell. Syst., 27: 189–216. doi: 10.1002/int.20517, 2012.

[19] Protégé Editor. http://protege.cim3.net Accessed [June 14th 2015].

[20] T. Martin and A. Majidian, "Finding Fuzzy Concepts for Creative Knowledge Discovery". Int. J. Intell. Syst., 28: 93–114. doi: 10.1002/int.21576, 2013.

[21] SWRL: A Semantic Web Rule Lang http://www.w3.org/Submission/2004/SUBM-SWRL-20040521/.

[22] F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi and P.F. Patel-Schneider, editors. "Description Logic Handbook". Cambridge University Press, 2003.

[23] Knowledge Interchange Format. http://cl.tamu.edu/discuss/kif-100101.pdf Accessed [Sept. 2015].

[24] Eclipse Java Runtime Environment Open Source, https://eclipse.org/downloads/ (Accessed August, 2015)

[25] Ontology Web Language Application Programming Interface - OWL API http://protegewiki.stanford.edu/wiki/ProtegeOWL_API_Programmers_Guide

[26] W. M. P. Van der Aalst, A. J. M. M. Weijters and L. Maruster, "Workflow Mining: Discovering Process Models from Event Logs." Int. Journal of IEEE transactions on Knowledge and Data engineering, Vol. 16(9). Pg. 1128-1142, 2004.

[27] W. M. P. Van der Aalst and K. M. Van Hee, "Workflow management: Models, Methods, and Systems." MIT Press, Cambridge, MA, 2004.

[28] S. Jablonski and C. Bussler, "Workflow management: Modelling Concepts, Architecture, and Implementation." In: Thompson Computer Press, London, 1996.

[29] E. H. J. Nooijen, B. F. Van Dongen and D. Fahland (2013) "Automatic Discovery of Data-Centric and Artifact-Centric Processes." BPM Workshops. Lecture Notes in Business Info. Processing Vol. 132, pp. 316-327, 2013.

[30] D. Piessens, "Event Log Extraction from SAP." ECC 6.0, 2011.

[31] A. Ramesh, "Process Mining." In: People Soft, 2006.

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

[33] 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.

[34] IEEE Task Force on Process Mining, "Process Mining Manifesto." BPM Workshops LNBIP, vol. 99, pp. 169-194. Springer, 2012.

[35] K. Okoye, A.R.H. Tawil, U. Naeem, R. Bashroush and E. Lamine, "A Semantic Rule-Based Approach Towards Process Mining for Personalised Adaptive Learning," In:

16th IEEE Intl. Conf. on High Performance Computing and Comms (HPCC, CSS, ICESS) Paris, France, pp. 929-936, 2014.

[36] 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., ESWC2008. Vol. 5021 of LNCS, Spr 288-302, 2008.

[37] J. Lehmann and P. Hitzler, "Concept learning in description logics using refinement operators". Machine Learning, 78:203-250, 2010.

[38] T. Mitchell, editor. "Machine Learning". McGraw Hill, 1997.

[39] J. Shawe-Taylor and N. Cristianini, editors. "Support Vector Machines and other kernel-based learning methods". Cambridge Uni. Press, 2000.

[40] L.H. Thom, M. Reichert and C. Iochpe, "Activity patterns in process-aware information systems: basic concepts and empirical evidence". J. of Business Process Integration and Mangt, Vol. 4(2), pp. 93-110, 2009.

[41] D.L Ferreira and L.H. Thom, "A Semantic Approach to the Discovery of Workflow Activity Patterns in Event Logs". Journal of Business Process Integration and Management, Inderscience Publisher, Vol. 6(1), 2012.

[42] C. Kemp et al. "Semi-supervised learning with trees. Advances in Neural Info. Processing Syst. MIT Press, Vol.(16), pp. 257–264, 2004.

[43] T.L. Griffiths, N. Chater, C. Kemp, A. Perfors, and J.B. Tenenbaum "Probabilistic models of cognition: exploring representations & inductive biases". In: Trends in Cognitive Sciences, Vol. 14(8) pp. 357-364, 2010.

[44] S.B. McGrayne, "The Theory That Would Not Die", p. 10., at Google Books, 2011.

[45] S.E Fienberg, "When did Bayesian Inference become "Bayesian?" Bayesian Analysis, vol.1 (1), pp. 1–40, 2006.

[46] J.B. Tenenbaum, T.L Griffiths and C. Kemp "Theory-based Bayesian models of inductive learning and reasoning". In: TRENDS in Cognitive Sciences Vol.10 No.7 July, 2006.

[47] R.G.L. Miani and E.R.H. Junior, "Exploring Association Rules in a Large Growing Knowledge Base" *International Journal of Computer Information Systems and Industrial Management Applications.* ISSN 2150-7988 Volume 7 (2015) pp. 106–114, 2015.

[48] M.H.A. Elhebir and A. Abraham, "A Novel Ensemble Approach to Enhance the Performance of Web Server Logs Classification" *International Journal of Computer Information Systems and Industrial Management Applications* ISSN 2150-7988 Volume 7 (2015) pp. 189-195, 2015.

Author Biographies

Kingsley Okoye received his Bachelors degree in Computer Science from Nnamdi Azikiwe University, Nigeria in 2007, a Masters in Technology Management at the University of East London, UK in 2011. He is currently a PhD candidate in Software Engineering at the School of Architecture Computing and Engineering, University of East London. He is MIET member at the Institution of Engineering and Technology, UK. Graduate Member at the Institute of Electrical and Electronics Engineers, IEEE. His Research interests includes Learning Process Automation, Process Mining, Semantic Technologies, Process Modelling, Internet Applications and Ontology.

Abdel-Rahman H. Tawil is currently a Senior lecturer at the School of Architecture, Computing and Engineering, University of East London, UK. He received his BSc in Computer Science from the University of Jordan, Amman, Jordan, and his PhD in Semantic Interoperability in Heterogeneous and Distributed Multiple-Information Servers Environment from Cardiff University, Cardiff, UK. Before he joined UEL, he worked as a lecturer at the School of Computing Science and Mathematics, Aston University, Birmingham, UK. He is a member of the Distributed Software Engineering Research Group. Abdel-Rahman worked as a lecturer at the University of West of England (UWE), Bristol; University of Wales Institute Cardiff (UWIC); and the Department of Computer Science, Cardiff University. He is also a visiting lecturer at ISIS, Engineering School of Health Information Systems, Jean-François Champollion University, Toulouse, France. He is Fellow of the Higher Education Academy (HEA) and Member of the Research Committee at the University of East London. Abdel-Rahman has extensive research expertise and a wide range of research interests with special emphasis being placed on semantic technology and its use for real-world applications with focus on the development of semantic techniques for assistive technologies. His current research interest includes using wireless sensor networks in Tele-health and Tele-homecare applications: exploiting the semantics of resources and activities descriptions available in a patient surrounding environment, adaptive learning and care actors profiling.

Usman Naeem is currently a Senior Lecturer within the School of Architecture, Computing and Engineering, University of East London. He was awarded his PhD in July 2009 from Queen Mary University of London. He is a Fellow of the Higher Education Academy, UK. Member of IEEE and the IET. He has won a number of Awards, Fellowships, and Grants which includes the Annual Fund Start-Up Research Grant 2010/2011 (£2,000), UEL Early Career Researcher Accelerator Fund 2011/12 (£5,000) and UEL Researcher Accelerator Fund 2012/13 (£5,000). He is a Reviewer for a number of conference and journals such as ACM Computing Surveys, and IEEE Communications Magazine. He has Research Collaboration with a number of institutions such as the Cross-disciplinary collaboration with School of Health, Sports and Bioscience - UEL for the Project: Brain Training and Its Role in Robot-Assisted Neurorehabilitation - Development of cloud based system for analysing behaviour of stroke patients. He also has research collaboration with the Swinburne University, Malaysia and Greenwich University, UK. Usman's current research interests includes but not limited to the areas of Pervasive/Ubiquitous Computing, Machine Learning, Semantic Reasoning, Activity Recognition, Task Recognition, RFID Sensors and Networks, Object Usage, Activities of Daily Life (ADL) modelling, Context Awareness, Assisted Living, Smart Homes and Environments.. Much of his research is focused on the development of assistive technologies to support independent living for the elderly community. The results of this research have been published in a number of journals, conferences and workshops.

Elyes Lamine is an Associate Professor at the Centre universitaire de formation et de recherche Jean-François Champollion · Département Informatique et systèmes d'information pour la santé, France · Castres. He is also a Research Associate at the École des Mines d'Albi-Carmaux - The Industrial Engineering Center IO France Albi. His research is focused on engineering and agility management of collaborative business processes for inter-organizational systems run in dynamic and uncertain environments. Targeted divisions are mainly within the health sector: Drugs circuit. His Research interest includes Artificial Intelligence, Computing in Mathematics, Natural Science, Engineering and Medicine, Information Systems (Business Informatics), Information Retrieval, Knowledge Management, Medical Informatics, Semantic Web and Knowledge Representation.