# Information Security and the use of Ontologies during the process of Document Recovery

**Jonida Shehu[1], Endri Xhina[2] , [*] and Diana Shehu [3]**

[1]*University of Tirana, Faculty of Natural Sciences, Department of Informatics, Tirana, Albania*

[2]*University of Tirana, Faculty of Natural Sciences, Department of Informatics, Tirana, Albania*

[3]*Agricultural University of Tirana, Faculty of Economy and Agribusiness, Department of Management of Rural Tourism, Tirana, Albania*

## Abstract

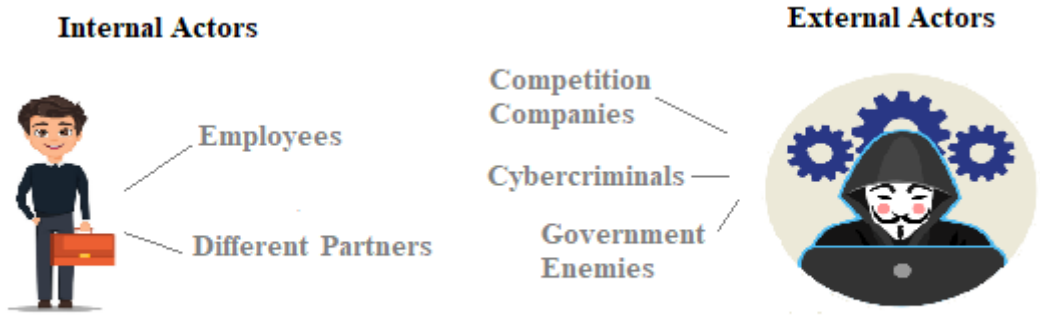
Information security is considered a matter of particular importance for information systems. Public administration, but also other e-government institutions possess essential documents related to their internal functioning or sensitive information about citizens. Attacks on information systems are diverse and use different methods of stealing, altering or destroying data. After the attack, the IT personnel responsible for the security of the systems should have some action plans on the reassessment of information, and the use of ontologies at this point could facilitate this process. If the data of an institution is structured using semantic web technologies, it is possible to analyze the missing information by performing several queries on the relevant ontology. Performing this search on the ontology provides the ability to reason on the content of information and logical connections that may exist between the changed or lost documents. In this paper, we identify and define the risks that threaten the information systems and discuss some of the attacks correlated to these risks. We focus on the risks that result in altering or losing information. Besides, we discuss semantic web technologies, their use in different application domains, especially the advantages of their use in the case of information security. Finally, we model an ontology in the domain of auditing, which includes information of document structure, auditors, and audits. With the semantic search on the ontology, we tend to discover corrupted documents by comparing them with the original information already stored in the ontology.

**Keywords:** information security, ontologies, semantic web technologies.

## 1. Introduction

Organizations, businesses, and government institutions invest in adapting the newest innovations of information technology. The employees of these institutions share data and services they utilize in their daily work. These data can be accessed everywhere on various devices such as desktops, laptops, mobile phones. Today's trial is to protect these data and services from external or internal actors who, for various reasons, attempt harm, fraud, or modification of information. It is often natural that businesses or organizations do not pay much attention to information security, assuming they will not be part of cyber-attacks, but as discussed by (Vacca, 2013), "businesses must understand that any network that is connected to the Internet is a potential target regardless of business type." Most of the attacks may have the intention of stealing personal information of employees such as passwords or credit card data for economic profits. Other types of attacks are performed just for enjoyment in order to demonstrate expertise. Social engineering is another worrying phenomenon that manipulates and deceives employees by making them believe that information transmitted by links or messages is secure. However, the company's employees, either intentionally or not, can cause loss or alteration of the information.

*Figure 1: Different actors in security attacks*




The responsibility of IT employees is to implement policies and strategies to protect the organization's information. The initial step is related to the identification of an attack, which in the case of passive attacks is very challenging to identify. In the next step, the company data must be protected, by dividing these data from the affected areas and resetting once again the rights of the authorizations. The last step is to identify the losses of the information stored in local documents or databases and applying methods to recover this information. In the process of this recovery, the knowledge represented through ontologies can be of great assistance. The primary purpose of ontologies is to model concepts and relations in order to reason on various concepts. We can perform queries on the ontology that answers questions like, i.e., which documents are related to documents affected by the attack. Alternatively, did the information on Audits changed? In this paper, we briefly discuss information security, different types of security attacks, and we propose an ontology in order to structure organization data and documents.

### 1.1 Security attacks

Employees of organizations transmit data and information by respecting the confidentiality and integrity of the data. However, as we discussed a bit earlier, this data sharing may often be vulnerable to attacks. As described by (Abomhara & Køien, 2015), attacks are actions taken to harm a system or disrupt normal operations by exploiting vulnerabilities using various techniques and tools.

Security attacks are of two types as explained in (Dhotre & Bagdad, 2009) and (Stallings, 2005):

- Passive Attacks
- Active Attacks

Passive attacks do not relate to modification of the system data; in this case, malicious people may, for example, monitor network activity or read employee emails to obtain information. Passive attacks consist of two types:

1. **Release of message contents** (In this type of attack, the malicious character can read the contents of an email or a document containing confidential information.)
2. **Traffic Analysis** (This type of attack is used to identify what information is being transmitted. Message patterns are monitored by the sender to the receiver).

In Active attacks malicious actors monitor and damage, modify or delete data, thus posing a risk of integrity and availability of data. There are four types of active attacks:

1. **Masquerade**-It is an attack when a malicious actor conceals his identity claiming to be another person.
2. **Replay-** It is an attack in which data is captured and replayed to the recipient.
3. **Modification of Message**- Modification of messages means that some portion of a legitimate message is altered, or those messages are delayed or reordered, to produce an unauthorized effect.
4. **Denial of Service**- The denial of service prevents or inhibits the normal use or management of communications facilities.

## 2. Literature Review

There are various researches on using ontologies for security support. The development of an ontology in the domain of information security has appeared very early on (RaskJn et al., 2001) by adapting existing concepts and adding new concepts to ontology, a particular focus has been adjusting the lexical entry senses as well as acquiring new entries in the English lexicon.

In (Almut et al., 2007), authors discuss the ambiguity in the domain of security; thus, such ambiguities could be mitigated by a common repository of domain knowledge. An ontology representing concepts of information security can be used as a dictionary in this domain by offering terms and links on threats, measures taken to protect against attacks.

In paper (Pereira & Santos, 2009) propose an ontology-based approach to firm up and unify the concepts and terminology in the security information domain, based on the relevant ISO/IEC_JTC1 standards. In this ontology are modeled some fundamental concepts for

information security such as threat, attack, impact, vulnerability, control, impact, and relations that exist between them.

When an attack on information systems occurs, it is difficult to identify the cause of the assault, the type of attack, or the actor, whether it is internal or external, resulting in the need for an automated solution. This solution can be achieved by building tools based on semantic web technologies that will enable cybercrime analysis and security measures. This approach is presented in (Bromander et al., 2016) when as explained: "A semantic model of threats will enable security analysts to work faster and more efficiently in terms of identifying threat agents and take advantage of previous experience and gathered intelligence when handling incidents caused by known or unknown threat agents."

## 3. Case Study and Results

Most of the ontology-based solutions and models in the above section based their research on the concepts of information security and the identification of various attacks. The ontology that we implemented relies on the structure of information stored in an audit institution, and we tend to offer analysis for the identification of the stolen or lost documents after an attack.

In the definition of (Bench-Capon, 2012), ontologies usually refers to the representations and descriptions of domain knowledge, by describing the types of objects found in the domain, the attributes which these objects may have, the relationships which these objects may enter into, the values that the attributes may have for particular types and not excluding also the axioms constraining these concepts.

The ontology graph in Figure 2 exported from Protégé editor, presents some fundamental concepts:

- ***Audit*** - The class of audits conducted by the institution.
- ***Auditor*** - The class of persons conducting audits. This class has two subclasses internal auditors and external auditors.
- ***TypesOfAudit*** - The class that holds information about audit types such as financial, performance, compliance, and IT.
- ***StagesOfAudit*** - The class that contains information about the stages of auditing, some representative individuals would be: planning, implementation, reporting, and archiving.
- ***Documents*** - The class will retain information on internal audit documents and external documents sent to the audited business.
- ***InternalDocuments***- The class of internal documents of the institution usually has more than 20 documents for an audit, but for demonstration purposes, we have modeled a few of them. This class contains subclasses: InformationOfConducted Audit, StructuredTasks, and WorkingPapers.

The queries performed in our ontology will assist in information analysis in cases where attacks can cause loss or change in the information. In order to perform SPARQL queries, we use Jena Fuseki SPARQL endpoint.

*Figure 2. Audit Ontology Graph*

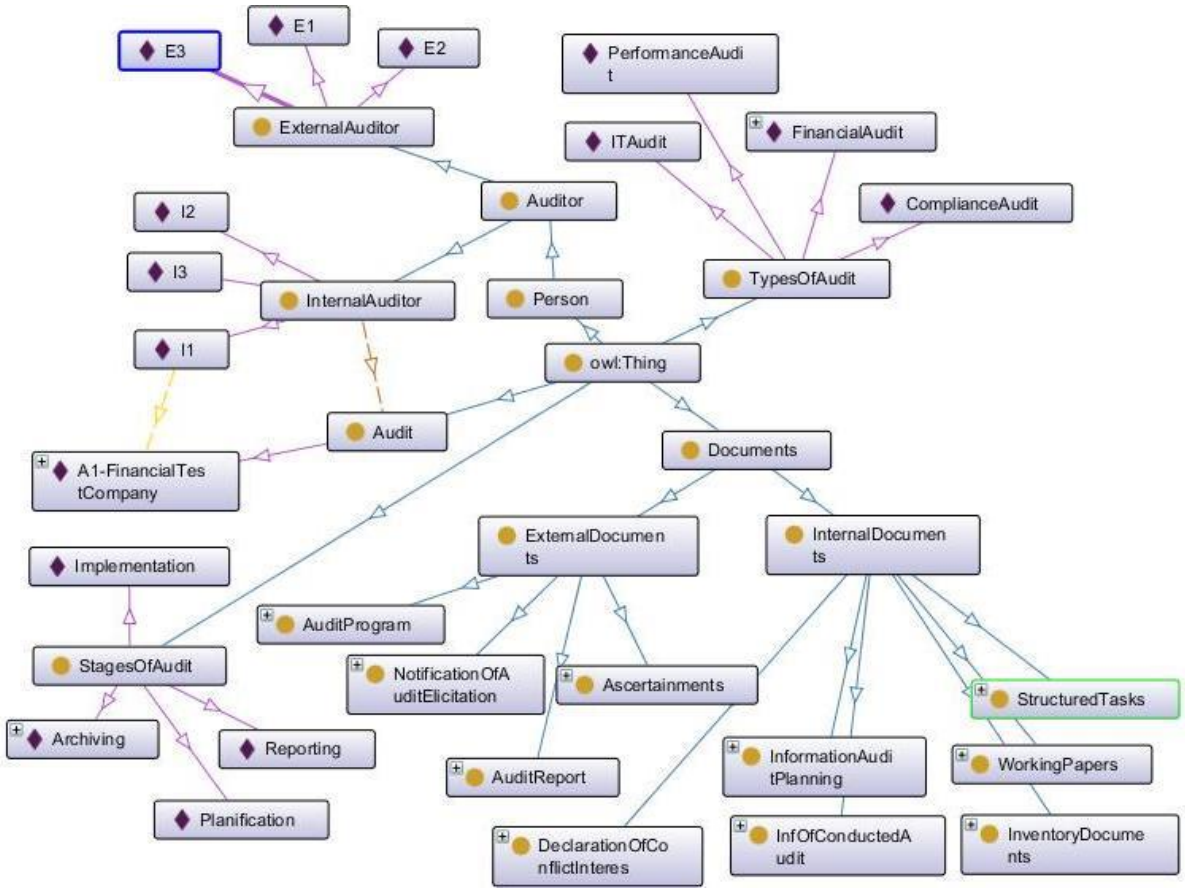



*Figure 3. Query to find the documents of Audit "A1-FinancialTestCompany"*

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX aud: <http://www.semanticweb.org/jonida-pc/ontologies/2019/4/audit#>

SELECT ?documents ?audit
 WHERE { ?audit rdf:type aud:Audit . ?audit aud:hasDoc ?documents
  FILTER ( ?audit IN (aud:A1-FinancialTestCompany))
}
```

The query presented in Figure 3, lists all the documents and the audit they do belong, here we filter only the documents of the audit ***"A1-FinancialTestCompany"***. The class Audit is associated with Documents with the object property***: hasDoc***, so we are confident that every individual of type document will be listed if it is related to this property and if it belongs to ***"A1-FinancialTestCompany"*** audit. The results in Figure 4 are illustrated from Jena Fuseki Interface. In other queries, we can explore the auditors or apply more filters about internal/external documents. Also, we can select other data properties connected to instances such as the date of creation or Document content.

*Figure 4. Results of A1-FinancialTestCompany audit and their documents*

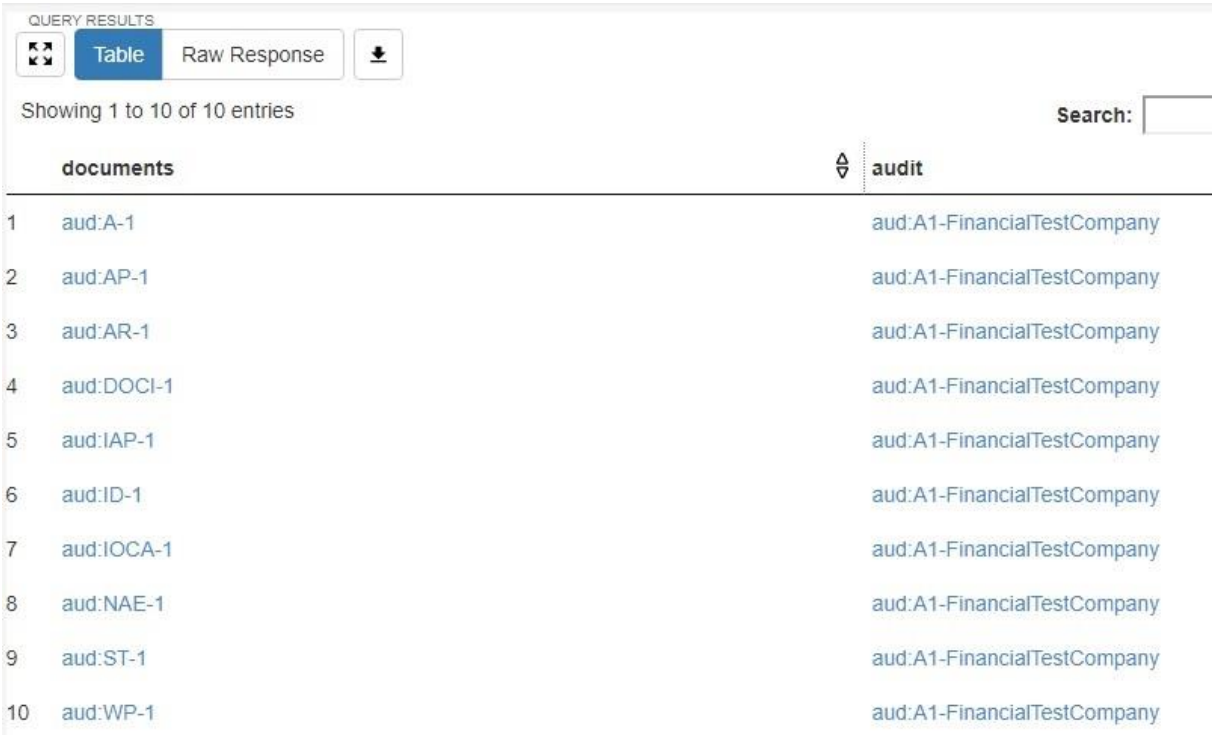

QUERY RESULTS

Table | Raw Response

Showing 1 to 10 of 10 entries

Search:

| | documents | audit |
|---|---|---|
| 1 | aud:A-1 | aud:A1-FinancialTestCompany |
| 2 | aud:AP-1 | aud:A1-FinancialTestCompany |
| 3 | aud:AR-1 | aud:A1-FinancialTestCompany |
| 4 | aud:DOCI-1 | aud:A1-FinancialTestCompany |
| 5 | aud:IAP-1 | aud:A1-FinancialTestCompany |
| 6 | aud:ID-1 | aud:A1-FinancialTestCompany |
| 7 | aud:IOCA-1 | aud:A1-FinancialTestCompany |
| 8 | aud:NAE-1 | aud:A1-FinancialTestCompany |
| 9 | aud:ST-1 | aud:A1-FinancialTestCompany |
| 10 | aud:WP-1 | aud:A1-FinancialTestCompany |

## 4. Conclusions

Attacks on information systems are frequent and of several types. The purpose of government institutions, businesses, or other organizations is to design appropriate strategies for protecting the data collected in these systems. Using ontologies as a way to store the information and logical connections that exist in it, is familiar to many application areas. Through a case study in which we modeled an ontology for an audit institution, we showed that it could reveal information on lost or modified audit documents, revised auditors' data, or deleted and many other aspects that may be related to the scope of application of this ontology. The advantages offered by ontologies compared to other modeling technologies such as UML or ER model is that it offers the opportunity to reason and deduce new knowledge from an existing one.

## References

[1] Vacca, J. R., (2013). *Book Computer and Information Security, Second Edition*, Morgan Kaufmann Publishers.

[2] Abomhara, M. and Køien, G. M. (2015). *Cyber Security and the Internet of Things: Vulnerabilities, Threats, Intruders and Attacks, Journal of Cyber Security and Mobility* Journal of Cyber Security, vol 4, pp. 65–88.

[3] Dhotre, I.A. and Bagad, V.S. (2009). *Information Security, First Edition*, Technical Publications.

[4] Almut, H., Shahmehri, N. and Duma, C. (2007). *An Ontology of Information Security*. *International Journal of Information Security and Privacy*. pp. 1-23.

[5] Pereira, T. and Santos, H. (2009). An Ontology Based Approach to Information Security, *Proceedings of Metadata and Semantic Research Conference*, Milan, Italy, pp. 183-192.

[6] RaskJn, V., Hempelmann, Ch. F., Triezenberg, K. E. and Nirenburg, S. (2001). *Ontology in Information Security: A Useful Theoretical Foundation and Methodological Tool,* NSPW.

[7] Stallings, W. (2005). *Cryptography and Network Security Principles and Practices, Fourth Edition*, Prentice Hall.

[8] Bromander, S., Jøsang, A. and Eian, M. (2016). *Semantic Cyberthreat Modelling, STIDS.*

[9] Bench-Capon, T. (2012). *Ontologies in AI and Law*.