Ontology-Based Framework Applied to Money Laundering Investigations

Gonçalo Carnaz[1], Vitor Nogueira[1] and Mário Antunes[2,3]

[1] Informatics Department, University of Évora
d34707@alunos.uevora.pt, vbn@di.uevora.pt
[2] School of Technology and Management, Polytechnic Institute of Leiria
mario.antunes@ipleiria.pt
[3] Center for Research in Advanced Computing Systems (CRACS), INESC-TEC;
University of Porto mantunes@dcc.fc.up.pt

Abstract. Criminal investigations face a deluge of structured and unstructured data obtained from heterogeneous sources like forensic reports or wiretap transcriptions. In these cases, finding relevant information can be a complex task. Ontologies have been successfully applied to several domains including legal, cybercrime and digital forensics. In this paper[4] it is proposed a framework based on ontology engineering, that provides an unified approach to represent and reason with the criminal investigation data. Moreover, this framework is applied to the specific use case of money laundering.

Keywords: ontology, knowledge representation, criminal investigation

1 Introduction and Motivation

Over the years, with the massive introduction of Information and Communications Technology (ICT) in different professional areas, where users leave their digital fingerprint, bringing new challenges to computer scientists. Crime, however reprehensible, is an activity that uses ICT as a tool of crime, or as proof of it. Criminal polices in general are currently facing new challenges, namely the huge amount of data produced during their investigations, resulting from heterogeneous sources. Concealment, hiding, dissimulation, economic system, illicit origin of assets are terms commonly used in the context of money laundering related crimes. These crimes can also be associated with other crimes, such as drug trafficking.

Computer science has a wide set of tools that may be used to automate analysis and correlation of investigations documents. Some of those tools include frameworks to retrieve data and to represent knowledge, as well as to collect

[4] This paper is for the assessment of Doctoral Seminar 1.

evidences and provides decision making during investigations.

Data is acquired daily as the occurrences and evidences take place. Depending on the type of crime, the types of sources can be challenged, regarding the origin of the documents, like paper, digital reports, handwritten transcripts of interrogations, social networks transcript messages and forensic logs, just to mention a few.

The contribution of this paper is thus to answer the following research question: *How to design and represent the information inherent to documents collected in money laundering investigations?*

Some questions, arises from research question above:

- It is possible to design a knowledge base, based on an ontology that represents the knowledge inherent to money laundering crimes in Portuguese Legal System;
- how can we detect patterns in the knowledge base that lead to money laundering schema's;
- how to find relevant data;
- how to deal with evidences sources that may support knowledge base;
- which visualization format will support users questions.

The remaining sections of this paper are organized as follows: section 2 depicts money laundering crimes and stages; different approaches to digital evidences analysis are described in section 3; in section 4.2 we describe approaches related with ontologies. In section 5 we present a framework to deal with money laundering. Finally, in section 6, we discuss the conclusions and delineate the future work.

2 Money Laundering Crimes

The best definition that can be enumerated for Criminal Investigation is described in the Criminal Investigation Organization Law, art. 1 of Portuguese Law 49/2008 of August 27 [13], and is defined by the *"set of measures that, under the terms of criminal procedural law , are intended to inquire the crime existence, to determine its agents and its responsibility, to discover and collect evidence in the course of the proceedings..."* [13].

In money laundering there are several definitions, all of which have in common the following main terms: concealment, dissimulation, economic system, illicit origin of assets. Thus, it is a process of cover-up or dissimulation through operations, supported by the economic/financial system, as a result of the large amount of money arising from illicit or criminal practices [7]. Basically, money laundering is based on a process of legitimate concealment of goods, products or

capital so that, at the end of the process, they have the appearance of legibility. Money laundering is supported by a process, called the "three-step model", implemented by the Financial Action Task Force (FATF) [36], described next:

- **Placing:** *"consists of the introduction of goods, products or capital that are to be laundered into the economic-financial system, using the most diverse means or instruments"* [7];
- **Circulation:** *"it will imply a set of procedures that provoke great rotation of ownership of the goods, with a view to the widest possible distance between their origin and the way of obtaining them, and the one that will eventually remain in their possession."* [7];
- **Integration:** *"is constituted by the integration of assets and/or values in the patrimonial sphere of the criminal to whom the values are due. It is completed when the illicit goods or values appear with the appearance of licit and are used freely by the criminal, Ahead of everyone, often even with high social consideration."* [7]

Returning to the definition stated above, we must identify three distinct objectives:

- the existence of a crime;
- their agents and their responsibilities;
- collect evidence, establishing the relationship between the act and its author.

Furthermore, any criminal investigation, including money laundering, should answer the following questions, the purpose of any investigation: Who?, Where?, When?, How? or Why?. In addition, researchers should look for patterns that may lead to the detection of criminal activity.

3 Digital Evidences Frameworks

The literature on frameworks for digital evidences analysis shows a variety of approaches, from academic to industrial ones. The following paragraphs will give us a selected work discussion focus in the academic approach. In [45] POLESTAR is described as a framework for knowledge management and collaboration tool for analysts, providing a framework from text documents and creating a documents repository to analysis. One of the main framework features is anomaly detection, alerting experts if any anomaly is detected in data retrieved from text documents. In [46], authors define an architecture that abstracts digital evidence, retrieving those evidences from multiple sources. They also realize that past reconstruction would be a requirement for the system, which facilitates the investigators' theories. The authors in [41] define a framework that crawls into Web blogs looking for relevant information. A three layer infrastructure is defined to support crawling and to store information for further analysis.

In [1] authors propose central repositories of integrated data from one or more heterogeneous sources concepts into a framework supported by a 5-steps interactive process: 1) Data identification; 2) Business Data Model Design; 3) Data Warehouse Model Design; 4) Testing and Analysing and 5) Data Marts Models Design. Supported by these 5 steps, they designed a relational tool that analyses crime activity, also organized police reports into a data warehouse, named by "police logs".

In [27] authors focuses their work in social criminal networks understanding, designing a framework that fetches data retrieved from Web and documents, as a result, they design criminal networks, detect crime hot stops and profiling criminal steps.

The authors [32] analyse criminal networks based on communication logs, they have used a interactive process for criminal network construction from smartphone call logs, based into phases: First one, data are clean by expert officers and data engineers; second, added metrics by social networks experts and finally a analysis is performed over the network supported by machine learning algorithms. This method is supported, initially, by human intervention and for learning, algorithms are applied to retrieve knowledge. A different approach made by [18], focus in criminal network visualization supported by mobile calls logs reconstruction. Every day, police officers produce text documents related to crime investigations and victims reports, the paper [51] authors developed a visual analytical tool that identifies entities on those documents and visualise them in multiple views (coordinated).

In [9] a system to predict survival techniques after a terrorist attack is described, by crawling into twitter, analysing the propagation of re-tweeting to map the necessity of survival, as a mechanism of defense. In [38], based on use scenario of Point-of-Sale (POS) Skimming, authors proposed a semantic framework to structured knowledge related to financial crimes.

Finally, there are industrial approaches, such as Analyst Notebook[5], Xanalysis Link Explorer [6] and Palantir [7] that may be consider to review.

4 Overview of Ontologies

Historically, the term "ontology" has its roots in two Greek words: "ontos", being, and "logos", word. Being the original word "category", applied by Aristotle in the sense of classification. Aristotle developed a list of categories that served as the basis for classifying any entity, dividing reality into entities: (i) individual substances and (ii) their qualities.

[5] http://www-03.ibm.com/software/products/pt/analysts-notebook
[6] http://www.xanalys.com/products/link-explorer/link-explorer-analysis/
[7] https://www.palantir.com/

From a philosophical point of view, the Oxford Dictionary of Philosophy, ontology is defined as: *"[...] the term derived from the Greek word for 'being', most used since the seventeenth century to refer to Branch of metaphysics that concerns what exists"*. Gruber defined: *"An ontology is an explicit specification of a conceptualization"* [24]. In this ontology, definitions associate names of entities in the universe of discourse (eg, classes, relations, functions, etc. with texts that describe what the names mean and the formal axioms that restrict the interpretation and use of those terms) [2]. From Gruber's definition, the term conceptualization as emerged, which corresponds to objects, concepts properties and other entities, that can be represented in several domains of knowledge. Therefore, conceptualization can be interpreted as abstraction, to represent the world in a simplified way. In 1997, Borst [6], defines ontology as: *"[...] Ontologies are defined as the formal specification of a shared conceptualization"* [6], while Gruber [24] defines ontology as: *"[...] An ontology is an explicit specification of a conceptualization [...] "* [24]. In computer science, ontologies have been developed in artificial intelligence in order to facilitate the sharing and reuse of information.Therefore, ontologies are applied to a wide range of computer science applications, and a significant contribution to the representation of the concepts, relationships and properties associated with the knowledge acquired in the different domains, so there are different areas of ontology using it, from Knowledge management [15] to the medical area [50].

4.1 Cybercrime and Forensic Ontologies

This section is an overview of related works that focus the cybercrime and forensic ontologies. In [54] proposed a dynamic and real-time forensic model based on ontologies and context information, where model is based on the authentication method that supports user's authentication, depending on the context. Therefore, authors added an ontology that describes the entities, authorizations and rules involved. Any police investigation process needs proofs, in [44] DCoDeOn ontology is defined, that allows preserving the cybercrime evidence, the so-called Chain of Custody[8], the authors defined a taxonomy diagram[9] that allows Chain of Custody representation. In [52] developed an ontology that seeks to represent knowledge in computer forensics field, namely:

- Digital forensic domain representation;
- Disciplines: software forensics, network, computer, database, multimedia and devices;

[8] in legal and police contexts, refers to the chronological documentation, showing analysis and disposition of physical or electronic evidence.

[9] a classification and naming in a ordered system that indicates relationships, in a form of a diagram.

- Sub-disciplines, such as: operating systems and applications forensics, or mobile forensics;
- Objects, such as forensic objects, i.e. web-services or authentication services;
- Sub-objects, i.e. access control systems with their logs.

In [43] authors proposed an ontology to supports safe operations in cyberspace. In this work, besides demonstrating the concepts related to the domain, they added the human factor as an important part of this technological field. They created the CRATELO ontology:

- Top level: the DOLCE SPRAY ontology allows the natural language understanding, capturing the primitive concepts inherent to the language;
- Middle level: the SECCO ontology defines security concepts in cyberspace;
- Lowest level: the OSCO ontology represents operations in cyberspace.

In [33] authors describe an approach to solve one of the main objectives of computer forensics: cause-effect in the acquisition of digital evidence.Therefore, authors developed a platform based on an ontology consisting in two layers: hardware and software.

- The hardware sub-layer: representing the digital equipment used in the investigations;
- The software sub-layer: representing forensic analysis tools and operating systems.

The authors [22] created an ontology that represents the culture around cyberspace security, and the relations between different entities. The knowledge base of this ontology has resulted in the information acquired about the culture on cyberspace, based on campaigns in the communities (users). The use of social networks is not limited to recreational or professional purposes, but also to criminal activities. Therefore, platforms like Facebook, YouTube, Linkedin, etc. are used by criminal investigation entities as data sources for analysis (future or even in real time), for the detection and proof of crime. The ontology SC-Ont [29] was proposed to support the criminal domain and its relations, based on social networks. The smartphone, one of the most used devices for communication. Therefore, the F-DOS ontology [34] allows the abstraction between the user and the data collected by smartphones. This ontology consists of:

- A core ontology, where the essential concepts of the domain are presented;
- Other domain ontologies: contacts, messages and research.

There is a growing concern to represent, analyze and process data collected in criminal investigations. Authors [12] proposed an ontology that seeks to answer essential questions in the presentation of evidence in Court House: what, who, when, where, why and how.

The WikiCrimes platform [20] [21] allows the collaborative use based on maps manipulation, in order to register the criminal movements. The architecture is based on two ontologies: Crime and Reputation. In [44] designed an ontology applied to criminal investigation in cyberspace, with the categories of cybercrime, laws, evidence and information of suspects. In these domains there is the difficulty to relate the type of crimes and the collection of evidence, with this ontology, the authors try to represent this correlation and thus to detect the associated crimes and the evidences evidenced. Based on a Semantic Web framework, [16] presented the problems inherent in the integration and correlation of digital evidence, trying to present the steps necessary for the representation, aggregation and integration of this digital evidence in an ontology. In [39] described an event-based ontology for cybercrime, defined the events and their relationships using 6-tuples[10]: Action, Participant, Time, Location, Instrument and Good. He divided the relationships between classified and non-classified. Starting from an example of online banking fraud, they proposed the OBM ontology [11] to map out criminal organizations and identify malware developers. Finally, they also defined rules of inference based on empirical knowledge that would meet some of the needs of the forensic analyst.
From the wide spectrum of ontologies describe below, from cybercrime to forensics domains, can helps us implementing ontologies regarding forensics evidences knowledge representation.

4.2 Legal Ontologies

This section is an overview of related works that focus the legal ontologies. In [3] proposed an ontology as a support for the representation of crime and/or criminal activity in Italy, aims to be an attempt to solve some problems found in ongoing projects that were not based on ontologies and that did not have a conceptual definition of a knowledge base in order to achieve a conceptual framework for the various projects, added a domain knowledge also draw the classes that allow the ontological representation of the concept of crime, they defined a suspect/criminal - a person who acts in a manner punished by criminal law, with a given behavior in a given time interval - Event, and the penalty applied to the perpetrated act. It will thus support the management of documents as metadata, identify and suggest a crime hypothesis to the Judge, and semantically map criminal laws using the XML[11] language. Authors [48] developed the integration of different ontologies, different domains, representing the heterogeneous data gathered in the different information and communication technologies, in order to solve the lack of specialization of some researchers in the domains in

[10] a ordered list of elements
[11] eXtensible Markup Language

question, leading to the creation of the FORE ontology. On the other hand, in [8]proposed two basic ontologies applied to the legal domain:

- FOLaw is based on legal knowledge and seeks to represent this same knowledge;
- LRI - CORE supports the construction of structured legal domains, to allow automatic indexing of legal texts.

LKIF ontology [26] emerged as part of an architecture for information systems in the legal domain. The ontology has two requirements:

- Translation between legal knowledge base represented in different formats and formalisms;
- Formal representation as part of an information system architecture.

This ontologies are represented diverse legal / juridical knowledge, such as: documents, norms, laws. Another important requirement was the attention to the different levels of knowledge of the users. From papers related below, a sorted of ontologies were developed to support knowledge from the legal domain, that we will take in account for our framework, in case we need. Thus, legal domain differ from one country to another, we have to adapt our ontology to country legal system.

Money Laundering In [30], authors define the steps for an ontology, named FF POIROT, which represents knowledge in the field of financial crime. In [47] tries to represent a suspicious financial transactions through an expert system, based on an ontology and a set of rules. Following the rules of design applied to ontologies, the authors created a set of classes, objects and properties that represent the transactions to be processed by the expert system. Additionally, a set of rules, using SWRL (Semantic Web Rules Language), in order to infer new knowledge through existing knowledge. In [37], the authors define an ontology that can map the knowledge inherent money laundering , constructing an ontology that can help discover money laundering schemes. They are defined entities: people, organizations, portfolio and messages and other auxiliary classes, objects and properties. In [4] presented a proposal, ontology and rules, that allow to represent the crime of money laundering, called "minimal model". With this representation, the authors intend to discover, through rules, the different roles of the actors, and their level of relationship (for the use case, it is extremely important, establish this relationship and its level). In addition, relationships between companies are also established, in order to prove relationships: entities, people and actions. In [38] the authors designed a tool supported by an ontology, that tries to represent the semantic information extracted in the forensic investigations. The ontology is supported by three levels: "Abstract Knowledge

Layer" represents the knowledge of the experts; "Knowledge Processing Layer" supports forensic knowledge;"Concrete Knowledge Layer" represents the data extracted and stored in digital format.

5 Ontology-Based Framework for Money Laundering

The methodology proposed is to transform the informal and unstructured data retrieved in heterogeneous police data sources, such as police reports, into a structured knowledge. In the proposed framework we integrate the different kind of data sources, and with that, we can correlate data to help police investigations, like detecting different stages of money laundering done by different entities, for example: smurfing schema [35], all this supported by a defined ontology that will represent all knowledge associated to domain. The framework is represented in Figure 1. It is important to mention here that the framework can collect data from data-sources in Portuguese language.

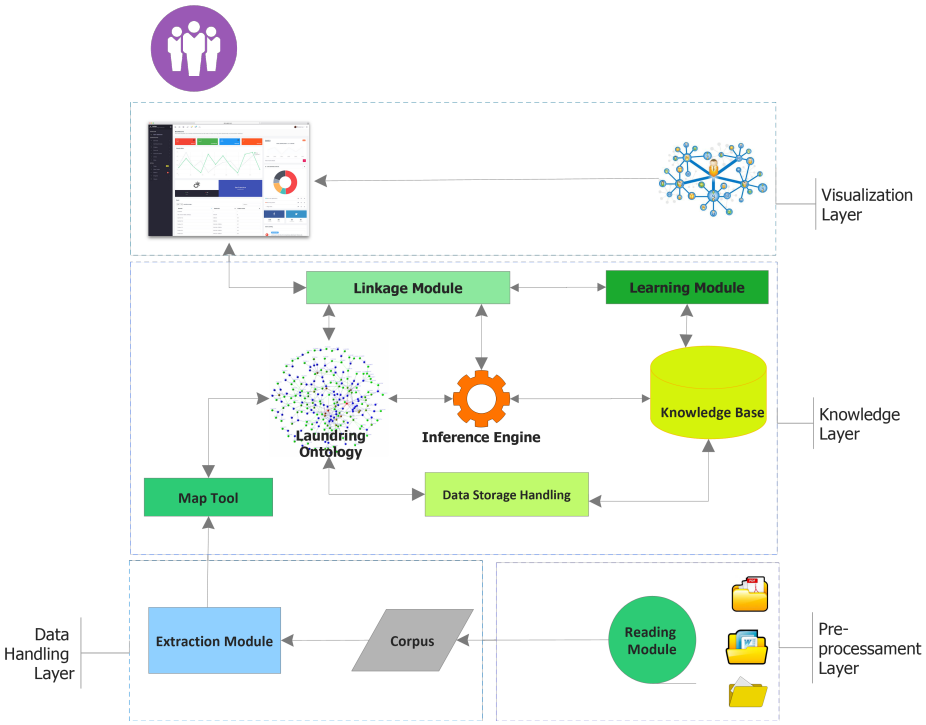

Fig. 1. Ontology-based framework for money laundering schema

5.1 Pre-processing Layer

Police repositories store millions of reports pages on crimes, offenders, and other intelligence. Thus, Pre-Processing layer define the data sources to collect, analyze and process all unstructured data. Currently, all police departments dedicated to money laundering produce reports, with different data types: text or numbers, and different formats: spreadsheets, text documents or forensic logs. There are some challenges associated with this layer:

- Deal with different data sources formats and types;
- How to deal with the asynchronous feeding, because all police cases are continuous updating evidences;

Also, this is a big data issue, framework must deal with the four big data dimensions: volume, variety, velocity and veracity [28]. To solve the enumerated challenges, a reading module will be added to framework, that will systematically browses documents, to retrieve and cleaning data. This module fits in batch processing definition, because documents represent chunks [28], that can be processed in parallel. Police reports are retrieved into plain text, forming the text corpus. This process is necessary for cleaning text purposes to exclude any noise from that, such as images or videos.

5.2 Data Handling Layer

This layer supports corpus analysis that was created from previous layer, there is a main challenge associated with this layer:

- Extract entities and relations in Portuguese language sources;

This is done by the Extraction Module. Using natural language processing (NLP) to retrieve information, such as entities and relations from text corpus. Done in two phases:

- Pre processing phase: remove unnecessary data, in order to organize the information extracted so that classification will be simpler and more efficient. For this, we use NLP techniques, such as tokenization, lower case, stopwords removal or stemming [23].
- Feature and classification phase: named entity recognition [40] techniques will be performed to classify entities and relations.

There are some studies [14] [49] [10] [31] to support natural language process understanding, in English, and also in Portuguese [42] [19]. Therefore, this layer will retrieve and identify entities and relations, analysing each sentence trying to extract context meaning on each word, in Portuguese language.

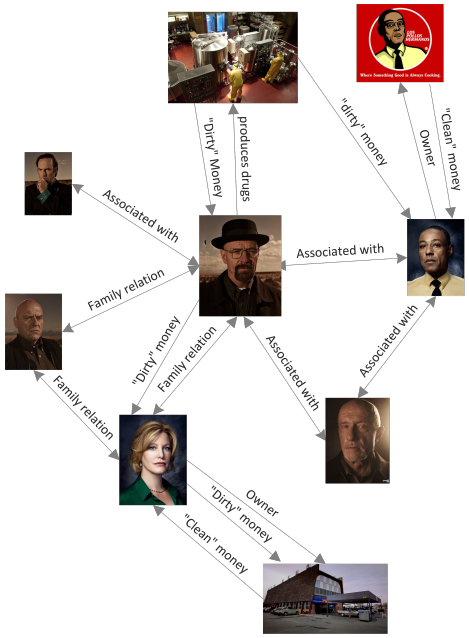

Fig. 2. Use Case - Breaking Bad

5.3 Knowledge Layer

The knowledge layer main objective is to define an ontology that allows the knowledge representation. In order to size an ontology that may support the knowledge inherent to the crime of money laundering, a use case was taken from the well-known American TV series - Breaking Bad [12]. In the Fig 2: **'Walter White'** that is associated with an **'Organization'**, maintaining family relationships, with **'brother-in-law'** and **'wife'**, delivers **'sums of money'** to the **'wife'**, who makes them go through the **'car wash'**, that she is the **'owner'**, thus clearing the money, which comes from **'drug traffic'**. Associated with, there are external signs of wealth, such as the purchase of a **'high-powered car'**, by **'Walter White'**. Therefore, from observation of use case, we need to dimension a ontology that supports knowledge inherent to the domain, representing actors, objects, actions, time and other relevant information. There are some features, that we can implement or improve from previous works:

[12] http://www.imdb.com/title/tt0903747/

- LKIF [26] ontology is an example applied to legal domains, representing legal/juridical knowledge, that can be suitable to our work, with a thesaurus[13] enriching our ontology;
- how time is represented [3];
- answer essential questions in the presentation of evidence in Court House [20].

In order to transform the collected data, commonly expressed is heterogeneous data formats, into their semantic representation, and match the extracted data to the instances, a Map Tool is defined to perform this task. Therefore, the ontology must map all entities and relations, also should allow the instantiation all data into ontology schema.

The ontology must be based on entities, agents and activities, something that crimes are also based on, as we can see in use case above. As well as the possibility of adding temporal and spatial properties, giving the possibility to draw a events timeline [17]. Summing up, ontology must reflect domain terminologies, entities, events, actors and relations between each others, with event time related. A inference engine will be added to define rules and perform inference queries against defined ontology, there are some related studies [25] [53] that may support our implementation. A data storage handling will be used to manage the CRUD[14] operations that instantiate data within the semantic representation model to a permanent database, with the semantic model as data schema, this database must reflect the heterogeneous environment.

The linkage module will support connections with different modules, acting as a bridge, between: visualization and knowledge layers, with laundering ontology, inference engine and the learning module.

Finally, learning module will support machine learning algorithms to learn from previous data and give a substantiated hypotheses to police activities, like a recommendation system, related to money laundering.

5.4 Visualization layer

This layer aims to knowledge visualization, how user will interact with extracted knowledge and visualises graph links and patterns. One way of content visualization is displaying it as a graph, because highlights patterns, and show clusters and connections, tools like [5]. Therefore, this block must give a visual analysis tool to:

- understand data that we are collecting ;
- understand relations between data elements;
- perform queries to knowledge base and visualize them;
- visualize networks of crime, based on data relations created in layers below.

[13] "lists words grouped together according to similarity of meaning" in Wikipedia - https://en.wikipedia.org/wiki/Thesaurus

[14] Create, Read, Update, Delete

5.5 Tools

The deployment of the framework benefits from using a wide set of tools in the various topics involved. The following list describes theses tools:

- Tika (`https://tika.apache.org/`) - detect and extract metadata and text from different sources.
- GATE (`https://gate.ac.uk/`) - retrieve data from text corpus using NLP.
- Protege (`protege.stanford.edu/`) - create, map and management of ontologies.
- CounchDB (`couchdb.apache.org/`) and Neo4j (`https://neo4j.com`) - database creation and management.
- Jena (`https://jena.apache.org`) - to engine inference.
- Gephi (`https://gephi.org/`) - for graph visualization.

6 Conclusion

In this paper we have provided an overview of different approaches for knowledge representation using ontologies. We made a comprehensive study of the literature and identified several challenges regarding the use of ontologies applied to the development of frameworks for crime investigation. That is, the amount of data related with criminal investigation, coming from distinct sources, are challenging computer science research to deploy and develop ontology-frameworks to identify and correlate terms and subjects related with a specific kind of crimes: money laundry.

We have proposed a framework based on a ontology to support knowledge representation related to money laundering. Since Portuguese is the default language that brought us new challenges regarding lexical and semantic features. The framework is composed by several components, organized in four layers: Pre-processing, data handling, knowledge and visualization. Based on previous research and domain requirements related to money laundering analysis, we present our initial design for the framework, from evidences retrieval, crossing knowledge representation and visualisation.

Future work consists on developing this framework and test it with real world scenarios in money laundering.

References

1. F. Albertetti and K. Stoffel. From Police Reports to Datamarts: Towards a Crime Analysis Framework. In *Proceedings of the 5th International Workshop, IWCF 2012, Tsukuba, Japan*, pages 48–59, 2012.

2. M. B. Almeida and M. P. Bax. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. *Ciência da Informação*, 32(3):7–20, 2003.
3. C. Asaro, M. A. Biasiotti, P. Guidotti, M. Papini, M.-T. Sagri, D. Tiscornia, and L. Court. A Domain Ontology: Italian Crime Ontology. *Proceedings of the ICAIL 2003 Workshop on Legal Ontologies and Web based legal information management*, pages 1–7, 2003.
4. J. Bak, C. Jedrzejek, and M. Falkowski. Application of an ontology-based and rule-based model to selected economic crimes: fraudulent disbursement and money laundering. In *International Workshop on Rules and Rule Markup Languages for the Semantic Web*, pages 210–224. Springer, 2010.
5. M. Bastian, S. Heymann, M. Jacomy, et al. Gephi: an open source software for exploring and manipulating networks. *ICWSM*, 8:361–362, 2009.
6. W. N. Borst. *Construction of Engineering Ontologies for Knowledge Sharing and Reuse*, volume PhD. 1997.
7. J. L. Braguês et al. O processo de branqueamento de capitais. *Observatório de Economia e Gestão de Fraude.[Em linha] Porto: Edições Húmus. Disponível em: http://www. fep. up. pt/repec/por/obegef/files/wp002. pdf,[Consult. 25 fev. 2013]*, 2009.
8. J. Breuker and R. Hoekstra. Epistemology and ontology in core ontologies: FOLaw and LRI-Core, two core ontologies for law. *Proceedings of the EKAW04 Workshop on Core Ontologies in Ontology Engineering*, pages 15–27, 2004.
9. P. Burnap, M. L. Williams, L. Sloan, O. Rana, W. Housley, A. Edwards, V. Knight, R. Procter, and A. Voss. Tweeting the terror: modelling the social media reaction to the Woolwich terrorist attack. *Social Network Analysis and Mining*, 4(1):1–14, 2014.
10. E. Cambria and B. White. Jumping nlp curves: a review of natural language processing research [review article]. *IEEE Computational Intelligence Magazine*, 9(2):48–57, 2014.
11. R. Carvalho, M. Goldsmith, and S. Creese. Applying semantic technologies to fight online banking fraud. In *Intelligence and Security Informatics Conference (EISIC), 2015 European*, pages 61–68. IEEE, 2015.
12. J. Ćosić and Z. Ćosić. The Necessity of Developing a Digital Evidence Ontology. *23th Central European Conference on Information ...*, (January 2012):325–330, 2012.
13. D. da República. Lei n.º 49/2008 de 27 de Agosto - Lei de Organização da Investigação Criminal, 2008.
14. S. Decherchi, S. Tacconi, J. Redi, A. Leoncini, F. Sangiacomo, and R. Zunino. Text clustering for digital forensics analysis. In *Computational Intelligence in Security for Information Systems*, pages 29–36. Springer, 2009.
15. J. Domingue. Tadzebao and Webonto: Discussing, Browsing and Editing Ontologies on the Web. In *11th Knowledge Acquisition Workshop*, page 20, 1998.
16. S. Dosis, I. Homem, and O. Popov. Semantic representation and integration of digital evidence. *Procedia Computer Science*, 22:1266–1275, 2013.
17. V. Ermolayev, S. Batsakis, N. Keberle, O. Tatarintseva, and G. Antoniou. Ontologies of time: review and trends. *Int. J. Comput. Sci. Appl*, 11(3):57–115, 2014.

18. E. Ferrara, P. De Meo, S. Catanese, and G. Fiumara. Visualizing criminal networks reconstructed from mobile phone records. *CEUR Workshop Proceedings*, 1210, 2014.
19. A. M. G. Ferreira. Ontospares: da linguagem natural às ontologias. contributos para a classificação automática de dados históricos (séc. xvi-xviii). 2016.
20. V. Furtado, L. Ayres, M. de Oliveira, E. Vasconcelos, C. Caminha, J. D'Orleans, and M. Belchior. Collective intelligence in law enforcement - The WikiCrimes system. *Information Sciences*, 180(1):4–17, 2010.
21. V. Furtado, L. Ayres, M. D. Oliveira, C. Gustavo, and J. Oliveira. Towards Semantic WikiCrimes: Motivation and Goals. *AAAI Spring Symposium: Social Semantic Web: Where Web 2.0 Meets Web 3.0*, pages 27–32, 2009.
22. N. Gcaza, R. V. Solms, and J. V. Vuuren. An Ontology for a National Cyber-Security Culture Environment. *Proceedings of the Ninth International Symposium on Human Aspects of Information Security & Assurance (HAISA 2015)*, (Haisa):1–10, 2015.
23. C. A. Gonçalves, C. T. Gonçalves, R. Camacho, and E. C. Oliveira. The impact of pre-processing on the classification of medline documents. In *PRIS*, pages 53–61, 2010.
24. T. R. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220, 1993.
25. V. Haarslev and R. Möller. Racer: A core inference engine for the semantic web. In *EON*, volume 87, 2003.
26. R. Hoekstra, J. Breuker, M. Di Bello, and A. Boer. The LKIF core ontology of basic legal concepts. *CEUR Workshop Proceedings*, 321:43–63, 2007.
27. J. Hosseinkhani, S. Chaprut, and H. Taherdoost. Criminal network mining by web structure and content mining. *Advances in Remote Sensing, Finite Differences and Information Security. In Proceedings of the 11th WSEAS International Conference on Information Security and Privacy (ISP '12)*, pages 210–215, 2012.
28. H. Hu, Y. Wen, T. S. Chua, and X. Li. Toward scalable systems for big data analytics: A technology tutorial. *IEEE Access*, 2:652–687, 2014.
29. E. Kalemi and S. Yildirim-Yayilgan. Ontologies for Social Media Digital Evidence. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 10(2):335 – 340, 2016.
30. G. Kul and S. Upadhyaya. Towards a cyber ontology for insider threats in the financial sector. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, 6(4):64–85, 2015.
31. W. G. Lehnert and M. H. Ringle. *Strategies for natural language processing*. Psychology Press, 2014.
32. X. Lou. Criminal Network Analysis with Interactive Strategies : A Proof of Concept Study using Mobile Call Logs.
33. A. Luthfi. The Use of Ontology Framework for Automation Digital Forensics Investigation. *International Journal of Computer, Control, Quantum and Information Engineering*, 8(3):423–425, 2014.
34. N. M. Karie. Building Ontologies for Digital Forensic Terminologies. *International Journal of Cyber-Security and Digital Forensics*, 5(2):75–82, 2016.

35. J. Madinger. *Money laundering: A guide for criminal investigators*. CRC Press, 2011.
36. D. Masciandaro. Money Laundering: the Economics of Regulation. *European Journal of Law and Economics*, 7(3):225–240, 1999.
37. M. Mehmet and D. Wijesekera. Ontological constructs to create money laundering schemes. In *CEUR Workshop Proceedings*, volume 713, 2010.
38. R. Merkel, C. Kraetzer, M. Hildebrandt, S. Kiltz, S. Kuhlmann, and J. Dittmann. A semantic framework for a better understanding, investigation and prevention of organized financial crime. In *Sicherheit*, pages 55–66, 2016.
39. M. Mudholkar. A Study on Significance of Event Ontology Approach in Web Crime Mining. 2(2):298–306, 2013.
40. D. Nadeau and S. Sekine. A survey of named entity recognition and classification. *Lingvisticae Investigationes*, 30(1):3–26, 2007.
41. M. Naghavi. A Proposed Architecture for Continuous Web Monitoring Through Online Crawling of Blogs. *International Journal of UbiComp*, 3(1):11–20, 2012.
42. G. Neto and A. Ferraz. Sentimentalista: Um framework para análise de sentimentos baseado em processamento de linguagem natural. 2016.
43. A. Oltramari, L. F. Cranor, R. J. Walls, and P. McDaniel. Building an ontology of cyber security. *CEUR Workshop Proceedings*, 1304:54–61, 2014.
44. H. Park, S. Cho, and H.-C. Kwon. Cyber Forensics Ontology for Cyber Criminal Investigation. In M. Sorell, editor, *Forensics in Telecommunications, Information and Multimedia: Second International Conference, e-Forensics 2009, Adelaide, Australia, January 19-21, 2009, Revised Selected Papers*, pages 160–165. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
45. N. J. Pioch and J. O. Everett. POLESTAR: Collaborative Knowledge Management and Sensemaking Tools for Intelligence Analysts. *Proceedings of the 15th ACM International Conference on Information and Knowledge Management*, pages 513–521, 2006.
46. S. Raghavan, A. Clark, and G. Mohay. FIA: An open forensic integration architecture for composing digital evidence. *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering*, 8 LNICST:83–94, 2009.
47. Q. Rajput, N. S. Khan, A. Larik, and S. Haider. Ontology based expert-system for suspicious transactions detection. *Computer and Information Science*, 7(1):103, 2014.
48. B. Schatz, G. Mohay, and A. Clark. Generalising Event Forensics Across Multiple Domains. *Australian Computer Network and Information Forensics Conference*, pages 1–9, 2004.
49. R. Sharnagat. Named entity recognition: A literature survey. *Center For Indian Language Technology*, 2014.
50. K. Sowkarthikaa and V. P. Sumathi. A Survey of Ontologies on Disease Classification. *International Journal of Science and Research (IJSR)*, 5(4), 2016.
51. J. Stasko, C. Görg, Z. Liu, and K. Singhal. Jigsaw: Supporting investigative analysis through interactive visualization. *VAST IEEE Symposium on Visual Analytics Science and Technology 2007, Proceedings*, (March):131–138, 2007.

52. A. M. Talib and F. O. Alomary. Towards a comprehensive ontology based-investigation for digital forensics cybercrime. *International Journal on Communications*, 5(5):263–268, 2015.
53. Z. Wu, G. Eadon, S. Das, E. I. Chong, V. Kolovski, M. Annamalai, and J. Srinivasan. Implementing an inference engine for rdfs/owl constructs and user-defined rules in oracle. In *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on*, pages 1239–1248. IEEE, 2008.
54. W. Yang. Dynamic Forensics Model based on Ontology and Context Information. 10(2):270–272, 2013.