International Research Journal of Pure and Applied Physics
Vol.4, No.3, pp.1-11, March 2016
Published by European Centre for Research Training and Development UK (www.eajournals.org)

A REPRESENTATIONAL FORMALISM FOR TRACKING CRIMINALS USING RESOURCE DESCRIPTION FRAMEWORK

Onyemauche U.C[1], Okonkwo O.R[1] and Nwosu Q.N[2] and Mbanusi, C.E[1]

[1]Department of Computer Science, Nnamdi Azikiwe University Awka, Anambra State, Nigeria.
[2]Department of Physical & Health Education, University Of Nigeria Nsukka, Enugu State, Nigeria.

ABSTRACT: *To effectively represent mass of events oriented data, automated methods for extracting event records and then classifying events and patterns of events into higher level terminology and vocabulary are necessary. Rich representation model and automated methods of correlating event information expressed in such models are becoming a necessity. The Resource Description Framework for Forensics (RDF) framework was developed with the strategic objective "to develop a means by which a consolidated repository of event information can be constituted and then queried in order to provide an investigator with post hoc event correlation.*

KEYWORDS: Framework, Correlation, Digital Evidence.

INTRODUCTION

RDF – the Resource Description Framework – is a foundation for processing metadata; it provides interoperability between applications that exchange machine-understandable information on the Web. RDF emphasizes facilities to enable automated processing of Web resources. RDF metadata can be used in a variety of application areas; for example: in *resource discovery* to provide better search engine capabilities; in *cataloging* for describing the content and content relationships available at a particular Web site, page, or digital library; by *intelligent software agents* to facilitate knowledge sharing and exchange; in *content rating*; in describing *collections* of pages that represent a single logical "document"; for describing *intellectual property rights* of Web pages, and in many others. RDF with *digital signatures* will be key to building the "Web of Trust" for electronic commerce, collaboration, and other applications.

Carrier (2006) introduced a model for representing RDF metadata and one syntax for expressing and transporting this metadata in a manner that maximizes the interoperability of independently developed web servers and clients. The syntax described in this document is best considered as a "serialization syntax" for the underlying RDF representation model. The serialization syntax is XML, XML being the W3C's work-in-progress to define a richer Web syntax for a variety of applications. RDF and XML are complementary; there will be alternate ways to represent the same RDF data model, some more suitable for direct human authoring. Future work may lead to including such alternatives in this document. The RDF data model is a syntax-independent way of representing RDF statements.

RDF statements that are syntactically very different could mean the same thing. This concept of equivalence in meaning is very important when performing queries, aggregation and a number of other tasks at which RDF is aimed. The equivalence is defined in a clean machine

ISSN 2055-009X(Print), ISSN 2055-0103(Online)

understandable way. Two pieces of RDF are equivalent if and only if their corresponding data model representations are the same.

Digital forensics has been defined as the use of scientifically derived and proven methods towards the preservation, collection, validation, identification, analysis, interpretation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal or helping to anticipate the unauthorized actions shown to be disruptive to planned operations. One important element of digital forensics is the credibility of the digital evidence and its representational format. Digital evidence includes computer evidence, digital audio, digital video, cell phones, digital fax machines to mention but a few.

Garfinkel (2010) uses resource description model meta data model to identify similar features across entire corpuses of drive, a technique which could prove useful for identifying computers with similar usage pattern. Finally, another useful form of classification is similarity.

RELATED RESEARCH

The RDF data model provides an abstract, conceptual framework for defining and using metadata. A concrete syntax is also needed for the purposes of authoring and exchanging this metadata. The syntax does not add to the model; APIs may be provided to manipulate RDF without reference to a concrete syntax. RDF uses the Extensible Markup Language (XML) encoding as its syntax. However, RDF will not require (and conforming implementations must not require) an XML Document Type Declaration for the contents of assertions.

Stallard et al (2010), employed an anomaly based expert systems approach to identifying semantic inconsistencies in investigation related data. Their approach translated MAC times generated by TCT and the UNIX last log into an XML representation, which was asserted into the HESS expert systems shell. Knowledge is encoded as heuristic rules which specify invariant conditions related to logins and potential file modifications.

Elsaesser et al (2006) employ an AI based approach to automated diagnosis of how an attacker might have compromised a system. Using a model of the topology of a network, the configuration of system, and a set of "action templates", a class of artificial reasoner called a "planner" generates hypothetical attack sequences which could have led to a particular situation. These hypothetical attack sequences are them run in a simulated environment, and the generated logs compared with the logs of the real world system. The action templates correspond to specifications of how a particular action will transit the state of the world from one state to the next.

Approaches to meta data correlation in the IDS and network management domains have focused on single domains of interest only, and have employed models of correlation that are very specific in nature. Repurposing these specific existing approaches to the more general task of event correlation in the CF domain is made difficult for a number of reasons. Existing event pattern languages do not necessarily generalize the application in wider domains. For example, while state machine based event pattern languages may work well for events related protocols, they do not work well with patterns where time and duration are uncertain. Most approaches focus exclusively on events, and ignore context related information such as

ISSN 2055-009X(Print), ISSN 2055-0103(Online)

environmental data and configuration information. Furthermore, few approaches have available implementations in a form that is readily modifiable.

Where we have modifiable implementations of RDF, we find that extension is complicated by the software paradigm underlying its implementation, and that the systems are weak on semantics.

Adding new vocabulary to the event language is slowed because of compilation and linkage overheads. Addition of concepts outside of the event pattern language require reengineering of the STATL LANGUAGE compiler and supporting framework.

Turner (2008) opined that the representation used to model events has a significant impact on the usability of correlation approaches, including conceptual expressiveness, extensibility, ease of integration of new information and maintainability. The MODEL language, a component of the DECS network management system, used an object oriented (OO) style model of classes of events related together in class/subclass relationship (which in this case was referred to as semantic generalization) . The event correlator translates from events patterns specified in the MODEL language directly to **C**++, and presumably, is encumbered by the maintainability characteristics of **C**++ software development and deployment. He further stated that Expert systems based approaches such as the EMERALD IDS combine a similar knowledge model, which support class/subclass models of events, with a rule language. The model however is dynamically constructed at run time, eliminating the **C**++ compile-link phase, resulting in simpler extensibility and more rapid evolution compared to the DECS approach.

A number of challenges were identified with the Resource Description Framework. The RDF data model intrinsically only supports binary relations. The approach does not incorporate notions such as semantic generalization in its modeling approach does not identify a methodology for mapping the detail, rich domain specific information contained in log files to the canonical form implied that every event was seen as a time-subject-object-action tuple (TSOA), a notion which proved to be an impediment when attempting to represent arbitrary event log entries. This canonical form was supplemented by the addition of shadow data an arbitrary set of name-value pairs which could be associated with a canonical entry.

RDF graph representation of the files shown in Listings 1 and 2.

ISSN 2055-009X(Print), ISSN 2055-0103(Online)

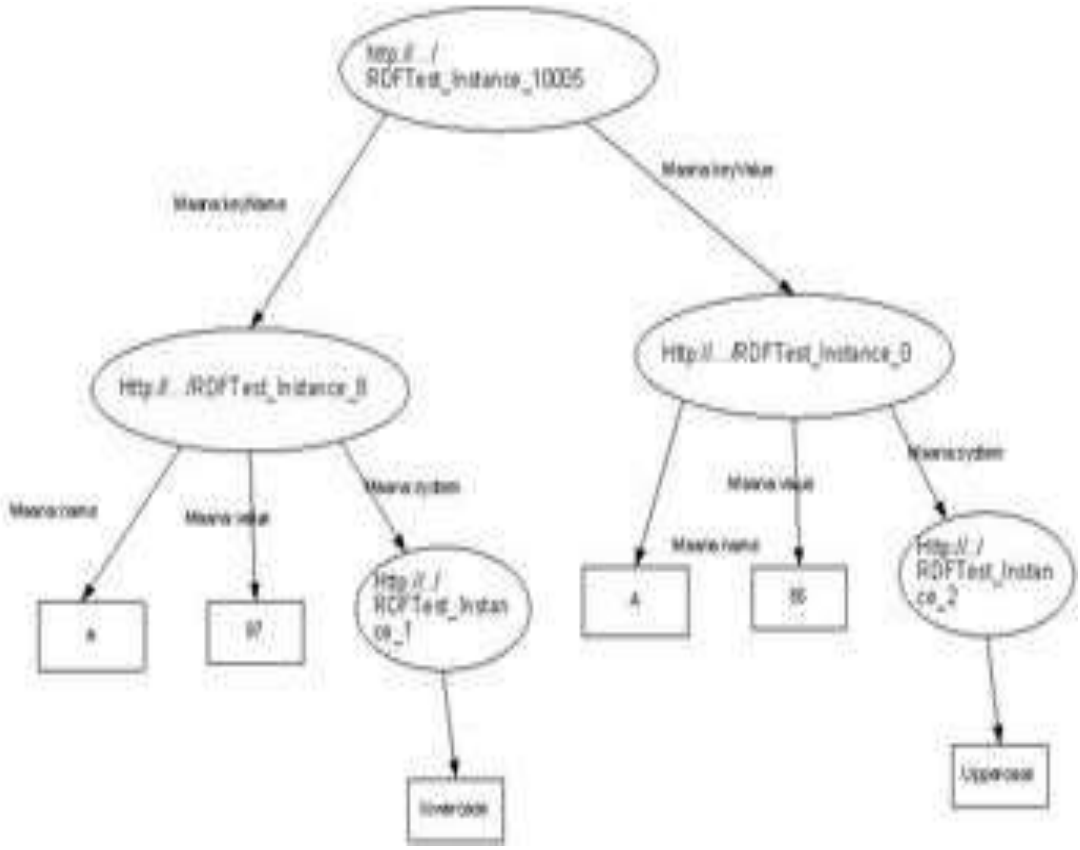


Figure 1. RDF graph representation for the sample RDF file. Click on thumbnail to view full-sized image.

An ellipse represents the resource, and a rectangle represents the literal. The resource (subject) is linked to another resource or literal (object or value) through an arc, or arrow, (predicate or property), which can be considered a *triple* and is called a statement.

The query below is an example RDQL query. The triple (**?x <http://www.vvasam.com/Maana#value> "97"**) in the query is a statement. The **x** is a bind variable that represents a resource; **http://www.vvasam.com/Maana#value** is a property with name **value**; and **97** is the value of the property.

ISSN 2055-009X(Print), ISSN 2055-0103(Online)

Knowledge Representation

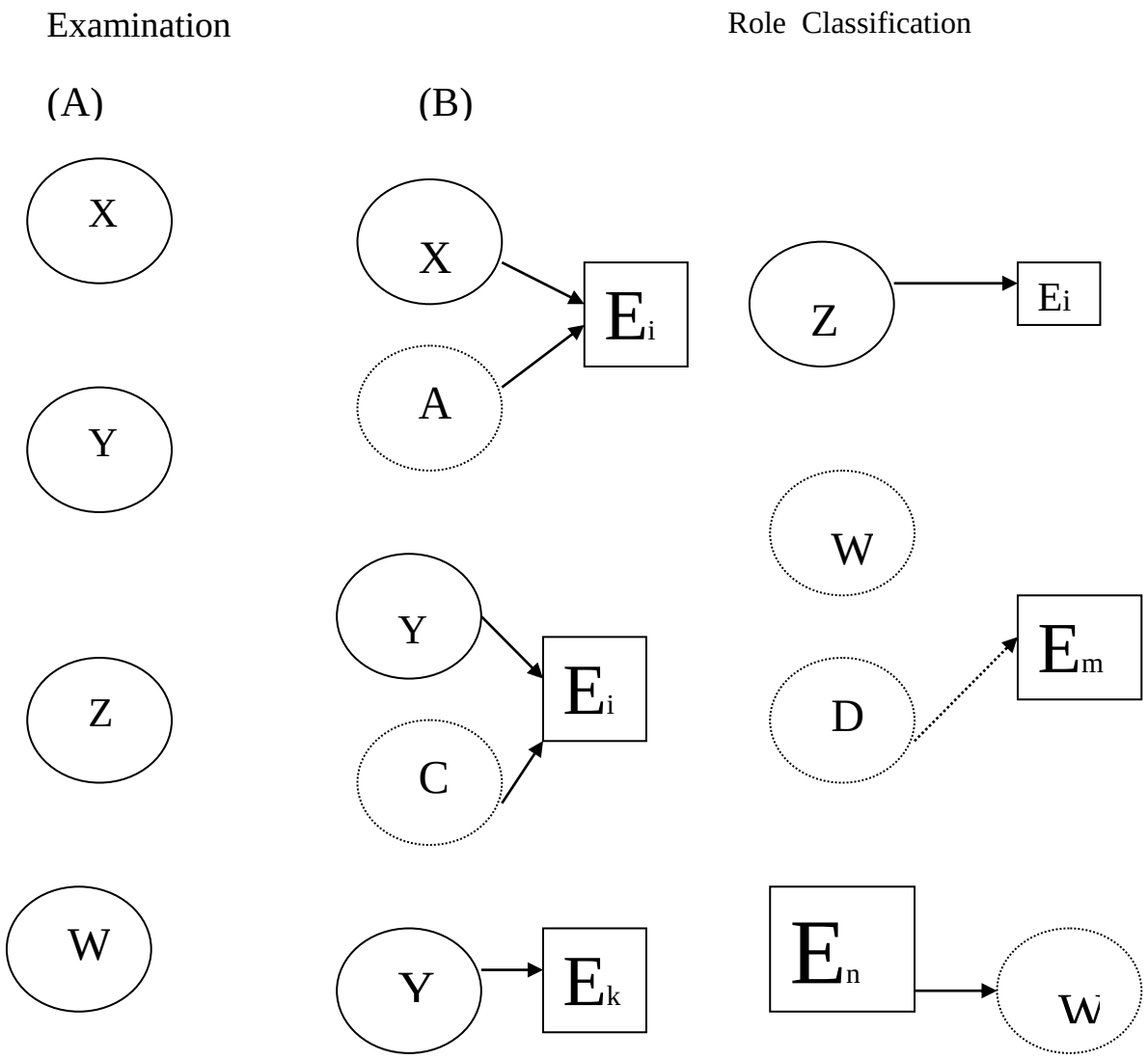


Fig. 2: Evidence Classification and Heuristics Rule Classification

Where X, Y,W, Z are Objects, all E's are events and A,C,D are Targets

After steps have been taken to preserve the state of the digital objects at the digital crime scene, the crime scene is searched for evidence. The goal of this phase is to recognize the digital objects such as X,Y,W and Z(digital computers) that may contain information about the incident. The first and foremost thing is to define a target that will be used to locate the evidence.For example, if you are looking for a file named foo.txt, then the target would have a name of foo.txt. If you are looking for a file with "bar" in the content, then the target would have "bar"in the content. Next is to extract data from the crime scene in some search pattern and then compare the extracted data with the target. After new evidence is found, updates of the general knowledge about the investigation so much recruited will be defined and reconstruction takes place giving birth to events (EK, Em, Ei).

ISSN 2055-009X(Print), ISSN 2055-0103(Online)

RESULTS

After all of the objects have been examined and their possible roles defined, event construction and testing groups the roles together to form events. Cause and effect roles are grouped together and if other objects must exist for the event to occur then they are searched for. The search may involve the objects that have been collected or it may involve a new search of the crime scene, if it is still available. After possible events have been constructed there may be objects that should exist, but could not be found. Hypotheses about the location of these objects are formulated.

RDFTest1.rdf

```
<?xml version='1.0' encoding='UTF-8'?>
<!DOCTYPE rdf:RDF [
  <!ENTITY rdf 'http://www.w3.org/1999/02/22-rdf-syntax-ns#'>
  <!ENTITY rdfs 'http://www.w3.org/TR/1999/PR-rdf-schema-19990303#'>
  <!ENTITY Maana 'http://www.vvasam.com/Maana#'>
]>
<rdf:RDF xmlns:rdf=""
  xmlns:Maana=""
  xmlns:rdfs="">
<Maana:ASCII rdf:about="RDFTest_Instance_0"
  Maana:Name="A"
  Maana:value="65"
  rdfs:label="A:65">
  <Maana:system rdf:resource="RDFTest_Instance_2"/>
</Maana:ASCII>
<Maana:System rdf:about="RDFTest_Instance_1"
  Maana:Name="lowercase"
  rdfs:label="lowercase"/>
<Maana:ASCII rdf:about="RDFTest_Instance_10000"
  Maana:Name="b"
  Maana:value="98"
  rdfs:label="b:98">
  <Maana:system rdf:resource="RDFTest_Instance_1"/>
</Maana:ASCII>
<Maana:ASCII rdf:about="RDFTest_Instance_10001"
  Maana:Name="B"
  Maana:value="66"
  rdfs:label="B:66">
  <Maana:system rdf:resource="RDFTest_Instance_2"/>
</Maana:ASCII>
<Maana:AscXRef rdf:about="RDFTest_Instance_10002"
  rdfs:label="b:98:B:66">
  <Maana:keyName rdf:resource="RDFTest_Instance_10000"/>
  <Maana:keyValue rdf:resource="RDFTest_Instance_10001"/>
</Maana:AscXRef>
<Maana:AscXRef rdf:about="RDFTest_Instance_10005"
  rdfs:label="a:97:A:65">
```


ISSN 2055-009X(Print), ISSN 2055-0103(Online)

```
  <Maana:keyValue rdf:resource="RDFTest_Instance_0"/>
  <Maana:keyName rdf:resource="RDFTest_Instance_8"/>
</Maana:AscXRef>
<Maana:System rdf:about="RDFTest_Instance_2"
  Maana:Name="uppercase"
  rdfs:label="uppercase"/>
<Maana:ASCII rdf:about="RDFTest_Instance_8"
  Maana:Name="a"
  Maana:value="97"
  rdfs:label="a:97">
  <Maana:system rdf:resource="RDFTest_Instance_1"/>
</Maana:ASCII>
</rdf:RDF>
```

Listing 2. RDFTest1.rdfs

```
<?xml version='1.0' encoding='UTF-8'?>
<!DOCTYPE rdf:RDF [
  <!ENTITY rdf 'http://www.w3.org/1999/02/22-rdf-syntax-ns#'>
  <!ENTITY system 'http://protege.stanford.edu/system#'>
  <!ENTITY Maana 'http://www.vvasam.com/Maana#'>
  <!ENTITY rdfs 'http://www.w3.org/TR/1999/PR-rdf-schema-19990303#'>
]>
<rdf:RDF xmlns:rdf=""
  xmlns:system=""
  xmlns:rdfs=""
  xmlns:Maana="">
<rdf:Property rdf:about="maxCardinality"
  rdfs:label="system:maxCardinality"/>
<rdf:Property rdf:about="minCardinality"
  rdfs:label="system:minCardinality"/>
<rdf:Property rdf:about="range"
  rdfs:label="system:range"/>
<rdfs:Class rdf:about="ASCII"
  rdfs:label="ASCII">
  <rdfs:subClassOf rdf:resource="Resource"/>
</rdfs:Class>
<rdfs:Class rdf:about="AscXRef"
  rdfs:label="AscXRef">
  <rdfs:subClassOf rdf:resource="Resource"/>
</rdfs:Class>
<rdf:Property rdf:about="Name"
  rdfs:label="Name">
  <rdfs:domain rdf:resource="ASCII"/>
  <rdfs:domain rdf:resource="System"/>
  <rdfs:range rdf:resource="Literal"/>
</rdf:Property>
<rdf:Property rdf:about="RDFTest_Slot_10003"
  rdfs:label="RDFTest_Slot_10003">
```


ISSN 2055-009X(Print), ISSN 2055-0103(Online)

```
  <rdfs:range rdf:resource="Literal"/>
</rdf:Property>
<rdfs:Class rdf:about="System"
  rdfs:label="System">
  <rdfs:subClassOf rdf:resource="Resource"/>
</rdfs:Class>
<rdf:Property rdf:about="keyName"
  rdfs:label="keyName">
  <rdfs:range rdf:resource="ASCII"/>
  <rdfs:domain rdf:resource="AscXRef"/>
</rdf:Property>
<rdf:Property rdf:about="keyValue"
  rdfs:label="keyValue">
  <rdfs:range rdf:resource="ASCII"/>
  <rdfs:domain rdf:resource="AscXRef"/>
</rdf:Property>
<rdf:Property rdf:about="system"
  rdfs:label="system">
  <rdfs:domain rdf:resource="ASCII"/>
  <rdfs:range rdf:resource="System"/>
</rdf:Property>
<rdf:Property rdf:about="value"
  rdfs:label="value">
  <rdfs:domain rdf:resource="ASCII"/>
  <rdfs:range rdf:resource="Literal"/>
</rdf:Property>
</rdf:RDF>
```


ISSN 2055-009X(Print), ISSN 2055-0103(Online)

Test Results

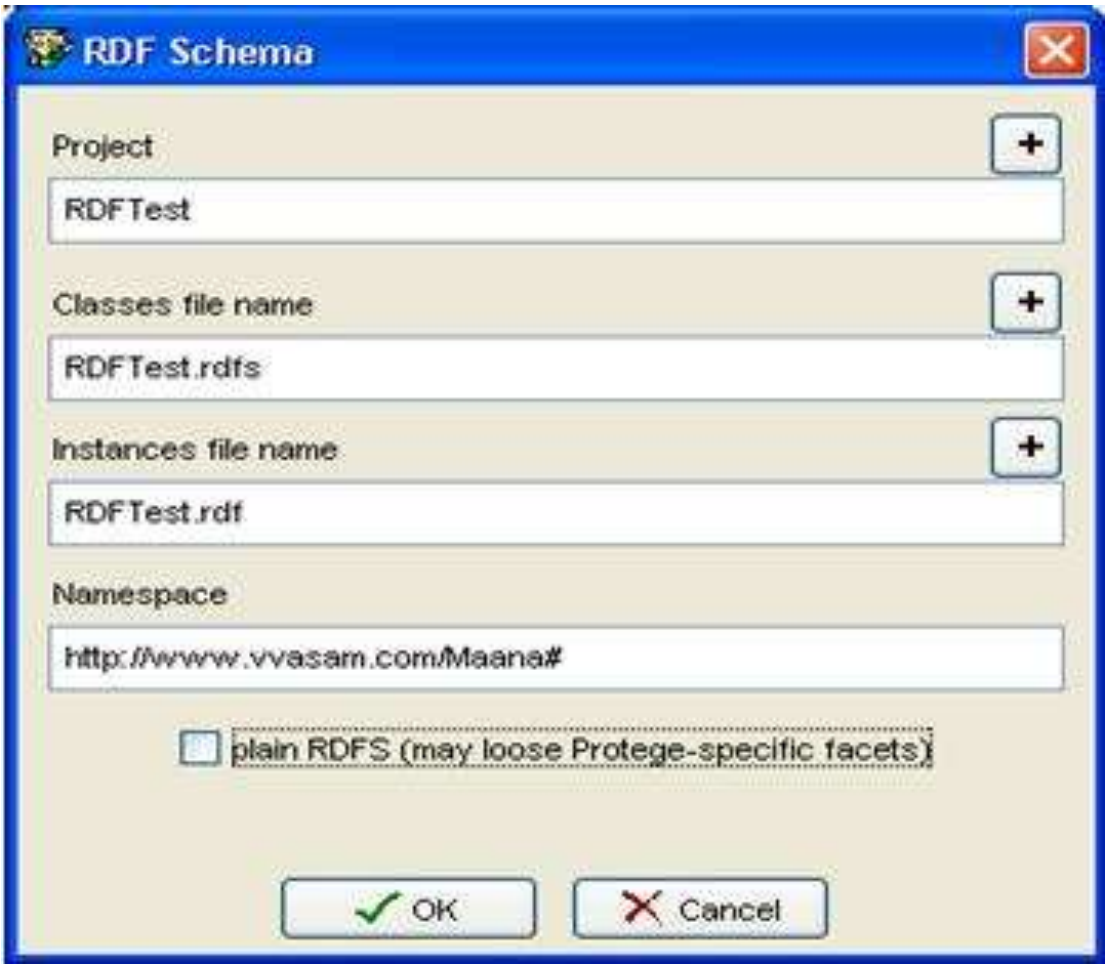
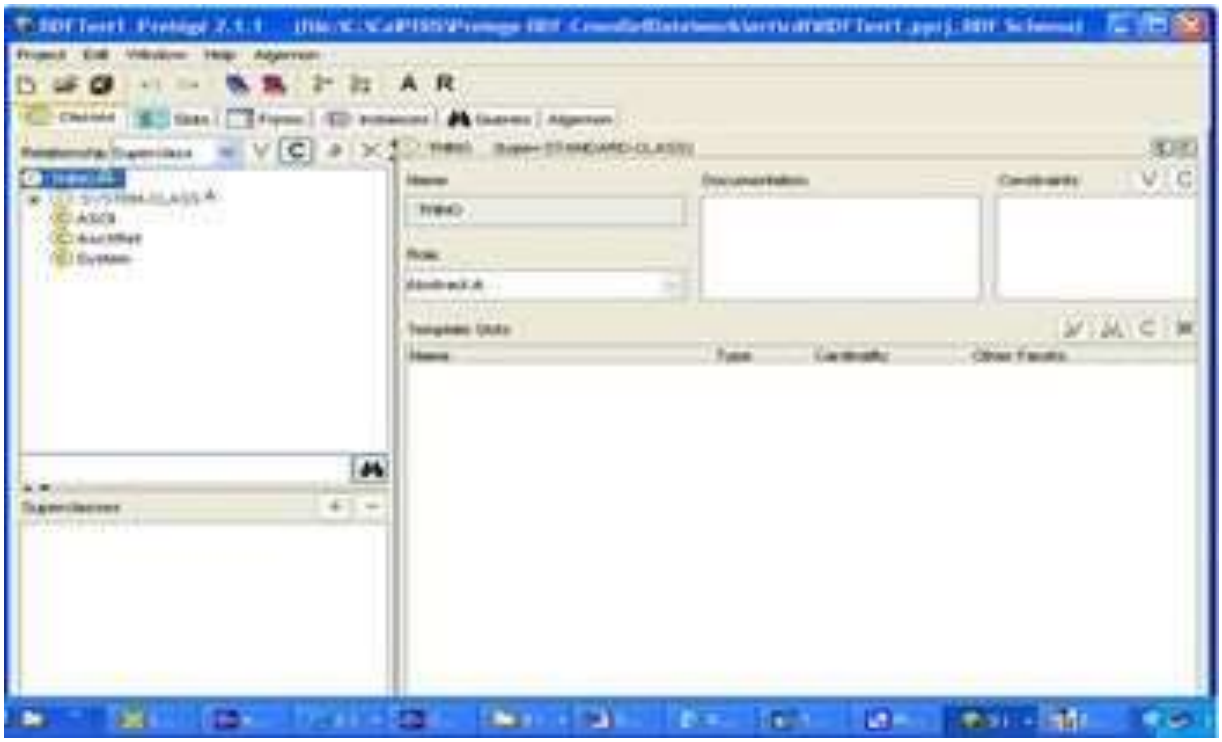


Figure 4. Prot g 's Save dialog

Figures 3 and 4 show the Prot g Classes and Instances tab, respectively, of the .rdf and .rdfs files shown in Listings 1 and 2. The files are created using Prot g RDF schema format.

ISSN 2055-009X(Print), ISSN 2055-0103(Online)



CONCLUSION

This paper proposes an abstract method of representing digital evidence using Resource Descriptive Framework. The digital forensic procedure is based on a new flow based specification methodology which uses XML (extensible Markup Language) serialization to integrate the meta data. It is shown through examples that the method can uniformly specify the forensic process in various phases and across roles. It also provides a more exact description where "things" (e.g., information, evidence) are separated into different streams of flow which finally will ensure information assurance.

Further research aims at experimenting with the method in real environments to build an information system to support stages of investigation.

REFERENCES

Carrier B.L.(2006). A cyber Forensics Ontology: Creating a new approach to studying cyber forensics. Proceedings in the 6th Digital Forensics Research Workshop. Lafayette, IN. pp 11-13.

Elsaesser, K.P, Gbendo, A.P.(2006). Dealing with Terabyte Datasets in Digital Investigations, Journal of *Research Advances in Digital Forensics*, Norwell: Springer, pp. 3-16.

Garfinkel, H, E.(2010). Unifying Computer Forensics modeling_approaches. A software engineering perspective. Proceedings in the 1st international workshop on systematic approaches to digital forensics engineering. Pp 10 -15.

ISSN 2055-009X(Print), ISSN 2055-0103(Online)

http://www.w3.org/TR/REC-xml
Stallard, L.P, Peter, M.K.(2010). Computer Profiling to Assist Computer Forensic Investigations, presented at RNSA Security Technology Conference, Canberra, pp. 34-30.
Turner, P.M.(2008) Unification of Digital Evidence from Disparate Sources (Digital Evidence Bags). in 5th Digital Forensics Research Workshop 2005. New Orleans. Pp 12 -16.
W3C. Extensible Markup Language (XML). 1998 [Viewed 7 Feb 2016]; Available from:

ISSN 2055-009X(Print), ISSN 2055-0103(Online)