UEL University of East London

The Graduate School

Semantic Process Mining Towards Discovery and Enhancement of Process Models and Event Logs Analysis: Application on Learning Process Domain

Doctoral Thesis Defence

by

Kingsley Okoye

Director of Studies: Dr. Syed Islam | Supervisor: Dr. Usman Naeem

Thesis Outline:

- **General Overview of the Research**
  - What…Why…How…?
  - Research Questions and Context
  - Research Contributions
- **Research Methodology**
- **Background Informations & Theory**
- **Proposed Methods and Design Framework**
  - 2-Dimensional Rhombus Approach
  - Architecture of the Semantic-based Approach and Algorithms.
  - Semantic Fuzzy Mining
- **Implementations and Experimentations**
- **Evaluation and Outcomes**
- **Summary**
- **Acknowledgements**

What the Research have done:

- The Research introduces a Semantic Fuzzy Mining approach that makes use of labels (i.e. concepts) within event logs about real time process to propose a method which allows for mining and improved analysis of the resulting process models through semantic - annotation, representation and reasoning.

Why…?

Syntactic vs Conceptual Model Analysis

- Most of the existing process mining techniques depend on tags or labels in event logs information about the processes they represent to discover process models.

Consequently, a common problem has been that majority of the existing techniques are to a certain extent limited or vague when confronted with unstructured data because they lack the abstraction level required from real world perspectives. This means that those techniques do not technically gain from the real knowledge (semantics) that describe the labels in the event logs of the domain processes.

Why… Contn’d

In principle, this research seek ways to prove how the analysis provided by the existing process mining techniques can be enhanced by adding semantic knowledge to the available event logs and the discovered process models.

How…?

- The research focus on extracting the streams of event logs from the real time processes and then propose algorithms, design frameworks and semantic-based formats that allows for mining and improved analysis of the captured datasets and the resulting process models.

How…? Contn’d

- **Qualitative Method of Analysis:**

  The study shows by using a case study of Learning Process - how the data from the various process domains can be extracted, semantically prepared, and transformed into mining executable formats to support the discovery, monitoring and enhancement of real-time processes through further semantic analysis of the discovered models.

How…? Contn’d

- **Quantitative Method of Analysis:**

  In addition, the research quantitatively assess the level of accuracy of the classification results of the proposed approach to predict behaviours of unobserved patterns or traces within the process knowledge-base.

How…? Contn’d

In summary, the research looks at:

- the level of impact and usefulness of the proposed semantic-based process mining approach
- validity of the classification results, and
- their influence compared to other existing benchmark algorithms and techniques for process mining.

Research Questions:

The following main research questions *RQ1* & *RQ2* forms the core validation study of the thesis and are addressed in Chapter 4 and 5. Primarily, the research explores the best possible ways towards the:

- ***RQ1:*** *Use of process mining techniques to discover, monitor and analyse event logs about some domain process by discovering useful and worthwhile process models? and*
- ***RQ2:*** *How effective semantic modelling and reasoning methods can be used to enhance process mining analysis from the syntactic level to a much more conceptual level?*

Research Questions Cont'nd…

Driven by such effort, the research in turn makes use of the case study of the learning process and data about a real-time business process to seek ways on *how* to do the following:

- ***RQ3*** *Extract data from process domains to show how we semantically synchronize the event log formats for various process domain data? (Chapter 4)*
- ***RQ4*** *Semantically prepare the data through an ontology driven search for explorative analysis of a learning process activities and executions? (Chapter 4)*

Research Questions Cont’nd…

- ***RQ5*** *Transform the extracted data into mining executable formats to support the discovery of valuable process models through our proposed technique for annotating unlabelled learning activity sequences using ontology schema/vocabularies?* (Chapter 4 and 5)

- ***RQ6*** *Provide techniques for accurate classification of unseen process instances (traces) within the process models, and useful strategies towards development of process mining algorithms that are more intelligent, predictive and robotically adaptive.* (Chapter 5)

Research Questions Cont’nd…

- ***RQ6*** *Monitor and enhance real-time processes through further semantic analysis of the discovered models*. (Chapter 5 and 6)
- ***RQ8*** *Importance of semantics process mining to augment information value of data about domain processes: case study of learning process.* (Chapter 6)
- ***RQ9*** *Application of process mining techniques to domain of learning process?* (the entire thesis)

Research Questions Cont’nd:

- ***RQ10*** *Provide real time semantic knowledge and understanding about domain processes (using the cases study of the learning process) that is useful towards the development of process mining algorithms that are more robust and intelligent with high level of effective conceptual reasoning capabilities?* (the entire thesis)

Main Components of the Research:

The main components and motive for implementing the proposed semantic-based process mining approach is summarised as follows:

- ***Event Logs:*** to show how process mining can be applied to improve the informative value of learning process data.
- ***Learning Model:*** describe how improved process models can be derived from the large volume of event data logs found within the learning process domain.

Main Components Cont’nd…

- ***Annotation:*** describe how semantic descriptions (annotation) of the deployed model can help enrich the result of the learning process mining and outcomes through discovering of new knowledge about the process elements.
- ***Ontology:*** use of ontologies with effective semantic reasoning to lift process mining analysis from the syntactic level to a more conceptual level.

Main Components Cont’nd…

- ***Semantic Learning Process Mining Algorithm (Semantic-Fuzzy Miner):*** reveals how references to ontologies and effective raising of process analysis from the syntactic to semantic level enables real time viewpoints on the learning process model - which in turn helps to address the problem of analyzing the learning process data based on concepts and to answer questions about relationships the learning objects (process instances) share amongst themselves within the knowledge-base.

Research Contributions:

The main contributions of the PhD are summarised as follows:

(1) Definition of a semantic-based fuzzy mining approach that exhibits a high level of semantic reasoning and capabilities.

(2) An algorithm that proves useful towards extraction, semantically preparation, and transformation of event log about any domain process.

(3) Design framework that highly influence and support the development of semantic process mining algorithms

(4) A process mining technique that is able to accurately classify and induce new knowledge based on previously unobserved behaviours.

Contributions Cont'nd…

(5) A method for formal structures on how to perform and present process mining results in a more intuitive and easy way.

(6) An ontology-based system that is able to perform information retrieval and query answering in a more efficient and effective way compared to other standard logical procedures.

(7) A series of case studies showing that semantic-based process mining can be used to enhance process mining results and analysis from the syntactic level to a much more conceptual level.

(8) Empirical evaluation of the impact of the Semantic Fuzzy mining approach and its outcomes compared to other benchmark algorithms for process mining.

Research Methodology:

The study makes use of both Qualitative and Quantitative research methods to carry out the investigations and proposals. In other words, the method is regarded as a fusion theory that is devoted to represent and analyse information in a qualitative and yet quantitative manner.

- In essence, the work utilizes both research methods for the purpose of validation and comparison by evaluating the level of impact and usefulness of the proposed approach and their influence compared to other existing benchmark algorithms and techniques that are closely related to the process mining field, using the case study of the Learning process and a training set and test log from a real time data about a business process for the cross-validation experiments.

Background Informations and Theory:

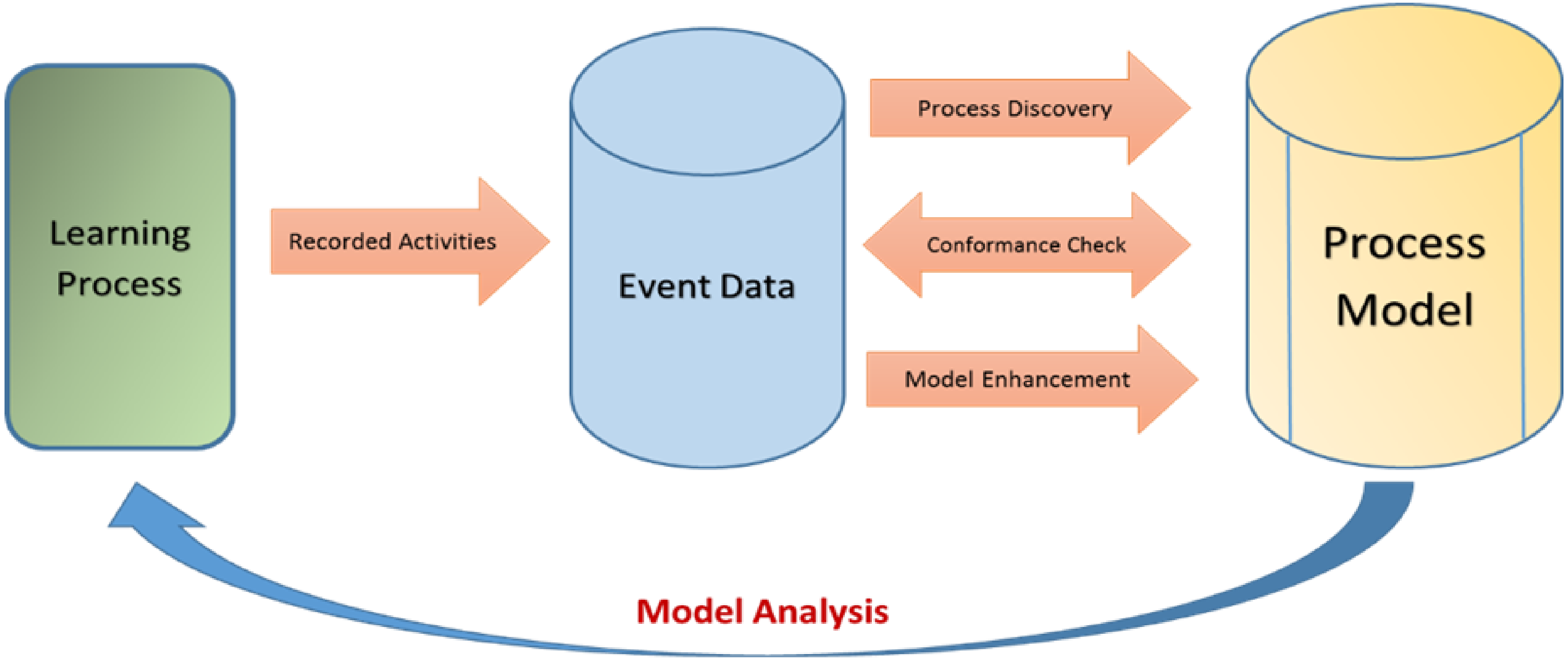
Process Mining is a new field that uses data mining techniques and process modelling to find out patterns or models from event logs, and predict outcomes through further analysis of the discovered models.

Types of Process Mining:

- ❑ Process Discovery
- ❑ Conformance Checking
- ❑ Model Enhancement

W. M. P. Van der Aalst (2003, 2004, 2011, 2016)

Application of the Process Mining Technique:



Application of the Process Mining Cont’nd…

Therefore, the main aspects of the process mining as shown in the above figure is described as follows:

- Process Discovery: applied to discover new process models from event Log about a learning process.
- Conformance Check: how much the data in the event log matches the presented behaviour in the deployed model?
- Model Extension: the need for both the model and its Logs to discover information that will enhance this model.
- Semantics Model Analysis: show how the analysis provided by the traditional process mining methods can be improved by adding semantic information to both the model and its logs based on the three basic building blocks: (i) Annotated Event Log/Model, (ii) Ontologies, and (iii) Semantic Reasoning.

Thus, the Research Plan & Key Core Elements:

- Process Mining: for extracting useful models from event Logs of a process, and augmenting information values of the resulting model through further semantic analysis of the discovered model

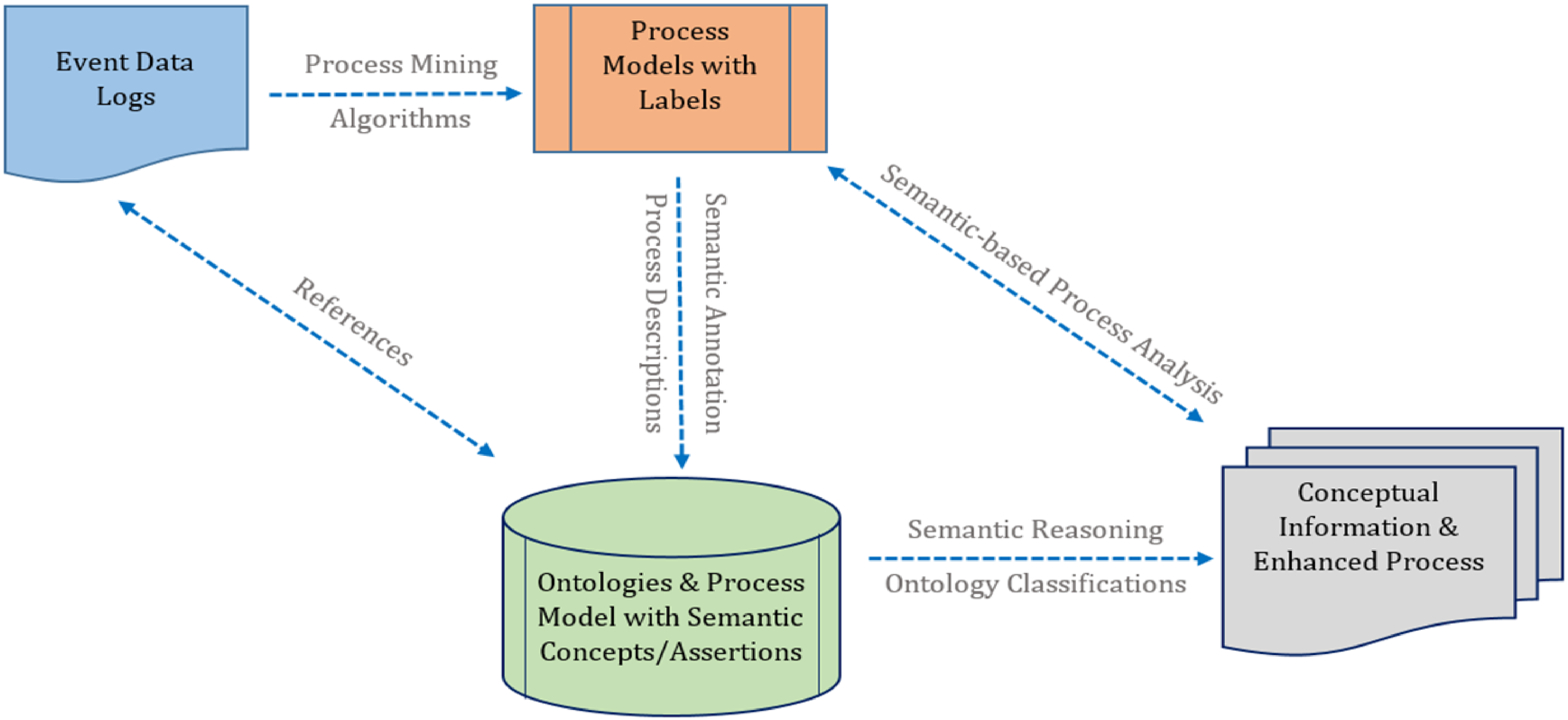
- Semantic Modelling: the process model and its logs enriched by using *Semantic Annotations* that links to concepts in an *Ontology* in order to extract useful patterns by means of *Semantic Reasoning*.

Proposed Method and Design Framework of the Thesis:

The work in this thesis claims that the quality augmentation of process models is as a result of employing process mining approaches that are capable of encoding the envisaged systems with the three rudimentary building blocks:

- *Semantic Labelling (annotation),*
- *Semantic Representation (ontology), and*
- *Semantic Reasoning (reasoner).*

The 2-Dimensional Rhombus Approach Framework:


Event Data Logs
Process Mining
Algorithms
Process Models with Labels
Semantic Annotation
Process Descriptions
References
Semantic-based Process Analysis
Ontologies & Process Model with Semantic Concepts/Assertions
Semantic Reasoning
Ontology Classifications
Conceptual Information & Enhanced Process

Design Framework Cont'nd…

Clearly, the 2-D Rhombus approach incorporates and informs the following:

- extraction of process models from event data logs: *the derived models are represented as a set of annotated terms that links and relates to defined terms in an ontology, and in so doing, encodes the process logs and the deployed models in the formal structure of ontology (semantic modelling).*
- the Reasoner (inference engine): *designed to perform automatic classification of task and consistency checking to validate the resulting model as well as clean out inconsistent results, and in turn, presents the inferred (underlying) associations.*

Design Framework Cont'nd…

- the inferred ontology classifications: helps associate meanings to labels within the event data logs and models by pointing to the concepts (references) defined within the ontology.
- the conceptual referencing: supports semantic reasoning over the ontologies in order to derive new information (or knowledge) about the process elements and the relationships they share amongst themselves within the knowledge base.

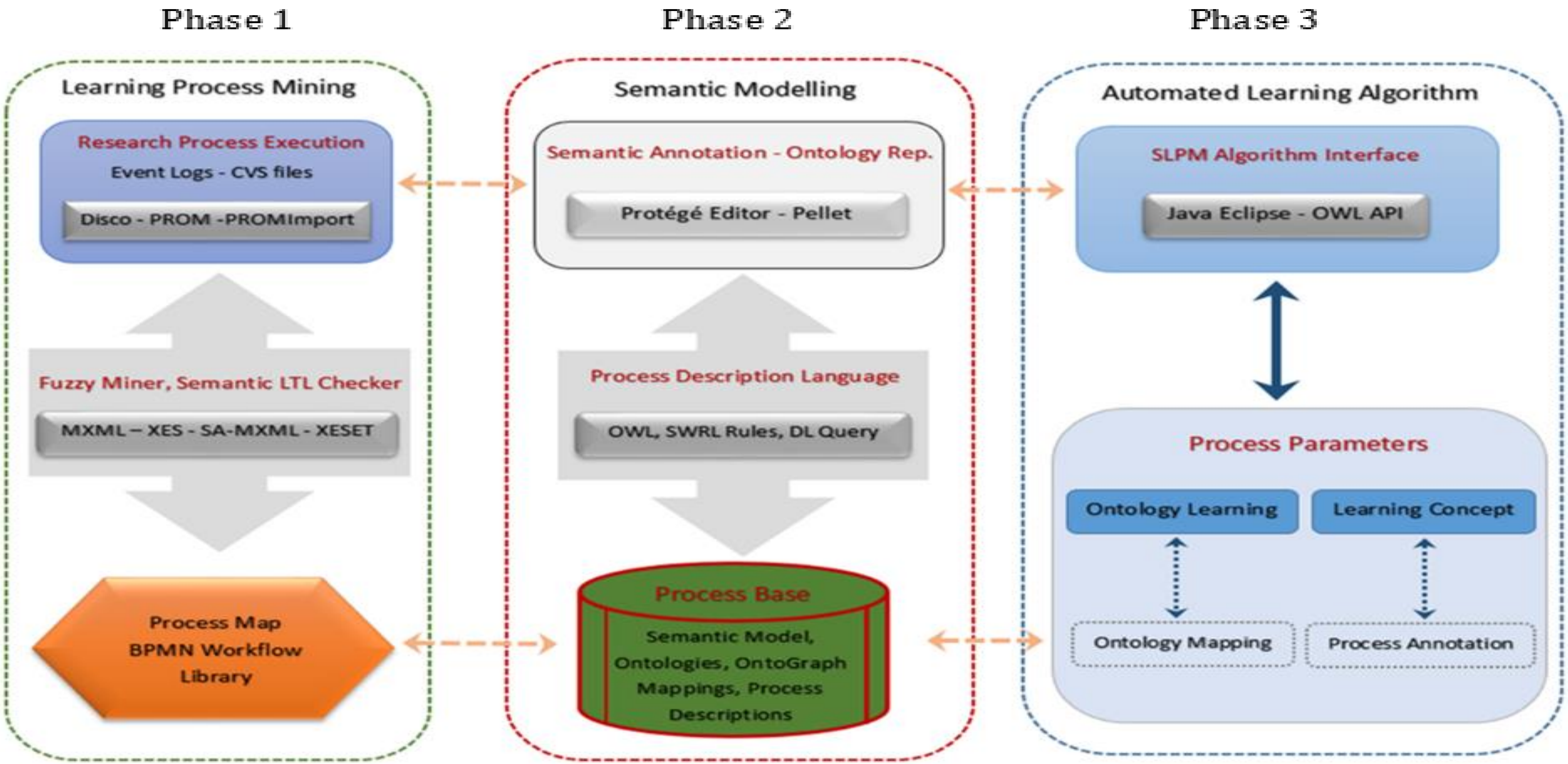
Summary of the Proposed Framework

To summarize the design framework, the work shows that the application of semantic-based process mining and analysis approaches must focus on feeding the mining algorithms with two key core elements:

(1) Event Logs and process models which their labels have references to concepts in an ontology, and

(2) Reasoners which are invoked to reason over the resulting ontologies produced from the logs and models.

Architecture of the proposed Semantic-based Process Mining Approach:



Phase 1

Learning Process Mining

Research Process Execution

Event Logs - CVS files

Disco - PROM -PROMImport

Fuzzy Miner, Semantic LTL Checker

MXML – XES - SA-MXML - XESET

Process Map
BPMN Workflow
Library

Phase 2

Semantic Modelling

Semantic Annotation - Ontology Rep.

Protégé Editor - Pellet

Process Description Language

OWL, SWRL Rules, DL Query

Process Base

Semantic Model,
Ontologies, OntoGraph
Mappings, Process
Descriptions

Phase 3

Automated Learning Algorithm

SLPM Algorithm Interface

Java Eclipse - OWL API

Process Parameters

Ontology Learning

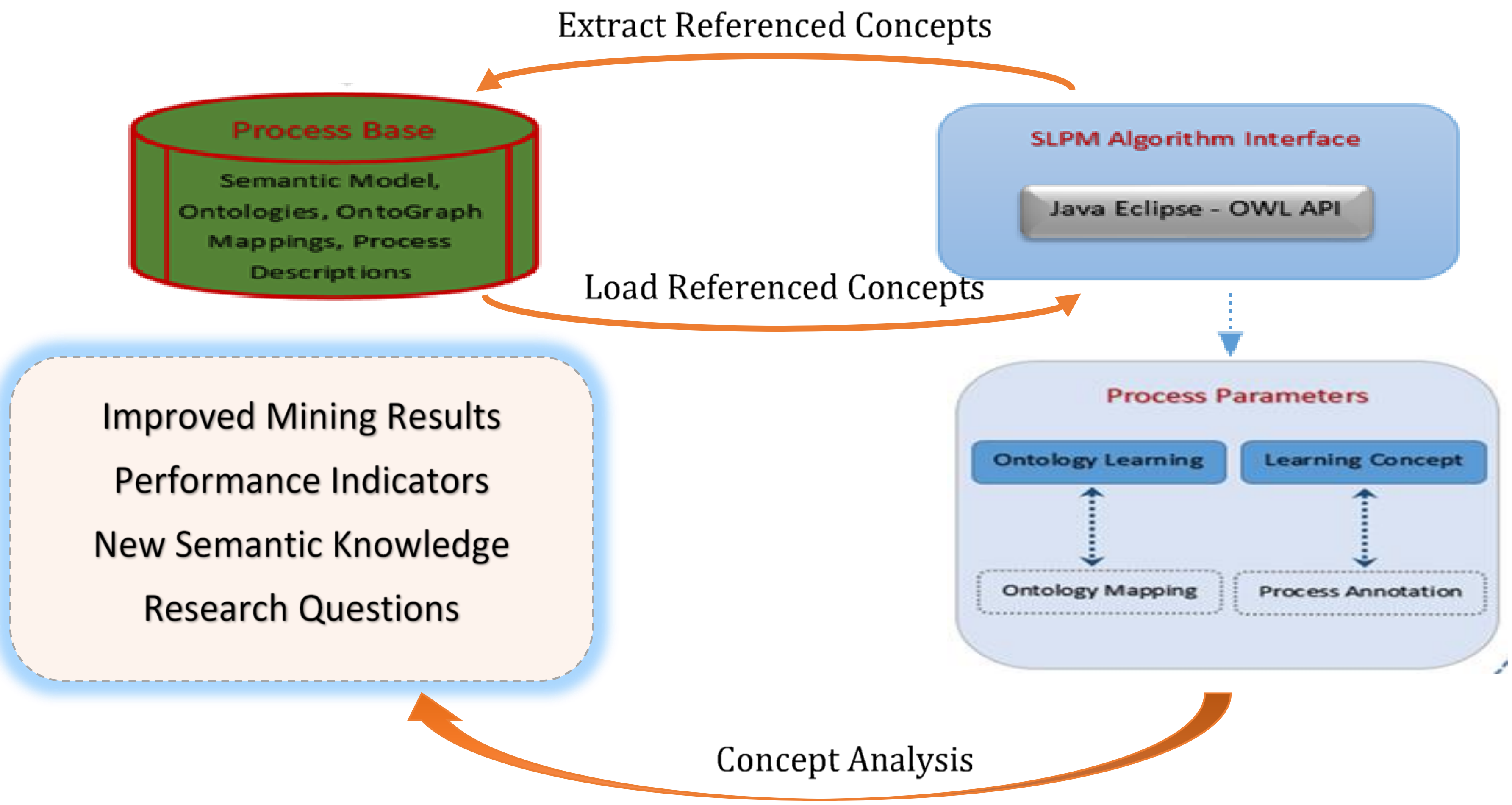
Learning Concept

Ontology Mapping

Process Annotation

Pre-modelling → **Modelling** → **Post-modelling**

Practical aspects of implementing the proposed system and its main functions



Extract Referenced Concepts

Process Base

Semantic Model, Ontologies, OntoGraph Mappings, Process Descriptions

SLPM Algorithm Interface

Java Eclipse - OWL API

Load Referenced Concepts

Improved Mining Results

Performance Indicators

New Semantic Knowledge

Research Questions

Process Parameters

Concept Analysis

Understanding the Different Phases of the Proposed Approach:

In Phase 1: the study applies the process mining techniques in order to make available the process mappings for the learning process, and check its conformance with the event logs based on the Fuzzy Miner. The main reason is that the resulting process map allows us to quickly, and interactively explore the processes into multiple directions and to show the learning activities workflows, and then provide platform for semantic annotation of the different process elements within the knowledge base.

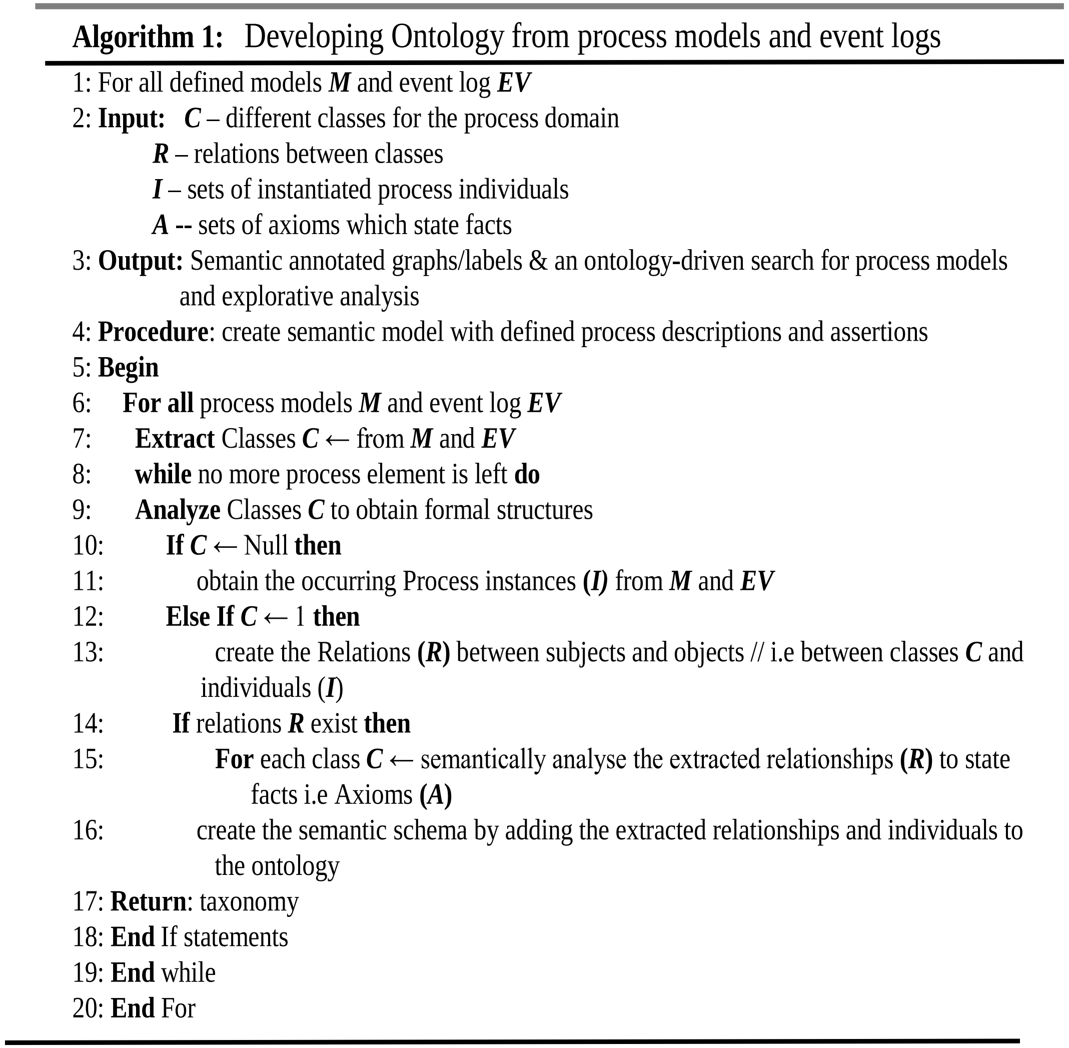
Phases Cont’nd…

In Phase 2: the work performs the semantic modelling of the resulting process mappings in terms of the annotated terms. Thus, the semantic model represents domain knowledge about the various learning activities and sequence workflows including the concepts defined in an Ontology by making use of process description languages such as the Ontology Web Rule Language (OWL) and Semantic Web Rule Language (SWRL), in addition to the conceptual reasoning capabilities of the Reasoner (i.e. Pellet) to infer the different process instances.

Phases Cont’nd…

In Phase 3: the research implements the semantic-based application used for extraction and automated mining of the learning concepts. The work uses the Eclipse developer tool to create the methods and interface for loading the Process Parameters. Essentially, the work makes use of the OWL API to extract and load the Inferred concepts. The purpose is to match the questions one would like to answer about the relationships the process instances share amongst themselves by linking to the inferred concepts within the learning ontology.

Proposed Semantic-based Algorithms and its Formalizations:



Algorithm 1: Developing Ontology from process models and event logs

```
For all defined models $M$ and event log $EV$
Input:  $C$ – different classes for the process domain
        $R$ – relations between classes
        $I$ – sets of instantiated process individuals
        $A$ -- sets of axioms which state facts
Output: Semantic annotated graphs/labels & an ontology-driven search for process models
        and explorative analysis
Procedure: create semantic model with defined process descriptions and assertions
Begin
  For all process models $M$ and event log $EV$
    Extract Classes $C \leftarrow$ from $M$ and $EV$
    while no more process element is left do
    Analyze Classes $C$ to obtain formal structures
      If $C \leftarrow$ Null then
        obtain the occurring Process instances ($I$) from $M$ and $EV$
      Else If $C \leftarrow 1$ then
          create the Relations ($R$) between subjects and objects // i.e between classes $C$ and
          individuals ($I$)
      If relations $R$ exist then
          For each class $C \leftarrow$ semantically analyse the extracted relationships ($R$) to state
            facts i.e Axioms ($A$)
        create the semantic schema by adding the extracted relationships and individuals to
          the ontology
Return: taxonomy
End If statements
End while
End For
```

Ultimately, from the described ***Algorithm 1***, we recognize that ontology is a quadruple, i.e.

Ont = (C,R,I,A)

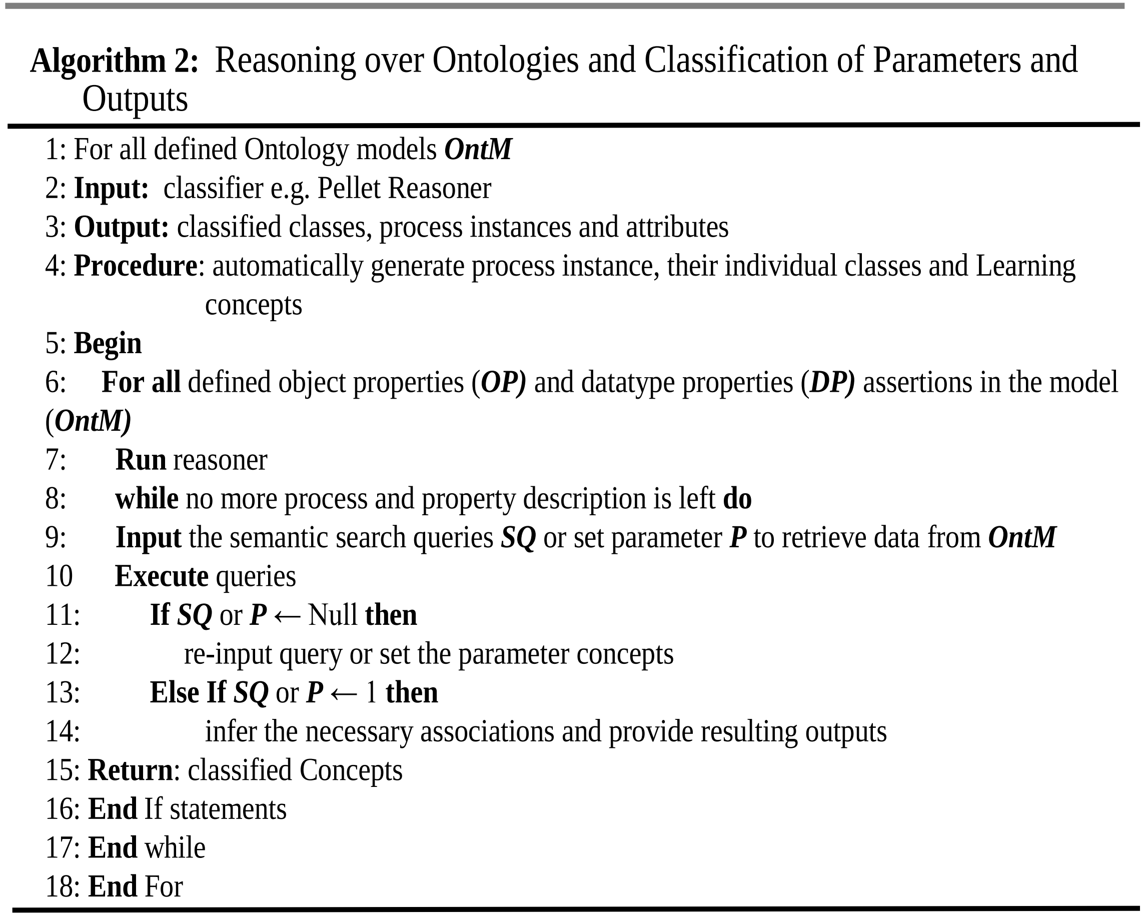
which consists of different ***classes, C,*** and ***relations, R,*** which trails to connect a set(s) of class with another class. Also, the classes are instantiated with a set(s) of ***Individual, I,*** and can likewise contain a set(s) of ***Axiom, A,*** which states fact (e.g. what is true and fitting within the model, or what is false and not fitting in the model).

Steps for the ***Algorithm 1*** Implementation:

To achieve this importance step in the study, it was necessary to:

- Create the various process domain ontologies, workflow ontologies, and the Individuals classes that will be inferred
- Provide Process Descriptions for all the Objects and Data Types that allows for Semantic Reasoning and Queries (i.e CLASS_ASSERTIONS; OBJECT_PROPERTY_ASSERTIONS; DATA_PROPERTY_ASSERTIONS)
- Create SWRL rules to map the existing class ontologies with concepts that are defined in the ontologies.
- Check for Consistency for all Defined Classes within the Model using the Description Logic Queries.

Algorithm 2: Semantic Reasoning

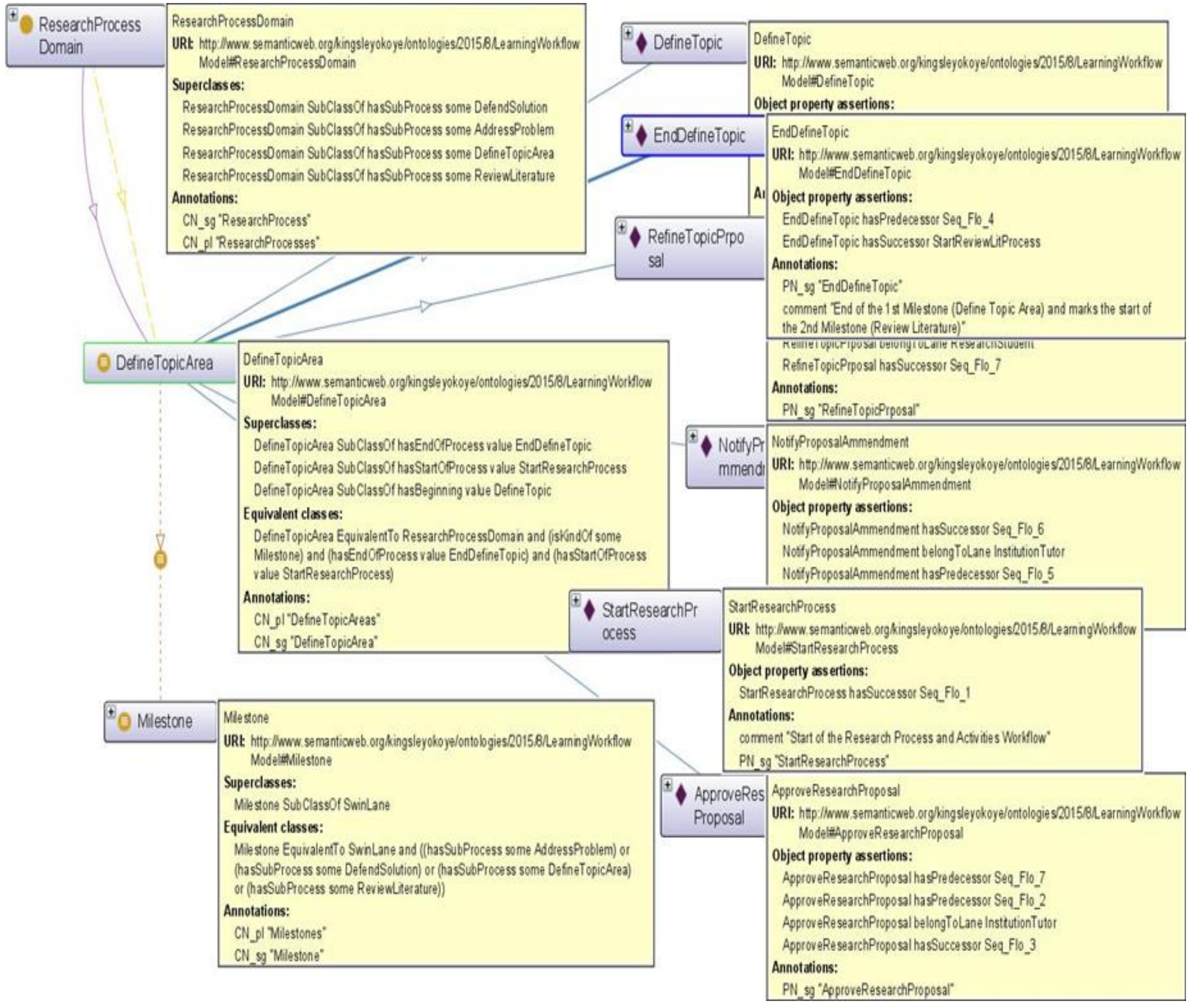


Algorithm 2: Reasoning over Ontologies and Classification of Parameters and Outputs

```
For all defined Ontology models OntM
Input:  classifier e.g. Pellet Reasoner
Output: classified classes, process instances and attributes
Procedure: automatically generate process instance, their individual classes and Learning
           concepts
Begin
    For all defined object properties (OP) and datatype properties (DP) assertions in the model (OntM)
      Run reasoner
      while no more process and property description is left do
      Input the semantic search queries SQ or set parameter P to retrieve data from OntM
      Execute queries
          If $SQ$ or $P \leftarrow$ Null then
              re-input query or set the parameter concepts
          Else If $SQ$ or $P \leftarrow 1$ then
                  infer the necessary associations and provide resulting outputs
Return: classified Concepts
End If statements
End while
End For
```

Indeed, as shown in the ***Algorithm 2***, the semantic reasoning helps to infer and associate meanings to labels within the defined ontologies by referring to the concepts assertions (i.e. Objects and Datatype properties) and sets of rules and/or expressions that are defined within the ontologies in order to answer and produce meaningful knowledge, and even in many cases, new information about the process elements and the relationships they share amongst themselves within the knowledge base.

Semantic Annotation:



ResearchProcessDomain
URI: http://www.semanticweb.org/kingsleyokoye/ontologies/2015/8/LearningWorkflowModel#ResearchProcessDomain
Superclasses:
ResearchProcessDomain SubClassOf hasSubProcess some DefendSolution
ResearchProcessDomain SubClassOf hasSubProcess some AddressProblem
ResearchProcessDomain SubClassOf hasSubProcess some DefineTopicArea
ResearchProcessDomain SubClassOf hasSubProcess some ReviewLiterature
Annotations:
CN_sg "ResearchProcess"
CN_pl "ResearchProcesses"

DefineTopic
URI: http://www.semanticweb.org/kingsleyokoye/ontologies/2015/8/LearningWorkflowModel#DefineTopic
Object property assertions:
A

EndDefineTopic
URI: http://www.semanticweb.org/kingsleyokoye/ontologies/2015/8/LearningWorkflowModel#EndDefineTopic
Object property assertions:
EndDefineTopic hasPredecessor Seq_Flo_4
EndDefineTopic hasSuccessor StartReviewLitProcess
Annotations:
PN_sg "EndDefineTopic"
comment "End of the 1st Milestone (Define Topic Area) and marks the start of the 2nd Milestone (Review Literature)"

RefineTopicPrposal belongToLane ResearchStudent
RefineTopicPrposal hasSuccessor Seq_Flo_7
Annotations:
PN_sg "RefineTopicPrposal"

DefineTopicArea
URI: http://www.semanticweb.org/kingsleyokoye/ontologies/2015/8/LearningWorkflowModel#DefineTopicArea
Superclasses:
DefineTopicArea SubClassOf hasEndOfProcess value EndDefineTopic
DefineTopicArea SubClassOf hasStartOfProcess value StartResearchProcess
DefineTopicArea SubClassOf hasBeginning value DefineTopic
Equivalent classes:
DefineTopicArea EquivalentTo ResearchProcessDomain and (isKindOf some Milestone) and (hasEndOfProcess value EndDefineTopic) and (hasStartOfProcess value StartResearchProcess)
Annotations:
CN_pl "DefineTopicAreas"
CN_sg "DefineTopicArea"

NotifyProposalAmmendment
URI: http://www.semanticweb.org/kingsleyokoye/ontologies/2015/8/LearningWorkflowModel#NotifyProposalAmmendment
Object property assertions:
NotifyProposalAmmendment hasSuccessor Seq_Flo_6
NotifyProposalAmmendment belongToLane InstitutionTutor
NotifyProposalAmmendment hasPredecessor Seq_Flo_5

StartResearchProcess
URI: http://www.semanticweb.org/kingsleyokoye/ontologies/2015/8/LearningWorkflowModel#StartResearchProcess
Object property assertions:
StartResearchProcess hasSuccessor Seq_Flo_1
Annotations:
comment "Start of the Research Process and Activities Workflow"
PN_sg "StartResearchProcess"

Milestone
URI: http://www.semanticweb.org/kingsleyokoye/ontologies/2015/8/LearningWorkflowModel#Milestone
Superclasses:
Milestone SubClassOf SwinLane
Equivalent classes:
Milestone EquivalentTo SwinLane and ((hasSubProcess some AddressProblem) or (hasSubProcess some DefendSolution) or (hasSubProcess some DefineTopicArea) or (hasSubProcess some ReviewLiterature))
Annotations:
CN_pl "Milestones"
CN_sg "Milestone"

ApproveResearchProposal
URI: http://www.semanticweb.org/kingsleyokoye/ontologies/2015/8/LearningWorkflowModel#ApproveResearchProposal
Object property assertions:
ApproveResearchProposal hasPredecessor Seq_Flo_7
ApproveResearchProposal hasPredecessor Seq_Flo_2
ApproveResearchProposal belongToLane InstitutionTutor
ApproveResearchProposal hasSuccessor Seq_Flo_3
Annotations:
PN_sg "ApproveResearchProposal"

Semantic Annotation ***(SemAn)*** is function that returns a set of concepts from the ontology for each node or edge in the graph. Thus,

$$\boldsymbol{SemAn :: N \cup E \rightarrow COnts}$$

where: ***SemAn*** describes all kinds of annotations which can be input, output, meta-model annotation etc.

Semantic Annotation Cont'nd…

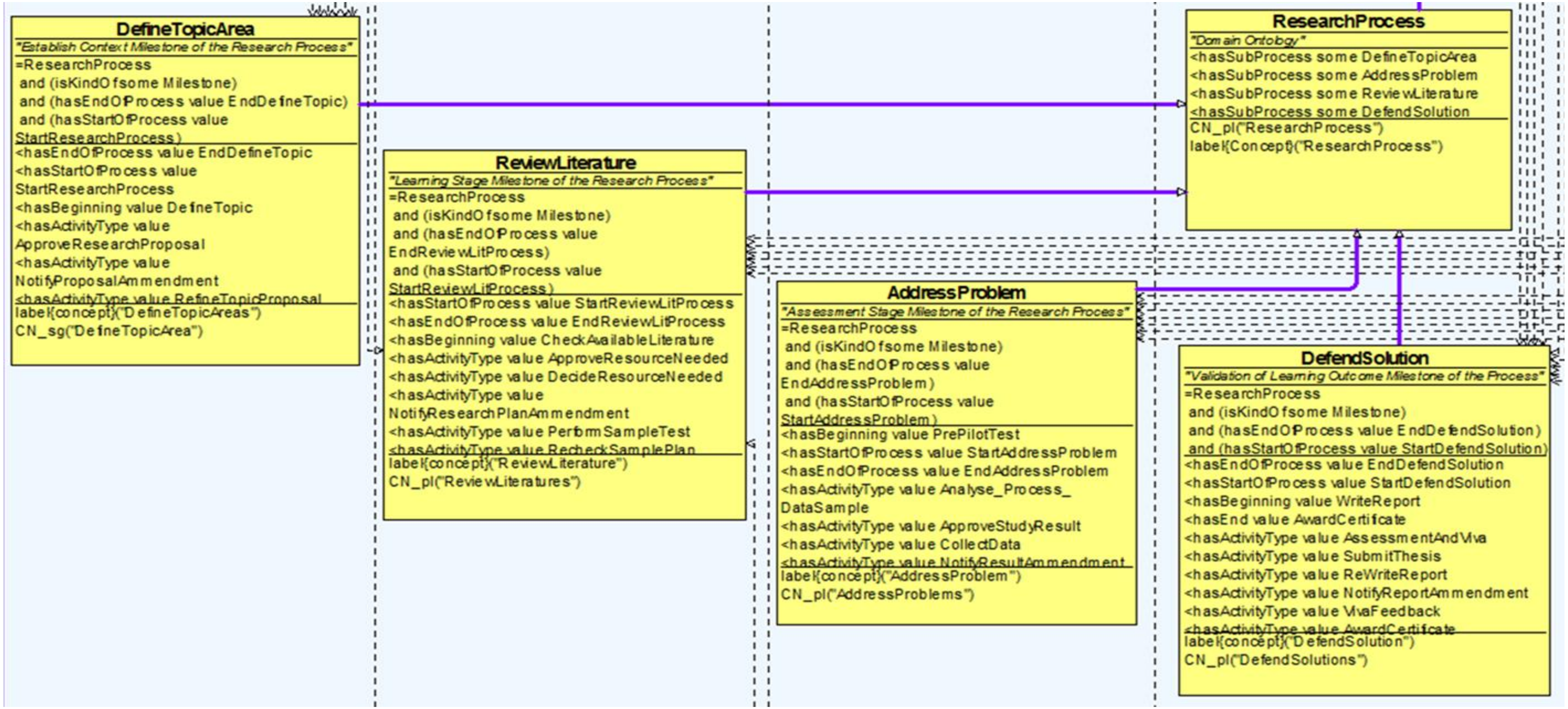
Typically, a *semantic annotated graph* is defined as follows:

- $Gsem = (Nsem, Esem, Onts)\ with\ Nsem = \{(n, SemAn(n)) | n \in N\}\ and\ Esem = \{(nsem, n\_sem) | nsem = (n, SemAn(n)) \land n\_sem = (n\_, SemAn(n\_)) \land (n, n\_) \in E\}$ (Lautenbacher, et al., 2009).

Thus;

- Let ***A*** be the set of all process actions. A process action ***a ∈ A*** is characterized by a set of input parameters ***Ina ∈ P***, which is required for the execution of ***a*** and a set of output parameters ***Outa ⊆ P***, which is provided by ***a*** after execution. All elements ***a ∈ A*** are stored as a triple (***namea, Ina, Outa***) in a process library ***libA***. (Lautenbacher, et al., 2009).

Use Case Scenario, Implementation & Experimental Setup:



DefineTopicArea

"Establish Context Milestone of the Research Process"

=ResearchProcess
and (isKindOfsome Milestone)
and (hasEndOfProcess value EndDefineTopic)
and (hasStartOfProcess value StartResearchProcess)

<hasEndOfProcess value EndDefineTopic
<hasStartOfProcess value StartResearchProcess
<hasBeginning value DefineTopic
<hasActivityType value ApproveResearchProposal
<hasActivityType value NotifyProposalAmmendment
<hasActivityType value RefineTopicProposal

label{concept}("DefineTopicAreas")
CN_sg("DefineTopicArea")

ReviewLiterature

"Learning Stage Milestone of the Research Process"

=ResearchProcess
and (isKindOfsome Milestone)
and (hasEndOfProcess value EndReviewLitProcess)
and (hasStartOfProcess value StartReviewLitProcess)

<hasStartOfProcess value StartReviewLitProcess
<hasEndOfProcess value EndReviewLitProcess
<hasBeginning value CheckAvailableLiterature
<hasActivityType value ApproveResourceNeeded
<hasActivityType value DecideResourceNeeded
<hasActivityType value NotifyResearchPlanAmmendment
<hasActivityType value PerformSampleTest
<hasActivityType value RecheckSamplePlan

label{concept}("ReviewLiterature")
CN_pl("ReviewLiteratures")

AddressProblem

"Assessment Stage Milestone of the Research Process"

=ResearchProcess
and (isKindOfsome Milestone)
and (hasEndOfProcess value EndAddressProblem)
and (hasStartOfProcess value StartAddressProblem)

<hasBeginning value PrePilotTest
<hasStartOfProcess value StartAddressProblem
<hasEndOfProcess value EndAddressProblem
<hasActivityType value Analyse_Process_DataSample
<hasActivityType value ApproveStudyResult
<hasActivityType value CollectData
<hasActivityType value NotifyResultAmmendment

label{concept}("AddressProblem")
CN_pl("AddressProblems")

ResearchProcess

"Domain Ontology"

<hasSubProcess some DefineTopicArea
<hasSubProcess some AddressProblem
<hasSubProcess some ReviewLiterature
<hasSubProcess some DefendSolution

CN_pl("ResearchProcess")
label{Concept}("ResearchProcess")

DefendSolution

"Validation of Learning Outcome Milestone of the Process"

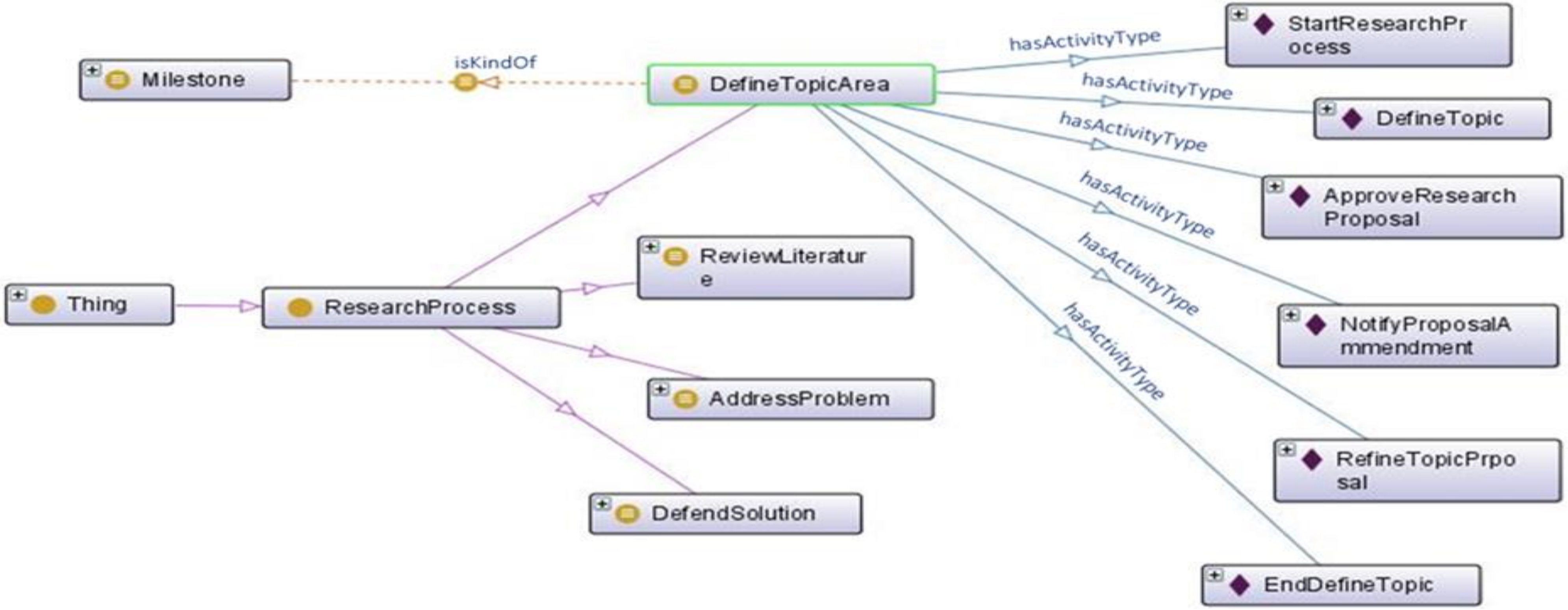
=ResearchProcess
and (isKindOfsome Milestone)
and (hasEndOfProcess value EndDefendSolution)
and (hasStartOfProcess value StartDefendSolution)

<hasEndOfProcess value EndDefendSolution
<hasStartOfProcess value StartDefendSolution
<hasBeginning value WriteReport
<hasEnd value AwardCertificate
<hasActivityType value AssessmentAndViva
<hasActivityType value SubmitThesis
<hasActivityType value ReWriteReport
<hasActivityType value NotifyReportAmmendment
<hasActivityType value VivaFeedback
<hasActivityType value AwardCertificate

label{concept}("DefendSolution")
CN_pl("DefendSolutions")

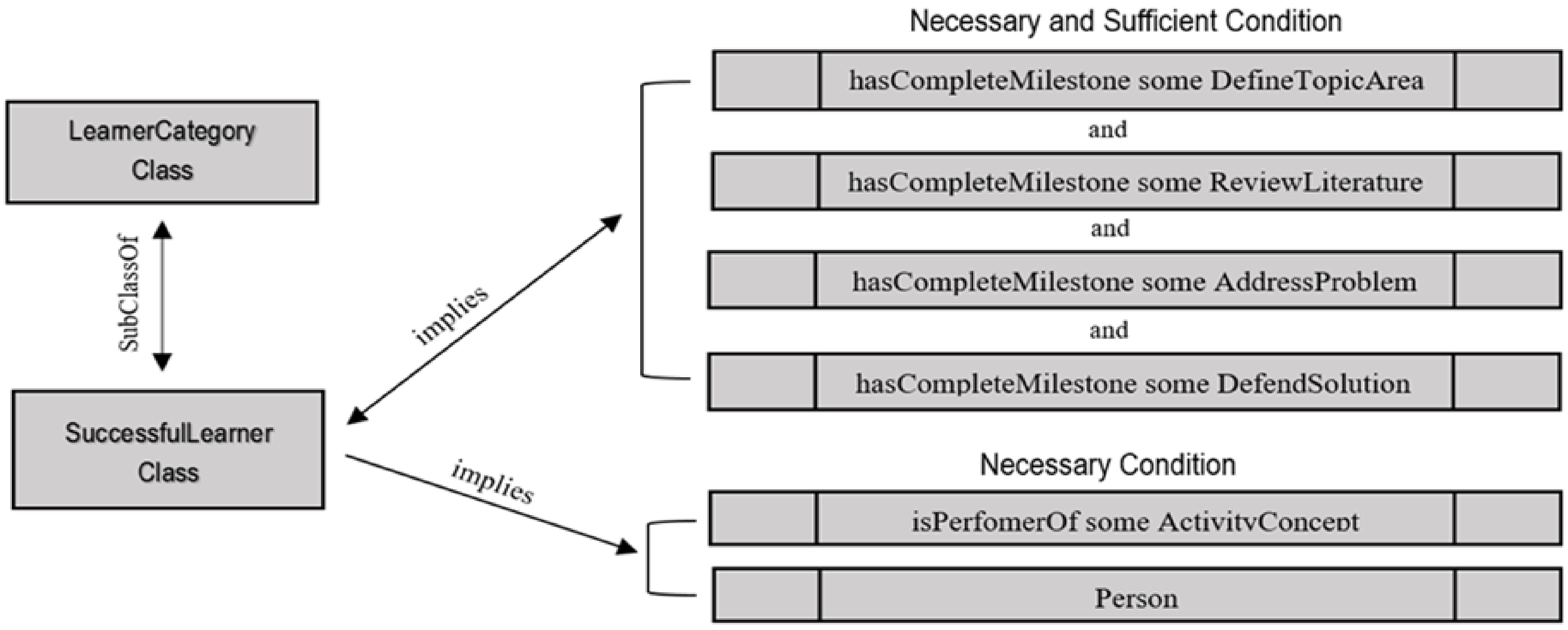
Semantic Representation and Modelling of Research Learning Process.

Example of OntoGraph and the ActivityConcept mapping for the DefineTopicArea Milestone.



Indeed, the drive for such semantic mapping of the activity concepts is that the method allows the meaning of the learning objects and properties to be enhanced through the use of ***property descriptions*** and ***classification of the discoverable entities*** (i.e. the inferred classes or concepts).

Description of Concepts: Example of a Successful Learner Class



Necessary and Sufficient Condition
LearnerCategory Class
SubClassOf
SuccessfulLearner Class
implies
hasCompleteMilestone some DefineTopicArea
and
hasCompleteMilestone some ReviewLiterature
and
hasCompleteMilestone some AddressProblem
and
hasCompleteMilestone some DefendSolution
implies
Necessary Condition
isPerfomerOf some ActivityConcept
Person

As shown in the **Figure** - the necessary condition is: if something is a Successful Learner, it is necessary for it to be a participant of the Learning ActivityConcept class and necessary for it to have a kind of sufficiently defined condition and relationship with the ResearchProcess subClasses: (i.e DefineTopicArea, ReviewLiterature, AddressProblem and DefendSolution) etc.

Classes & Concepts Cont’nd…

For example, the following are description of the implemented ontology concepts and axioms for the “successful learner” class within the learning model including the OWL XML file syntax as follows:

1: **ontology** ResearchProcess

2: **concept** SuccessfulLearner

3: hascompleteMilestone **ofType** {DefineTopicArea, ReviewLiterature, AddressProblem, DefendSolution}

4: isPerformerOf **some** LearningActivity

5: is **ofType** Person

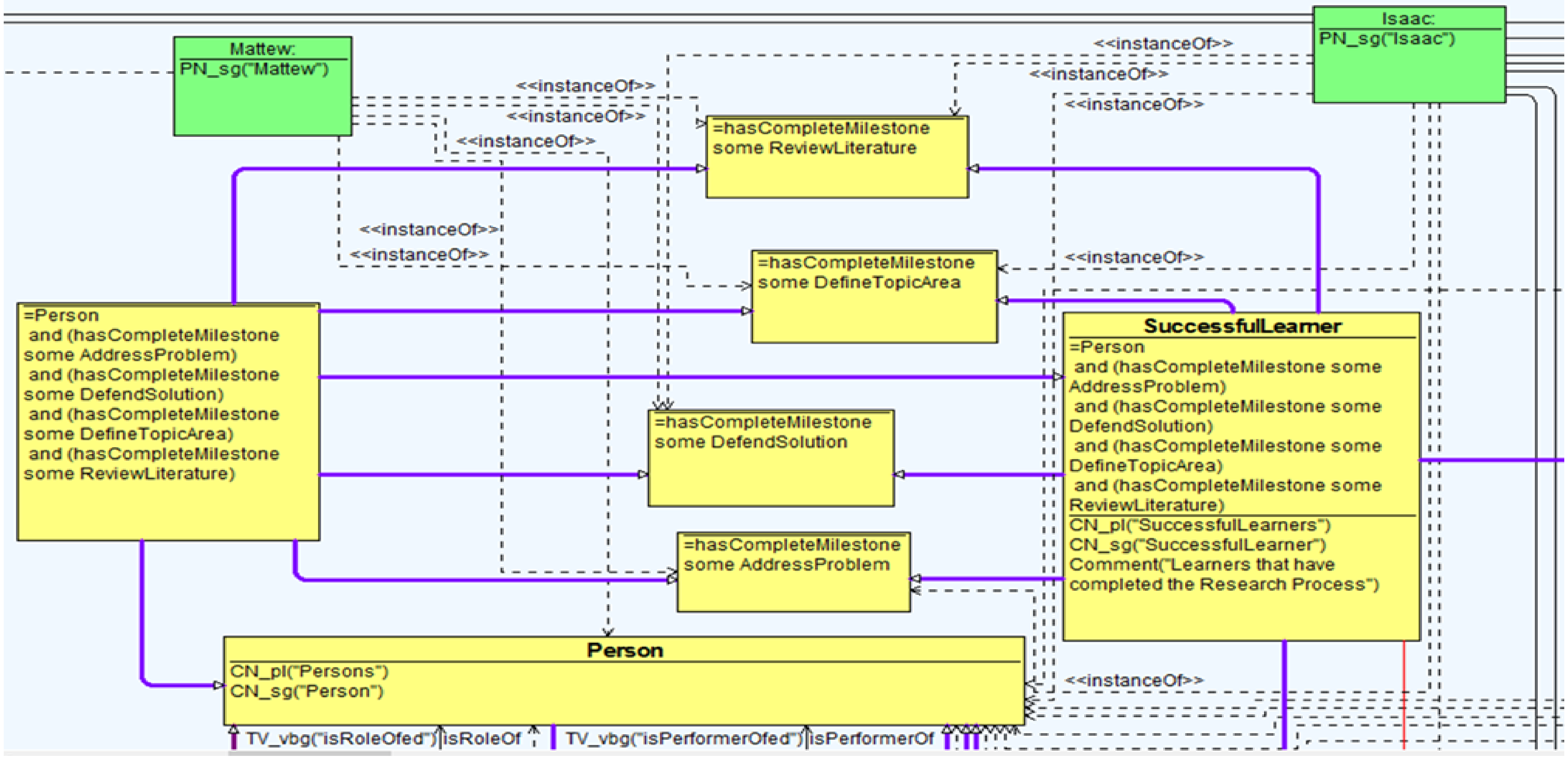
6: hasInstance **members** {Mattew, Isaac}

7: **axiom** DefinitionOfSuccessfulLearner

Classes & Concepts Cont’nd….

```
<EquivalentClasses>
        <Annotation>
            <AnnotationProperty
IRI="http://attempto.ifi.uzh.ch/acetext#acetext"/>
            <Literal datatypeIRI="&xsd;string">Every SuccessfulLearner
is a Person that hasMilestones an AddressProblem and that
hasMilestones a DefendSolution and that hasMilestones a
DefineTopicArea and that hasMilestones a ReviewLiterature. Every
Person that hasMilestones an AddressProblem and that hasMilestones a
DefendSolution and that hasMilestones a DefineTopicArea and that
hasMilestones a ReviewLiterature is a SuccessfulLearner.</Literal>
        </Annotation>
</EquivalentClasses>
```

Concept assertions and the different formal relationships for the **SuccessfulLearner Class**



Mattew:
PN_sg("Mattew")

Isaac:
PN_sg("Isaac")

=hasCompleteMilestone
some ReviewLiterature

=hasCompleteMilestone
some DefineTopicArea

=Person
and (hasCompleteMilestone
some AddressProblem)
and (hasCompleteMilestone
some DefendSolution)
and (hasCompleteMilestone
some DefineTopicArea)
and (hasCompleteMilestone
some ReviewLiterature)

SuccessfulLearner
=Person
and (hasCompleteMilestone some
AddressProblem)
and (hasCompleteMilestone some
DefendSolution)
and (hasCompleteMilestone some
DefineTopicArea)
and (hasCompleteMilestone some
ReviewLiterature)
CN_pl("SuccessfulLearners")
CN_sg("SuccessfulLearner")
Comment("Learners that have
completed the Research Process")

=hasCompleteMilestone
some DefendSolution

=hasCompleteMilestone
some AddressProblem

Person
CN_pl("Persons")
CN_sg("Person")

Fuzzy-BPMN Approach: Experimentations and Implementation

The research shows how it practically apply current tools that supports process mining by participating in the **First Process Discovery Contest** (Carmona et al., 2016) organised by the **IEEE CIS Task Force on Process Mining.**

- 10 different **Event Logs** **(**each for the ***Training Log*** and ***Test logs*****)** generated from a business process models that shows different behavioural characteristics were provided by the group for the contest. Each of the test event logs are characterised to have 10 different traces that can be replayed and other 10 traces that cannot be replayed. Making a total of 20 traces for each test event log. i.e

10 test logs x 20 traces which equals to a total of = 200 Traces

where: 100 traces are replayable and other 100 traces are not replayable by the original model.

Process Discovery Contest Cont’nd

The aim of the contest and the submission was to carry out a classification task to determine the individual traces that makes up the ***Test event logs*** and then cross-validated against the ***Training Log*** in order to determine which traces that can be replayed by the original model. In other words;

- **Given a trace (t) representing real process behaviour, the process model (m) classifies it as allowed, or**
- **Given a trace (t) representing a behaviour not related to the process, the process model (m) classifies it as disallowed.**

In the following **Table 1:** the study presents the classification results of the Fuzzy-BPMN miner approach where each individual cell indicates if the discovered model classifies the corresponding trace as fitting (allowed) or not fitting (disallowed).

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
Trace_1	TP *	TN *	TP *	FP	TN *	FP	TP *	TP *	TP *	TP *
Trace_2	TN *	TN *	TP *	TP *	TP *	TP *	TP *	TN *	TP *	TP *
Trace_3	TP *	TP *	TP *	TN *	TN *	FP	FP	TP *	TP *	TN *
Trace_4	TP *	TP *	FP	TP *	TN *	TP *	TN *	TP *	TP *	FP
Trace_5	TN *	FP	FP	TP *	TN *	TP *	TN *	TP *	TP *	TN *
Trace_6	TP *	FP	FP	TP *	TN *	TP *	TP *	TN *	TN *	TP *
Trace_7	TN *	TP *	TP *	TN *	TN *	TP *	TN *	TP *	TN *	TN *
Trace_8	TN *	TP *	TP *	FN	TN *	FP	TP *	TP *	TP *	TP *
Trace_9	TP *	TN *	TP *	TN *	TP *	FP	TP *	TP *	TN *	TP *
Trace_10	TP *	FP	TP *	TN *	TN *	FP	TP *	TP *	TP *	TP *
Trace_11	TN *	TP *	TP *	FN	TP *	TN *	TN *	FP	TN *	TP *
Trace_12	TP *	FP	FP	TP *	TP *	TP *	TP *	FP	TP *	TN *
Trace_13	TP *	TP *	FP	TN *	TP *	FP	TN *	TN *	TN *	TP *
Trace_14	TN *	TP *	TN *	TN *	TN *	FP	TN *	TP *	TN *	TP *
Trace_15	TP *	TN *	TN *	TN *	TP *	TP *	TN *	TN *	TN *	TN *
Trace_16	TN *	TN *	FP	TP *	TP *	FP	TN *	FP	TP *	TN *
Trace_17	TP *	TP *	TP *	TP *	TP *	TP *	TP *	TN *	TN *	TP *
Trace_18	TN *	TP *	FP	TN *	TP *	TP *	TP *	TN *	TN *	TN *
Trace_19	TN *	TP *	TP *	TP *	TN *	TP *	TP *	TP *	TN *	TN *
Trace_20	TN *	TN *	FP	TN *	TP *	FP	TN *	TN *	TP *	TN *

True Positive (TP) :	10	10	10	8	10	10	10	10	10	10
False Positive (FP):	0	4	8	1	0	9	1	3	0	1
True Negative (TN):	10	6	2	9	10	1	9	7	10	9
False Negative (FN):	0	0	0	2	0	0	0	0	0	0
NO. of traces correctly classified	20	16	12	17	20	11	19	17	20	19

The cells colours indicates the classification attempt for each of the traces discovered from the test event logs. Also, the cells with gold sign * indicates the traces that were correctly classified by the Fuzzy-BPMN Miner with total of 171 traces out of 200.

Performance Metrics:

The following performance metrics (Van der Aalst, 2016) were used to measure the fitness of the individual traces for the datasets, where:

- **TP** is the number of **true positives** i.e. instances that are correctly classified as positive
- **FN** is the number of **false negatives** i.e. instances that are predicted to be negative but should have been classified as positive
- **FP** is the number of **false positives** i.e. instances that are predicted to be positive but should have been classified as negative
- **TN** is the number of **true negatives** (i.e. instances that are correctly classified as negative)

Indeed, the final result after scoring by the committee (panel of judges) shows that the Fuzzy-BPMN miner approach has correctly classified 171 out of 200 (85.5%) traces in the original process model.

Semantic-Fuzzy Mining: Experimentations Outcomes and Analysis:

The research also makes use of the event logs used for the IEEE CIS Task Force on Process Mining contest to describe how the work expounds the Fuzzy-BPMN approach in order to weigh up the performance of the proposed Semantic-based Fuzzy miner being able to perform a more accurate classification of the individual traces within the process base.

- This includes the capability to integrate ontological concepts and the semantic annotations in order to perform semantic reasoning capable of discovering worthwhile models with abstraction levels of information (i.e. semantic knowledge) given the datasets (*training set* and *test set*) for the cross-validation experiments.

Semantic-Fuzzy Mining Approach cont'nd…

Indeed, the semantic fuzzy mining approach and application references a number of different OWL ontologies (e.g. training model ontology, test set ontology, traceFitness Classification ontology etc.) which were created for the experiment.

- For each ontology, all concepts in their turn were considered by the reasoner and are checked for consistency by referencing the process parameters.
- The corresponding traces were computed and recorded according to the reasoner response, and the classification process was tested on the resulting individuals by assessing its performance with respect to correctly classified traces. For each result of the classification process, the replayable (true positives) and non-replayable (true negatives) traces were learned.

Semantic-Fuzzy Mining Approach cont’nd…

For instance, the work executes the DL queries below as a set of input parameters to output the set of traces for the example “TestLog_Apri_1” within the model that has 'TrueTrace_Fitness_(TP)' and 'FalseTrace_Fitness_(TN)' respectively.

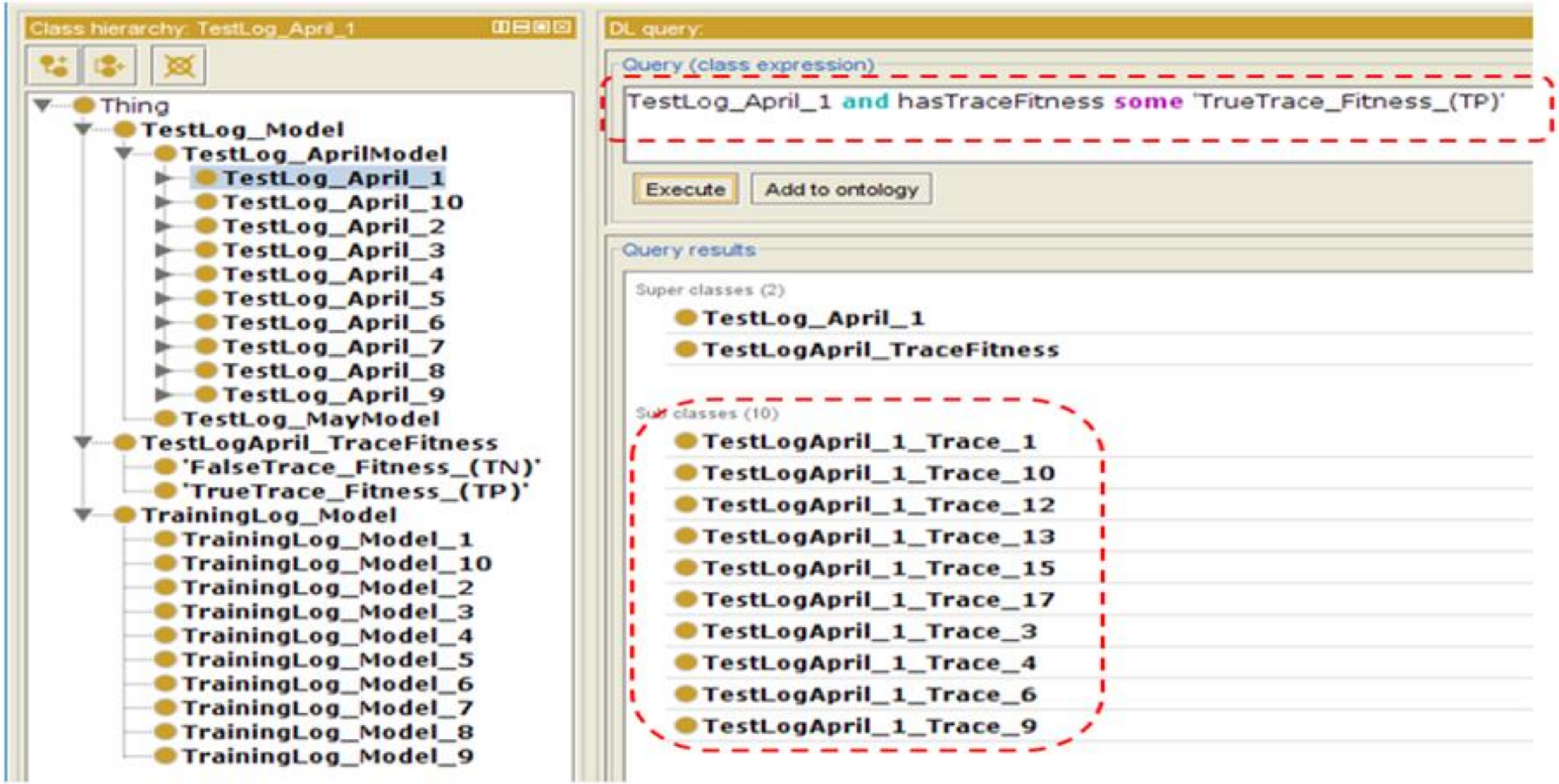
Thus:

“TestLog_April_1 and hasTraceFitness some 'TrueTrace_Fitness_(TP)'”

“TestLog_April_1 and hasTraceFitness some 'FalseTrace_Fitness_(TN)'”

The results of computing the input and output parameters, for example, the 'TrueTrace_Fitness_(TP)' are as shown in the following Figure.

Example of the TrueTrace_Fitness_(TP) classification for the TestLog_April_1 with the correctly classified traces.



Class hierarchy: TestLog_April_1

- Thing
  - TestLog_Model
    - TestLog_AprilModel
      - TestLog_April_1
      - TestLog_April_10
      - TestLog_April_2
      - TestLog_April_3
      - TestLog_April_4
      - TestLog_April_5
      - TestLog_April_6
      - TestLog_April_7
      - TestLog_April_8
      - TestLog_April_9
    - TestLog_MayModel
  - TestLogApril_TraceFitness
    - 'FalseTrace_Fitness_(TN)'
    - 'TrueTrace_Fitness_(TP)'
  - TrainingLog_Model
    - TrainingLog_Model_1
    - TrainingLog_Model_10
    - TrainingLog_Model_2
    - TrainingLog_Model_3
    - TrainingLog_Model_4
    - TrainingLog_Model_5
    - TrainingLog_Model_6
    - TrainingLog_Model_7
    - TrainingLog_Model_8
    - TrainingLog_Model_9

```
TestLog_April_1 and hasTraceFitness some 'TrueTrace_Fitness_(TP)'
```

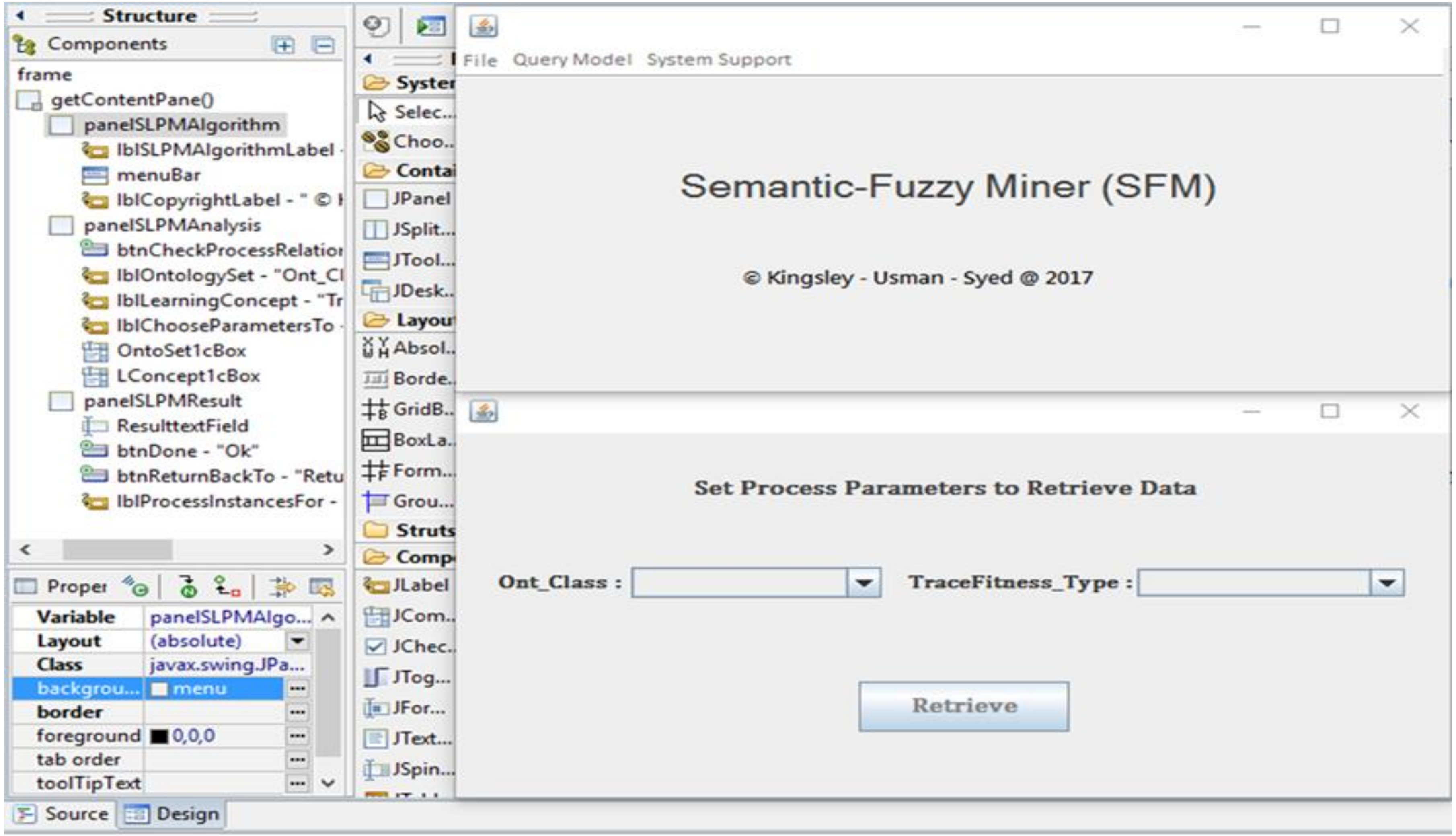
Super classes (2)

- TestLog_April_1
- TestLogApril_TraceFitness

Sub classes (10)

- TestLogApril_1_Trace_1
- TestLogApril_1_Trace_10
- TestLogApril_1_Trace_12
- TestLogApril_1_Trace_13
- TestLogApril_1_Trace_15
- TestLogApril_1_Trace_17
- TestLogApril_1_Trace_3
- TestLogApril_1_Trace_4
- TestLogApril_1_Trace_6
- TestLogApril_1_Trace_9

Application Interface for the semantic-fuzzy miner (SFM) in Eclipse



Experimental Outcome of the Semantic-Fuzzy Mining Approach:

The outcome of the experiments with regards to the defined models and the classification of the corresponding individual traces occurring in each test set are as reported in the next following Table 2.

- The study observes that for every run set of parameters, the commission error, i.e. false positives (FP) and false negatives (FN) was null, thus equal to 0. This means that the classifier did not make critical mistakes. For example, settings where a trace is deemed to be an instance of a class while it really is an instance of another class.
- At the same time, the study observes that the trace accuracy rates was very high i.e. for the true positives (TP) and true negatives (TN), and were consistently observed for all the test sets.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
Trace_1	TP *	TN *	TP *	TN *	TN *	TN *	TP *	TP *	TP *	TP *
Trace_2	TN *	TN *	TP *	TP *	TP *	TP *	TP *	TN *	TP *	TP *
Trace_3	TP *	TP *	TP *	TN *	TN *	TN *	TN *	TP *	TP *	TN *
Trace_4	TP *	TP *	TN *	TP *	TN *	TP *	TN *	TP *	TP *	TN *
Trace_5	TN *	TN *	TN *	TP *	TN *	TP *	TN *	TP *	TP *	TN *
Trace_6	TP *	TN *	TN *	TP *	TN *	TP *	TP *	TN *	TN *	TP *
Trace_7	TN *	TP *	TP *	TN *	TN *	TP *	TN *	TP *	TN *	TN *
Trace_8	TN *	TP *	TP *	TP *	TN *	TN *	TP *	TP *	TP *	TP *
Trace_9	TP *	TN *	TP *	TN *	TP *	TN *	TP *	TP *	TN *	TP *
Trace_10	TP *	TN *	TP *	TN *	TN *	TN *	TP *	TP *	TP *	TP *
Trace_11	TN *	TP *	TP *	TP *	TP *	TN *	TN *	TN *	TN *	TP *
Trace_12	TP *	TN *	TN *	TP *	TP *	TP *	TP *	TN *	TP *	TN *
Trace_13	TP *	TP *	TN *	TN *	TP *	TN *	TN *	TN *	TN *	TP *
Trace_14	TN *	TP *	TN *	TN *	TN *	TN *	TN *	TP *	TN *	TP *
Trace_15	TP *	TN *	TN *	TN *	TP *	TP *	TN *	TN *	TN *	TN *
Trace_16	TN *	TN *	TN *	TP *	TP *	TN *	TN *	TN *	TP *	TN *
Trace_17	TP *	TP *	TP *	TP *	TP *	TP *	TP *	TN *	TN *	TP *
Trace_18	TN *	TP *	TN *	TN *	TP *	TP *	TP *	TN *	TN *	TN *
Trace_19	TN *	TP *	TP *	TP *	TN *	TP *	TP *	TP *	TN *	TN *
Trace_20	TN *	TN *	TN *	TN *	TP *	TN *	TN *	TN *	TP *	TN *

True Positive (TP) :	10	10	10	10	10	10	10	10	10	10
False Positive (FP):	0	0	0	0	0	0	0	0	0	0
True Negative (TN):	10	10	10	10	10	10	10	10	10	10
False Negative (FN):	0	0	0	0	0	0	0	0	0	0
Number of traces correctly classified	20	20	20	20	20	20	20	20	20	20

The cells colours indicates if the specified trace has been classified as true positives (TP) or true negatives (TN). All the cells with gold sign * indicates traces that were correctly classified by the Semantic-Fuzzy Miner with total of 200 traces out of 200.

Evaluation of Research Outcomes:

Qualitative Evaluation of the Semantic Fuzzy mining Approach and Outcomes

Evidence from the research design framework, algorithms and experimentations shows that the semantic-based approach sparks methods that highly influence and support:

(i) the application of process mining techniques to domain processes, and

(ii) provision of real time semantic knowledge and understanding about the domain processes (e.g. case study of learning process) which are useful towards the development of process mining algorithms that are more intelligent with high level of effective conceptual reasoning capabilities.

Semantic-fuzzy miner vs Semantic LTL Checker (deMedeiros, et al., 2008)

	Semantic LTL Checker	Semantic-Fuzzy Miner
Data Input	Takes event Logs concepts as input to parameters of Linear Temporal Logic (LTL) formulae	Takes process models derived from fuzzy mining of event log as input to learn and reason about the domain process
Ontology	Ontologies are defined in WSML format	Ontologies are defined in OWL and SWRL format
Reasoning	Integrated using the WSML2Reasoner (W2RF)	Integrated using the Pellet Reasoner
Functionality	Uses LTL properties or formulae defined in LTL Template files (i.e. contains the specification of properties written in the special LTL language)	Uses process description properties (*CLASS_ASSERTIONS; OBJECT_PROPERTY_ASSERTIONS;* and *DATA_PROPERTY_ASSERTIONS)* defined using OWL and SWRL Language/schema.
GUI	There is option to select *concepts* for the parameter values	There is option to select *concepts* for the parameter values
Support	Supports *concepts* as a value (i.e when a concept is selected, the algorithm will test whether the attribute is an *instance of* that concept, and concepts can only be specified for set attributes).	Supports *concepts* as a value (i.e. when a concept is selected, the algorithm will test whether the attribute is an instance of that concept, and concepts can only be specified for set attributes).

Quantitative Evaluation and Analysis of the Semantic Fuzzy Miner

To assess performances of the Semantic-Fuzzy Mining Approach being able to correctly classify and analyse the individual traces within the models:

- The work refers to the results as recorded in Table 2 and the final outcome of the experimentation and cross-validation were carried out on other existing benchmark algorithms and techniques for process mining which includes namely:
  - Inductive Miner and Decomposition (Ghawi, 2016)
  - DrFurby Classifier (Verbeek & Mannhardt, 2016),
  - Heuristic Alpha+ Miner (Shteiner, et al., 2016)
  - Fuzzy-BPMN miner (Okoye, et al., 2016)

Evaluation Cont'nd…

The study utilize the standard Percent of Correct Classification (PCC) (Baati, et al., 2017) to assess the performance of the classifier. Henceforth, the standard Percent of Correct Classification for the test log is defined as follows:

Log_PCC = (number of correctly classified traces) / (total number of traces) x 100

For example, for *training_model_7* as previously shown in Table 1, the *Log_PCC* for the April test log for the initial result of the Fuzzy-BPMN miner is determined as follows:

Training_Model_7 (PCC) = (19) / (20) x 100

= 0.95 x 100

= 95%

Evaluation Cont'nd…

On the other hand, the *Log_PCC* for the training_model_7 as shown in Table 2 for the Semantic-Fuzzy miner approach is as follows:

Training_Model_7 (PCC) = (20) / (20) x 100

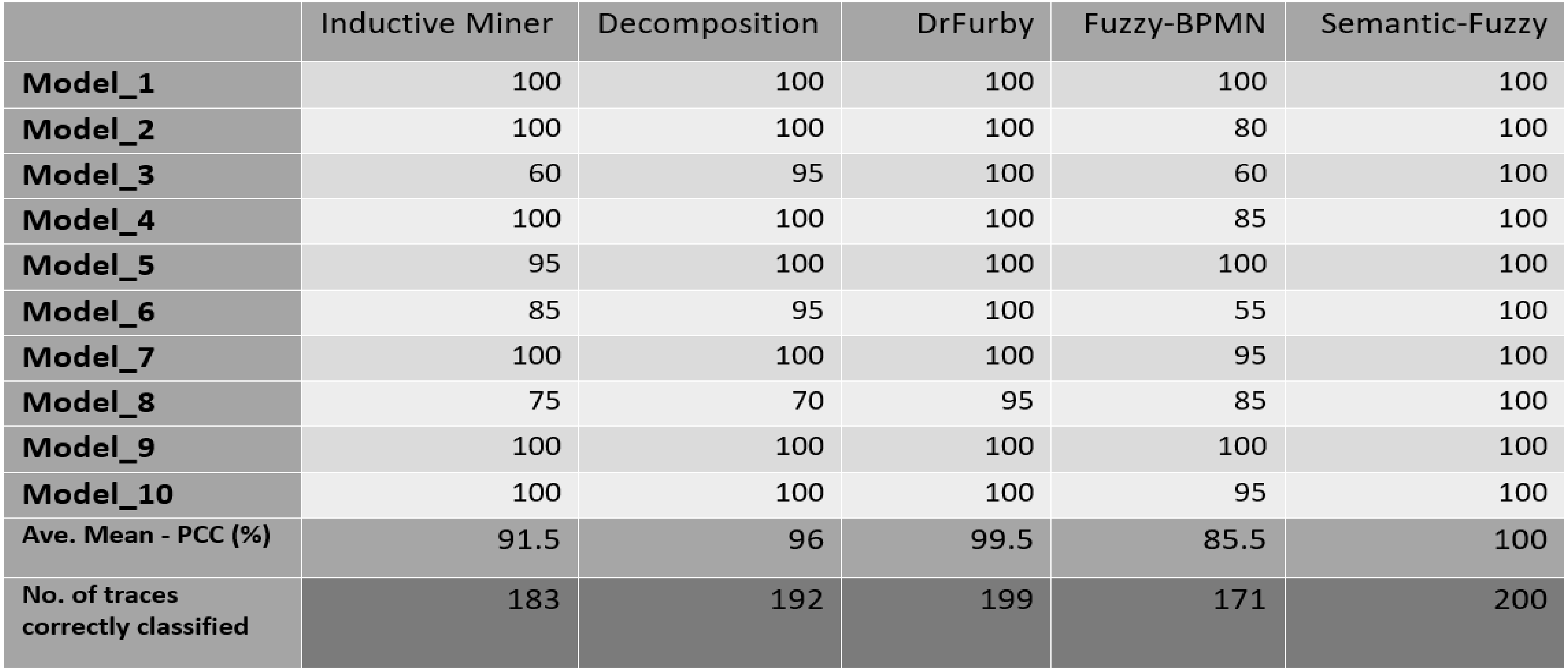
= 1 x 100

= 100%

Therefore, using the logical formula i.e. standard Percent of Correct Classification (PCC) (Baati, et al., 2017) the research measures and analyse in the following Table 3 the sophistication of the other existing benchmark algorithms including the initial result of the Fuzzy-BPMN miner to weigh up the proposed Semantic-Fuzzy mining approach and experimental results.

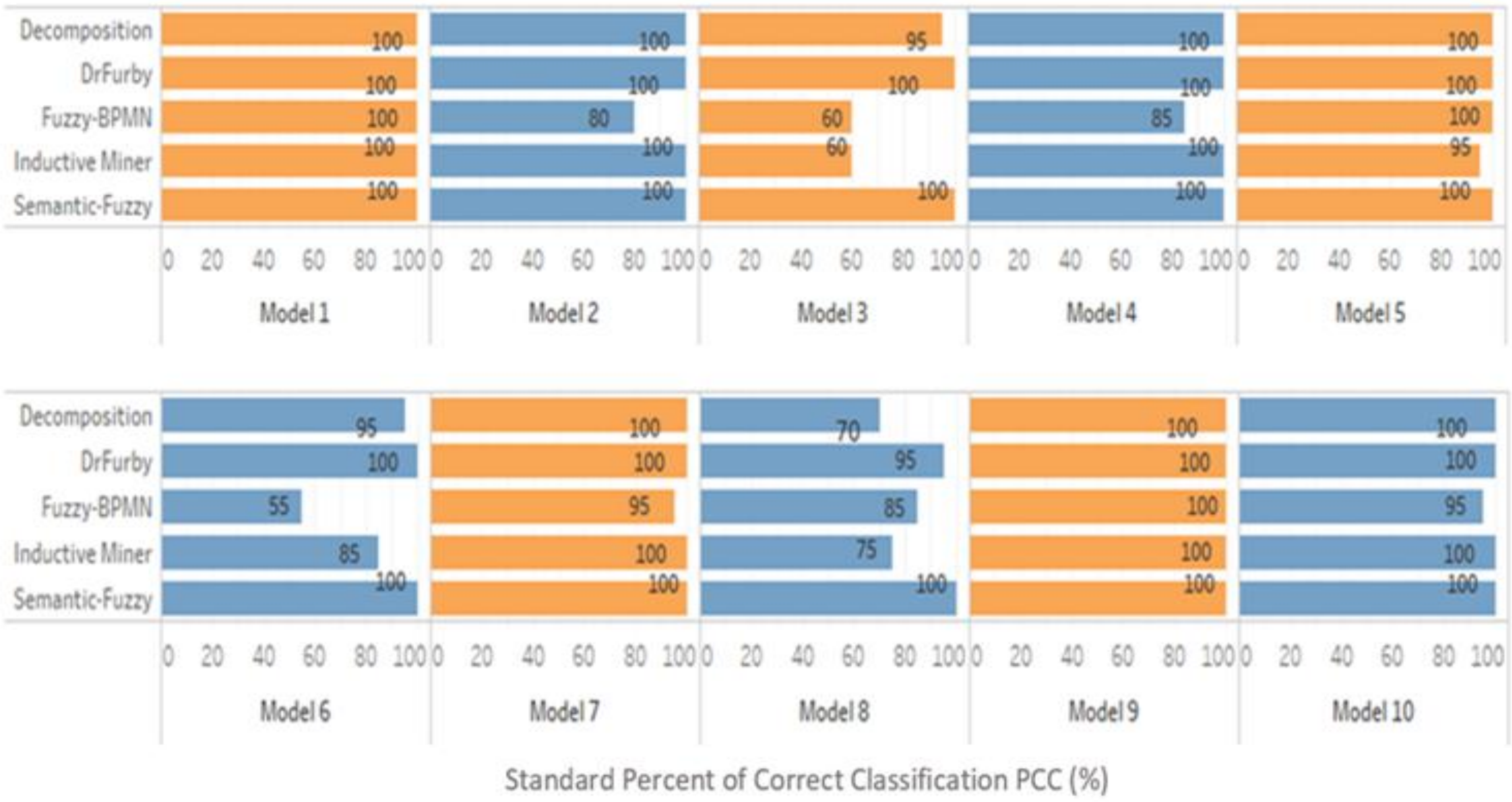
Evaluation Cont’nd…

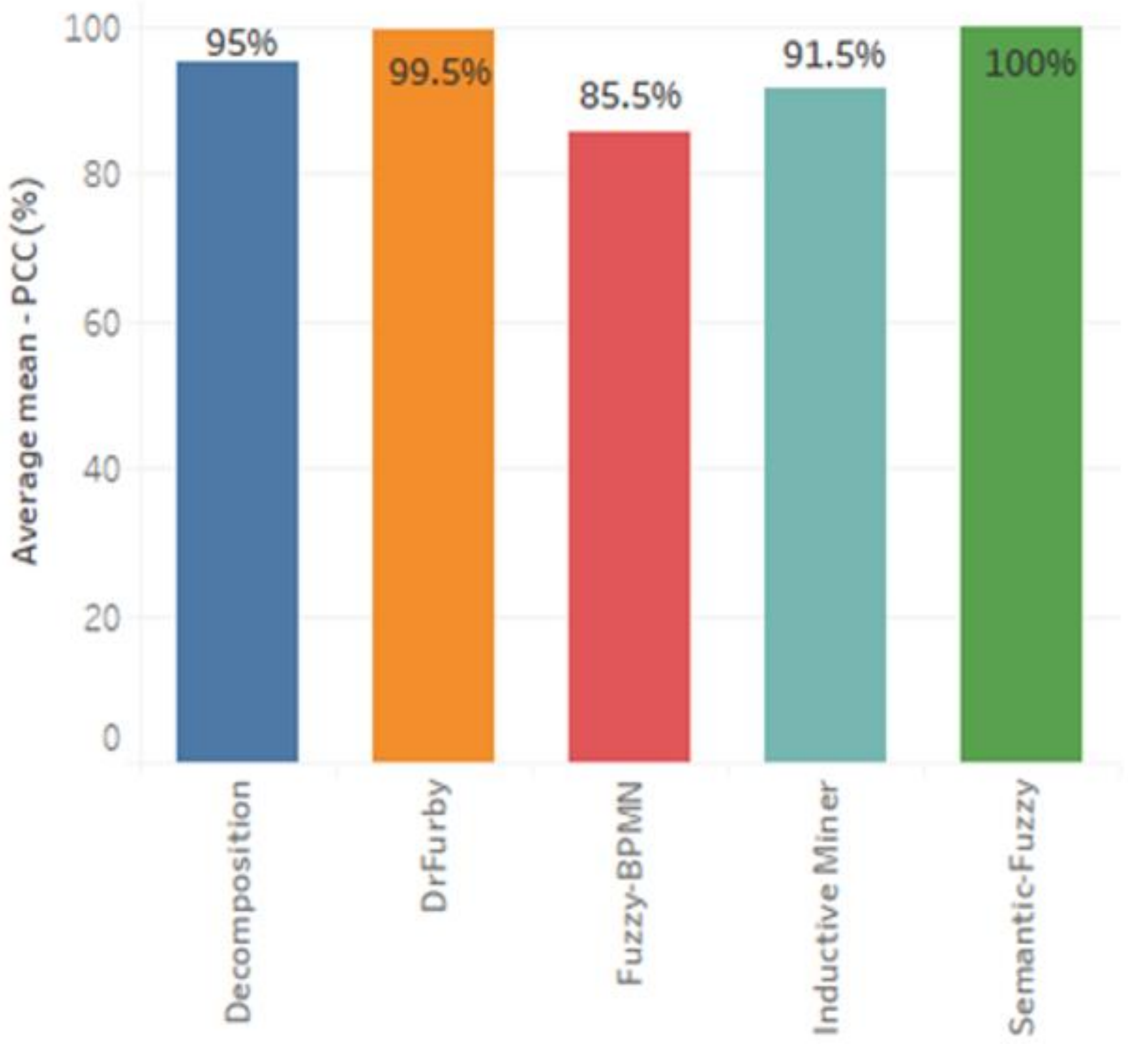
The outcome from the different benchmark techniques and the classification results are as shown in the following Table 3.



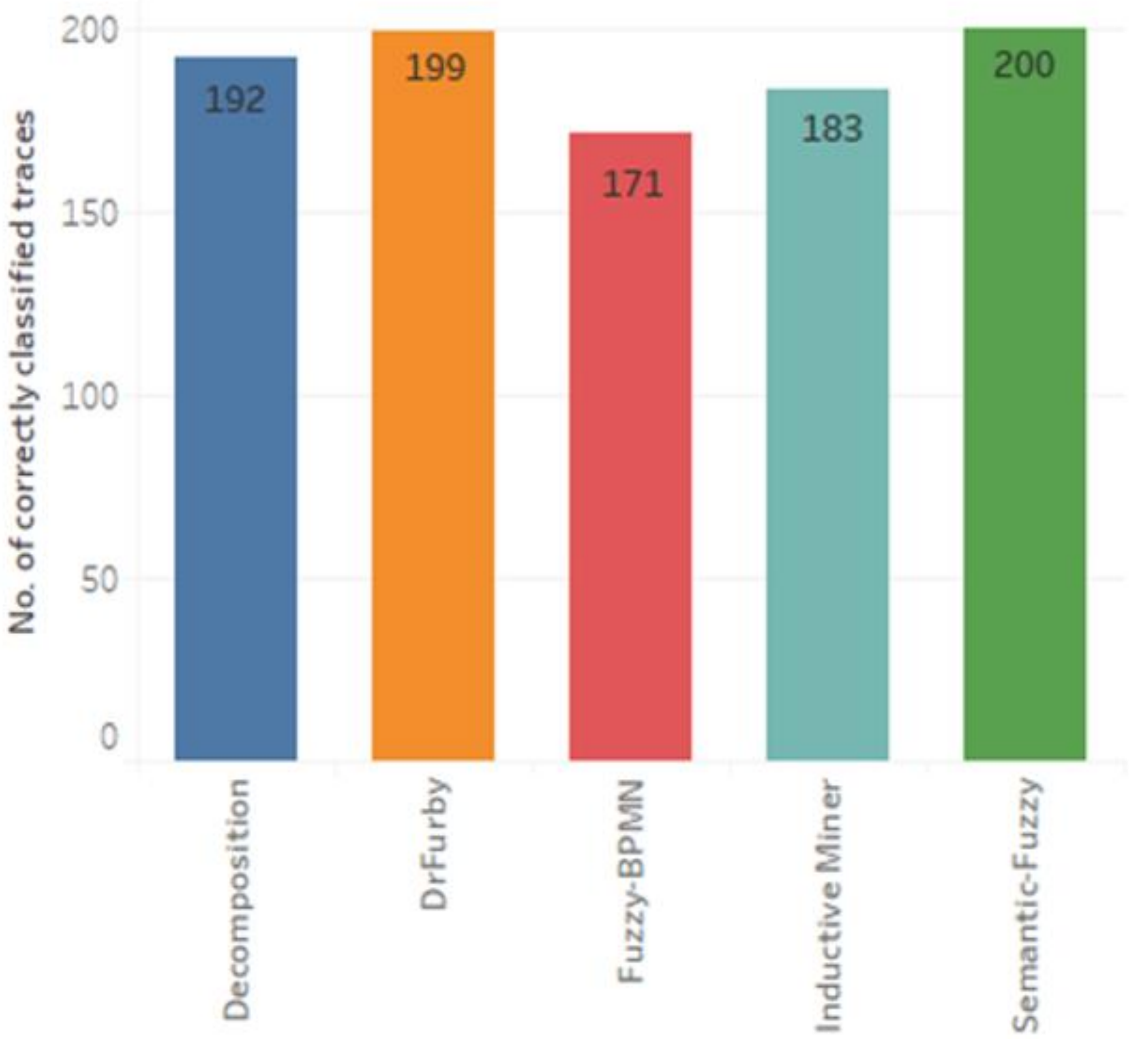
	Inductive Miner	Decomposition	DrFurby	Fuzzy-BPMN	Semantic-Fuzzy
Model_1	100	100	100	100	100
Model_2	100	100	100	80	100
Model_3	60	95	100	60	100
Model_4	100	100	100	85	100
Model_5	95	100	100	100	100
Model_6	85	95	100	55	100
Model_7	100	100	100	95	100
Model_8	75	70	95	85	100
Model_9	100	100	100	100	100
Model_10	100	100	100	95	100
Ave. Mean - PCC (%)	91.5	96	99.5	85.5	100
No. of traces correctly classified	183	192	199	171	200

Chart showing the sum of correctly classified traces by the various algorithms for each Model 1 to 10 - using the **standard Percent of Correct Classification PCC** (%).





Sum of Average mean - PCC (%) for each of the Algorithms



Total number of traces correctly classified by each algorithm

Evaluation Outcome and Conclusions

- Indeed, from the evaluation results in **Table 3**, and the plots in the **charts**: the study observe that the Semantic-Fuzzy miner considerably outperform respectively the Inductive miner and Fuzzy-BPMN miner, even though, the two algorithms Decomposition and DrFurby stands for the state of the art classifiers amongst the existing process mining techniques when compared to analysis of the classifications results and outcomes.

Performance Measurement and Indicator:

Classifier Name	**Formula**
tp-rate	*tp/p*
fp-rate	*fp/n*
Error	*(fp + fn)* / *N*
Accuracy	*(tp + tn)* / *N*
Precision	*tp/p’*
Recall	*tp/p*
F1 Score	(2 x *Precision* x *Recall*) / (*Precision* + *Recall*)

Performance measures formula for the Classifiers (Van der Aalst 2016)

Evaluation Outcome Cont'nd…

More so, the semantic-based approach has shown an error free performance indicator when measured using the classifier formula:

i.e. *Error = (fp + fn)/N)* where *fp* = 0 and *fn* = 0, thus, *Error* = (0 + 0) / 200 = 0.

In addition, the semantic fuzzy mining approach has shown a high level of accuracy through the classifier formula:

i.e. Accuracy = (tp + tn)/N) where *tp* = 100 and *tn* = 100, thus, *Accuracy* = (100 + 100) / 200 = 1.

Obviously, going by Accuracy & F1 Score = 1, and the error-rate =0, the Precision and Recall of the Semantic-Fuzzy miner classifications are indeed efficient.

Summary:

- The work in this thesis shows through the semantic fuzzy mining approach that by semantically annotating and encoding process models with rich semantics and the integration of semantic reasoning, that it is possible to specify useful domain semantics capable of bridging the semantic gap conveyed by the traditional process mining techniques.
- Henceforth, with the semantic-based process mining approach, useful information (i.e. semantics) about how activities depend on each other in a process domain is made possible, and essential for extracting models capable of creating new and valuable knowledge.

Summary Cont’nd…

- The main idea and lessons from the study - is that for any semantic-based process mining approach, these aspects of aggregating the task or computing the hierarchy of the process models should not only be machine-readable, but also machine-understandable.

- Besides, the unabridged notion of the proposed approach, design framework, algorithms and experimental results proves that semantic concepts (i.e. annotation, ontology, and reasoning) can be layered on top of existing information asset (i.e. process models, event data logs etc.) to provide a much more easy and accurate way of analysing real time processes capable of providing real world insights and answers that can be more easily grasp by the process owners, process analyst, system developers, software vendors etc.

Thus, the Research Hypotheses

The study claims and demonstrate that:

“It is possible to apply effective Reasoning Methods to make Inferences over a Process Knowledge-Base (e.g. the case study of the Learning Process) that leads to automated discovery of meaningful models, patterns or process behaviours”.

Research Publications: https://www.researchgate.net/profile/Kingsley_Okoye

Recent Survey on Educational Process Mining (EPM) Approaches @ 2017

Acknowledgements:

- **Dr. Syed Islam** – Senior Lecturer, School of Architecture Computing and Engineering, University of East London (*Director of Studies*)

- **Dr. Usman Naeem** – Senior Lecturer, School of Architecture Computing and Engineering, University of East London (*Supervisor*)

- **Dr Paolo Falcarin** – Reader in Computer Science at the University of East London (*Exam Chair*)

- **Dr Saeed Sharif** – Senior Lecturer, School of Architecture Computing and Engineering, University of East London (*Internal Examiner*)

- **Dr Islam Chowdhury** – Associate Professor in Computer Science at the Kingston University, UK (*External Examiner*)