Key steps for the construction of a glossary based on FunGramKB Term Extractor and referred to international cooperation against organised crime and terrorism

Master Thesis

Student: María Belén Carmona Reche

Supervisor: Ángel Felices Lago

Department of English and German Philologies

Master in English Literature and Linguistics

University of Granada

September, 2014

Acknowledgments

My most sincere gratitude goes to Ángel Felices Lago for his supervision and guidance during the entire academic year at the University of Granada, regarding to the practicum in the Faculty of Economic and Business Sciences as well as to the dissertation. Without his professional advice and his kind attitude the completion of this dissertation would not have been possible.

I am indebted to other lecturers from the Master in English Literature and Linguistics, from whose expertise and advice I have greatly benefited: Salvador Valera Hernández, Manuel Jiménez Raya, Belén Soria Clivillés, Pillar Villar Argáiz, Cristobal Lozano Pozo, Encarnación Hidalgo Tenorio, and Celia Wallhead.

I am also grateful to the researcher in the National University of Distance Education María de los Llanos Carrión Varela who sent me her dissertation as well as her thesis at a great speed during the summertime.

I would also like to thank the Director of the Master in English Literature and Linguistics José María Pérez Fernández for making it possible as well as for his commitment. Thanks must also go to my peer Miriam Castillo Arroyo for her advice during the entire academic year.

My gratitude also goes to my family, to my mother, to my father and to my grandmother for their generous support, and especially to my mother for her emotional support.

Needless to say, any errors and oversights are my own.

Table of Contents

List of Abbreviations

CCJ	Criminal Court of Justice
COREL	Conceptual Representation Language
EU	European Union
EUROPOL	European Police Office
EUROJUST	European Judicial Cooperation Unit
FunGramKB	Functional Grammar Knowledge Base
FGKBTE	FunGramKB term extractor
GCTC	Global Crime Term Corpus
NLP	Natural Language Processing
OAS	Organisation of American States
OECD	Organisation for Economic Cooperation and Development
OSCE	Organisation for Security and Co-operation in Europe
UN	United Nations
USA	United States of America

List of Figures

List of Tables

Abstract

The employment of new technological instruments for the processing of natural languages is crucial to improve the way humans interact with machines. The Functional Grammar Knowledge Base (FunGramKB henceforth) has been designed to cover Natural Language Processing (NLP henceforth) tasks in the area of Artificial Intelligence. The multipurpose lexical conceptual knowledge base FunGramKB is capable of combining linguistic knowledge and human cognitive abilities within its system as a whole. The conceptual module of FunGramKB contains both common-sense knowledge (Ontology), procedural knowledge (Cognicon) as well as knowledge about named entities representing people, places, organisations or other entities (Onomasticon). The Onomastical component is used to process the information from the perspective of specialised discourse. The definition in Natural Language of a consistent list of encyclopaedic terms existent referred to the legislation and to entities which fight against organised crime and terrorism existent in the GCTC would be the stepping stone for the future development of the Onomasticon. The FunGramKB Term Extractor (FGKBTE henceforth) is used to process the information. To cope with the inclusion of the terms in the Onomasticon according to the Conceptual Representation Language (COREL henceforth) schemata, the DBpedia project has been of paramount importance to develop specific patterns for the structure of the definitions.

***Keywords*:** legal ontology, FunGramKB, GCTC, terminology, Onomasticon, DBpedia.

Resumen

El empleo de nuevas herramientas tecnológicas para el Procesamiento del Lenguaje Natural (PLN en adelante) es fundamental para mejorar la forma en que las máquinas se relacionan con los seres humanos. FunGramKB ha sido diseñada para abordar tareas de PLN inmersas en el área de la Inteligencia Artificial. La base de conocimiento léxico conceptual multipropósito FunGramKB es capaz de combinar el conocimiento lingüístico con las habilidades cognitivas humanas dentro de su sistema como conjunto. El modulo conceptual de FunGramKB se basa en el sentido común (Ontología) y en el conocimiento procedimental (Cognicón), a la vez que en el conocimiento sobre entidades nombradas que representan personas, lugares, organizaciones u otras entidades (Onomasticon). La definición en Lenguaje Natural de una lista consistente de términos enciclopédicos concerniente tanto a instrumentos legales como a organizaciones que luchan contra el crimen organizado y el terrorismo que se ha incluido en el GCTC supondrá un gran adelanto en aras al futuro desarrollo del Onomasticon. El FGKBTE se usa para procesar la información. Con vistas a incluir los términos en el Onomasticón de acuerdo al esquema COREL, el proyecto DBpedia ha sido de una importancia fundamental para desarrollar patrones determinados con los que estructurar las definiciones.

Palabras clave: ontología legal, FunGramKB, GCTC, terminología, Onomasticon, DBpedia.

Everything that is possible demands to exist.

Gottfried Wilhelm Von Leibniz

I have a dream for the Web in which computers become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers. A “Semantic Web”, which makes this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The "intelligent agents" people have touted for ages will finally materialize.

Timothy John Berners-Lee

Introduction

The purpose of this M.A. thesis is to contribute to the compilation of an English encyclopaedic glossary in the fields of Global Organised Crime and Terrorism based on the GCTC inserted in one of the key utilities of FunGramKB: FGKBTE. To begin with, the list of terms included in the glossary would help to build a terminological subontology and would be the stepping stone for the future development of the specialized Onomasticon within FunGramKB conceptual level. Both the creation of an exhaustive body with specialised knowledge and the development of multipurpose tools for structuring such knowledge are essential for easy retrieval of information in professional contexts as well as for the solving of problem-oriented tasks in real situations. This way, it complies with the upsurge of new social demands coming from educational, academic and other professional contexts (Ureña Gómez-Moreno, Alameda Hernández & Felices Lago, 2011: 302).

Thanks to the Internet, modern society has access to a massive volume of information. Access to such a wealth of information often complicates and delays the process. This is partly because the machines that process the information do not comprehend the information which is being processed. Those machines are limited to transfer the information from one user to another, without having the capacity to process human queries which might be relevant. Due to the lack of accuracy, the search by the machine results in a loss of time. The current searching-machines of the Internet lack the semantic knowledge needed. Those machines are capable of identifying and recognizing the chain of graphic symbols and words. However, they are not capable of identifying what is hidden behind the symbols or knowing the true objective of the user (Carrión Varela, 2010: 6). Actually, for the machine it is merely symbols without meaning.

The lack of ability to infer data may be due to the fact that it may not have been anticipated before in a systematic way and so the machine ignores its existence. The machines need the ability to apply common sense to be able to resolve problems and answer questions based on semantic knowledge of the human query. It will be after having applied common sense that the machine will not only give back words but also answers. In this sense, it is necessary to implement the web of semantic knowledge and therefore to get into the field of Artificial Intelligence (Carrión Varela, 2010: 2). It means to carry out the Semantic Web. There have been attempts to implement semantic knowledge into texts written in a natural

language. There have been many attempts to build a more accessible and intelligent web, but there is still a strong need for a truly efficient model of communication between humans and machines. Nevertheless, a significant step towards this direction has been taken with the development of mechanisms of conceptual reasoning as that contained within the knowledge base FunGramKB. In this vein, the mechanisms of conceptual reasoning that FunGramKB is actually developing will be pretty useful for institutions and professionals determined to access to relevant information in a more smart and quick way (Ureña Gómez-Moreno, 2014: 513).

This dissertation contains five main sections. First of all, a theoretical standpoint in which the fundamental background of concepts discussed will be taken. An explanation of what an ontology and a subontology are will be given. The reason is that this dissertation has amongst its main objectives to help in the population of a legal subontology. Hence, I take into account the content of the organised crime and terrorist offenses which the legal ontology is built of. Besides, since the research I try to put forward is aimed at contributing with the GCTC, I provide some basic notions about this system. In addition, the architecture of FunGramKB is described. Moreover, I will also introduce the tool FunGramKB Term Extractor (FGKBTE henceforth), which turns out to be a cornerstone for the semi-automatic extraction and management of terminology from corpus data. I endeavour to concentrate specifically on the theoretical and practical aspects of the conceptual module and the building of the "Onomasticon" in particular. For the purpose of this dissertation, it is essential to concentrate on the Onomasticon because it is determined to develop the encyclopaedic knowledge of specialised fields. Finally, I explore the terminological aspects concerning the DBpedia project which appears to be fundamental for strengthening the goals of FunGramKB.

As to the methodology of the research, it is divided into three parts or stages. The first phase is about a previous search of corpus. I will refer to the encyclopaedic sources and references. The second stage is about the term extraction from FGKBTE. I will describe two tools included in it. I endeavour to explain the function called "search" used to see the amount of relevant terms chosen in the previous search included in GCTC. Besides, I commit to explaining the "corpus" function included in FGKBTE, which helps to find out the amount of terms that were chosen as relevant and which have been deleted accidentally. Lastly, I concentrate on how to decide upon the properties that demarcate the terms according to

DBpedia project. Furthermore, I focus on how to draft the definitions of the terms that will be defined in natural language. This will prepare their inclusion in FunGramKB´s Onomasticon.

Thirdly, the output is included. To begin with, a table of results with the selection of terms defined is displayed. A distinction is drawn between "Organised Crime", "Terrorism" or "Organised Crime and Terrorism". Besides, the terms with their corresponding properties to contribute to the building of the Onomasticon are presented. To sum up, the entire glossary or the subsequent repository of terms is showed. In addition, I have analysed the compliance of the compiled corpus with the two main criteria to build the GCTC, namely, balance and representativeness. And finally, in the remainder of this dissertation, the conclusions and the future avenues of research related to this study are presented.

I. Theoretical framework

1. Ontologies and subontologies

There are different meanings of the term ontology and different types of ontologies depending on the way they are represented or structured (Valente, 2005: 66). Nevertheless, there seems to be common features to any ontology. All of them gather relevant information of a particular field or domain in the world, concerning the entities that build it, the entities´ features, the events that occur and the relationships amongst all these concepts (Carrión Valera, 2010: 7).

Gruber refers to the ontologies as a knowledge-based system, or knowledge-level agent committed to some conceptualization. He adds that

> "*In such an ontology, definitions associate the names of entities in the universe of discourse (e.g., classes, relations, functions, or other objects) with human-readable text describing what the names are meant to denote, and formal axioms that constrain the interpretation and well-formed use of these terms"* (Gruber,1993: 2).

In particular, concerning the features or functions that the ontologies made in the field of the knowledge engineering should or may include, those have been described by Bateman as follows:

> *"The following list gives an idea of the range of functions adopted in NLP. Ontologies are often expected to fulfill at least one (and often more) of:*
> *• organizing 'world knowledge',*
> *• organizing the world itself,*
> *• organizing 'meaning' or 'semantics' of natural language expressions,*
> *• providing an interface between system external components, domain models, etc. and NLP linguistic components,*
> *• ensuring expressability of input expressions,*
> *• offering an interlingua for machine translation,*
> *• supporting the construction of 'conceptual dictionaries'.*
> *Moreover, an ontology is seen as a very general organizational device: i.e., one that provides a classification system for whatever area of application the ontology is applied to. The organizational resource offered by an ontology has to be re-usable*" (Bateman, 1991: 51).

In this vein, an ontology is a representation of human knowledge which computers can use to process and mine information in a connected and efficient way (Periñán-Pascual & Arcas-Túnez, 2014: 114). It is composed by a hierarchy of concepts, attributes and agreed relations that is determined to establish semantic nets of relations, as those contained within FunGramKB will prove. As a consequence, ontological engineering is a discipline that is actually having a great application because of the ontologies´ computational implementation for NLP tasks and their application in a number of areas, including Data Mining, Machine Translation, Semantic Web[1] technologies and Artificial Intelligence.

A major field of work of the ontological endeavour concerns the building of specialised ontologies, which are also called domain-level ontologies. Domain-level ontologies are taxonomies containing conceptual information specific to highly expert areas of human activity such as medicine or law. Domain-level ontologies used in the field of ontological engineering have the potential for the automatic solving of reasoning tasks in

[1] The main purpose of the Semantic Web is driving the evolution of the current Web by enabling users to find, share, and combine information more easily. http://en.wikipedia.org/wiki/Semantic_Web. Accessed on 10 September 2014.

which expertise is commonly required, as well as the design of artificial agents which simulate human expert decisions. In this context, terminological work plays a crucial role in domain-specific ontology engineering because it provides the lexical substratum upon which conceptualisation is based. It implies that, in order to implement domain-level ontologies it is essential to identify the specific concepts used by experts in their professional activities alongside the technical words (Felices Lago & Ureña Gómez-Moreno, 2014: 252).

Concerning the field of law, the search for information is one of the most time-consuming tasks in the daily work of legal professions. In this regard, advanced computer-based applications of natural language processing in the field of law are being developed. To implement those, models such as the model of specialised knowledge representation driven by the deep semantics of the knowledge base FunGramKB are built. The NLP systems based in deep semantics related to the law field may offer functionalities such as the information extraction of sentences´ summaries, the automatic store of documents in a judicial process, the identification of the implicated parties and of the adopted resolutions and may even serve as a guide of the parties´ testimonies reasoning. In this vein, the legal ontology of FunGramKB under construction will be implemented to understand legal discourse automatically (Periñán Pascual & Arcas Túnez, 2014: 114, 131) as well as to serve as a tool for the organisation and conceptualisation of legal information (Ureña Gómez-Moreno, Alameda Hernández & Felices Lago, 2011: 67).

2. Definitions of Organised Crime and Terrorism

As the glossary and the subontology chosen to work in this dissertation lies within the framework of the offenses in Criminal Law called organised crime and terrorism, here it is necessary to talk about their corresponding scopes. To sum up, this legal subdomain of organised crime and terrorism has been selected both for its international relevance nowadays as well as for the scarce references available with the aim of working with and populating ontologies (Ureña Gómez Moreno, Alameda Hernández & Felices Lago, 2011: 303). Consequently, the population of ontologies in this field is proving necessary.

According to Interpol, definitions of what constitutes organised crime vary widely from one country to another. Despite this fact, Interpol indicates that the activities in which

organised networks are involved may include trafficking in human beings, illicit goods, weapons and drugs, armed robbery, counterfeiting and money laundering. Organised networks are said to engage in many different types of criminal activity spanning several countries. Furthermore, the Federal Bureau of Investigation[2] (FBI henceforth) defines organised crime as any group having some manner of a formalized structure and whose primary objective is to obtain money through illegal activities. Such groups maintain their position through the use of actual or threatened violence, corrupt public officials, graft, or extortion.

On the other hand, terrorism is defined by the United Nations General Assembly[3] in its 1994 United Nations Declaration on Measures to Eliminate International Terrorism annex to UN General Assembly resolution 49/60, "Measures to Eliminate International Terrorism", of December 9, 1994, UN Doc. A/Res/60/49. It condemns criminal acts intended or calculated to provoke a state of terror in the general public, a group of persons or particular persons that for political purposes are in any circumstance unjustifiable, whatever the considerations of a political, philosophical, ideological, racial, ethnic, and religious or any other nature that may be invoked to justify them.

3. GCTC

This acronym stands for "Global Crime Term Corpus". It is a text collection which has been compiled for the creation of the Satellite Ontology on terrorism and organised crime within the knowledge base FunGramKB. The GCTC contains a selection of reliable sources from institutions such as the European Union, the Council of Europe, the International Criminal Court, the Europol and the Organization for Security and Co-operation in Europe. Furthermore, those organisations dealing with international cooperation against organised crime and terrorism offer an enriching representation of the technical aspects as well as of the specialised vocabulary used in official documents as to combat criminal acts in a global and democratic world (Carrión Delgado, 2012: 192).

[2] http://en.wikipedia.org/wiki/Federal_Bureau_of_Investigation. Accessed on 18 July 2014.
[3] http://www.un.org/en/ga/. Accessed on 18 July 2014.

At a certain point, the decision to complete and close the edition of the corpus in the GCTC and stop the possibility of including more data had to be taken in order to proceed with the statistical analysis in FGKBTE. That is the reason why the data gathered in the GCTC had to be chosen carefully beforehand. Two main criteria were taken into account for the compilation of the corpus, namely, representativeness and balance. The representativeness has to do with a wide variety of legal text types, amongst which there may be international treaties, fact sheets, rules, resolutions, conventions and acts. Besides, the corpus had to be balanced with respect to the number of texts on the domains under study, paying attention to the rich variety of those texts focused on terrorism, those dealing with organised crime, and the percentage that accounts for texts on both types of subject areas. (Felices Lago & Ureña Gómez-Moreno, 2014: 253). The GCTC currently contains 621 texts in English, of which 49% deal with organised crime, 34 % deal with terrorism, while the remaining 17 % contains texts concerning these two topics simultaneously (Ureña Gómez-Moreno, Alameda Hernández and Felices Lago, 2011: 304). Furthermore, the process of statistical relevance of the terms to be inferred as candidate terms will be explained in the section "FGKBTE".

Besides the function of the tailor-made GCTC to serve as a repository in the population of a legal subontology of the abovementioned domains, the terminological glossary will also be implemented to be suitable for application by humans by means of a dictionary interface. Moreover, this specialised corpus is intended to be applied by a machine in NLP tasks such as Data Mining, Machine Translation, Semantic Web technologies and Artificial Intelligence (Ureña Gómez-Moreno, Alameda Hernández & Felices Lago, 2011: 301). Moreover, the deep semantics applied to this terminological ontology, together with the implementation of the cognitive mechanisms which act as the human reasoning, show GCTC as a practical resource for its exploitation in NLP systems that require the understanding of the language, particularly that concerning to the development of intelligent agents composed by a communication interface person-machine (Periñán Pascual & Arcas Túnez, 2014: 131).

4. FunGramKB

The multipurpose lexical conceptual knowledge base for NLP systems FunGramKB[4] is a development within the field of Artificial Intelligence and the Semantic Web. The term "knowledge base" is defined in this context as a computational repository where conceptual and linguistic knowledge is stored and accessed in a connected, meaningful and efficient way. FunGramKB has been designed because it even adds a value to the languages with semantic labels such as Web Ontology Language (OWL henceforth). The OWL has been invented to ease the retrieval and extraction of information from the World Wide Web in an intelligent way. Thus, FunGramKB is meant to strength the field of the Semantic Web.

In summary, FunGramKB is provided with a model of representation with a powerful semantic foundation. This model is capable of producing linguistic labels with complete meaning. In this regard, FunGramKB is presented as a robust knowledge base provided with deep semantics instead of surface semantics. It means that it is machine-usable, meaning that it is able to understand the consultation made in a natural language and then retrieve the information accurately as required. In this vein, the Lexical Constructional Model (LCM henceforth) is the linguistic model that can explain the relation between both the lexical and the conceptual levels of FunGramKB. It is a usage-based meaning construction model that explains the relationship between syntax and all the different facets of meaning construction (Felices Lago, 2012: 25). In Ruiz de Mendoza Ibáñez & Mairal Usón words,

> "*The primary concern of the LCM is to develop a usage-based, comprehensive theory of meaning construction that aims to give explanations of how all aspects of meaning, including those that go beyond so-called core-grammar (e.g., traditional implicature, illocutionary force, and discourse coherence) interact among one another"* (Ruiz de Mendoza Ibáñez y Mairal Usón, 2008: 355).

In this scenario, FunGramKB is composed of structures of human beings´ cognitive system rendered by means of conceptual representations. Moreover, FunGramKB enriches the applications for NLP because it acts as an intelligent agent for the processing of information,

[4] "FunGramKB is part of the research carried out in the project "Implementación de una base de conocimiento para el razonamiento del sentido común" funded by Universidad Católica San Antonio de Murcia, code number PMAFI-PI-11/1C/05" (Periñán. Pascual & Arcas Túnez, 2005: 243)

as a prototype of automatic translation as well as a dictionary based on conceptual searches, which could be catalogued as "dictionary of the third millennium" (Periñán Pascual & Arcas Túnez, 2004: 38). Figure 1 captures a screenshot with the FunGramKB Editor website which contains four main sections: FunGramKB Modules, Resources, Utilities, Reviewers and Administrator.

Figure 1. FunGramKB Editor

In the words of Periñán Pascual and Arcas Túnez:

> *"FunGramKB Suite*[5] *is a user-friendly online environment for the semiautomatic construction of a multipurpose lexico-conceptual knowledge base for NLP systems, and more particularly for natural language understanding*" (Periñán Pascual and Arcas Túnez, 2010: 2667).

It is multipurpose in the sense that is multifunctional and multilingual. In this sense, FunGramKB can be used in many NLP tasks and so in several computational tasks such as machine translation. Besides, it has been created to work with any human language. Moreover, quoting Periñán Pascual and Arcas Túnez:

> "*On the other hand, our knowledge base comprises three major knowledge levels, consisting of several independent but interconnected modules:*

[5] "FunGramKB Suite" refers to the knowledge-engineering tool (www-fungramkb-com) and "FunGramKB" to the resulting knowledge base.

Lexical level:

- *The Lexicon stores morphosyntactic, pragmatic and collocational information about lexical units.*
- *The Morphicon helps our system to handle cases of inflectional morphology.*

Grammatical level:

- *The Grammaticon stores the constructional schemata which help Role and Reference Grammar (RRG) to construct the semantics-to-syntax linking algorithm (Van Valin and La Polla, 1997; Van Valin, 2005).*

Conceptual level:

- *The Ontology is presented as a hierarchical catalogue of the concepts that a person has in mind, so here is where semantic knowledge is stored in the form of meaning postulates. The Ontology consists of a general-purpose module (i.e. Core Ontology) and several domain-specific terminological modules (i.e. Satellite Ontologies).*
- *The Cognicon stores procedural knowledge by means of scripts, i.e. conceptual schemata in which a sequence of stereotypical actions is organised on the basis of temporal continuity, and more particularly on the basis of Allen´s temporal model (1983).*
- *The Onomasticon stores information about instances of entities and events, such as Bill Gates or 9/11. This module stores two different types of schemata (i.e. snapshots and stories), since instances can be portrayed synchronically or diachronically*" (Periñán Pascual and Arcas Túnez, 2010: 2667).

Unlike the lexical and the grammatical module, the conceptual module is shared by all languages in the sense that covers all those properties which are universal. It means that computational linguists must build one lexicon, one morphicon and one grammaticon for each different language but they build only an ontology, one cognicon and one onomasticon to process any language input conceptually. In this scenario, FunGramKB has taken on a conceptualist stance because the Ontology is the pivotal module for the whole architecture (Periñán Pascual, 2010: 2667).

Figure 2 shows the architecture of the three modules of FunGramKB. The larger arrows refer to the interconnected nature of each module and the thinner arrows refer to the interconnected nature of each sublevel of description.

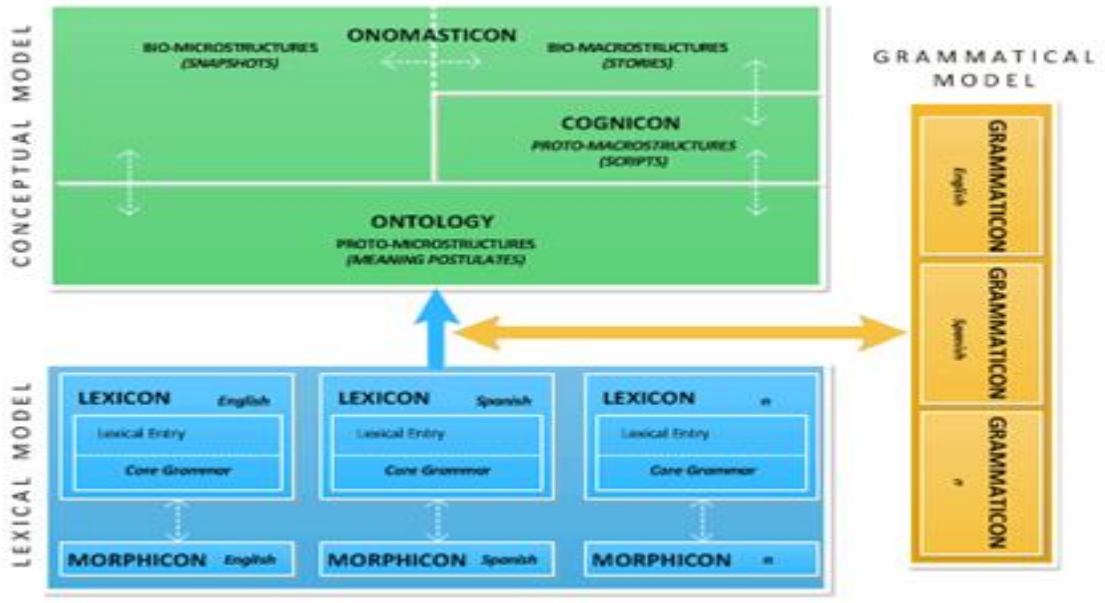

Figure 2. The architecture of the knowledge-base FunGramKB

5. FGKBTE

The acronym FGKBTE stands for FunGramKB Term Extractor. It is an innovative corpus-based tool essential in the process of Satellite Ontology creation. It helps with the automatic and the semi-automatic retrieval of terms from domain-specific corpora. Thus, it helps with the construction of Satellite ontologies (Felices Lago & Ureña Gómez-Moreno, 2014: 252). FGKBTE is included in the FunGramKB Suite, which constitutes a user-friendly environment for the semiautomatic construction of the knowledge base. This terminological extractor is responsible for both the management of the GCTC as well as the extraction and editing of the terminological units contained within it. Furthermore, FGKBTE consists of six tabs which in turn refer to several functions such as corpus management, extraction and editing of terminological units. Figure 3 captures the main menu.

Figure 3. FunGramKB Term Extractor

The "Pre-processing" tab allows the testing and assessing of new features for future implementation by the extractor. The "Processing (indexing)" tab handles the structured uploading of the texts of a corpus to the extractor alongside basic documentary information as it is a title, a field tag and a short description of the main contents of the corpus files. Nevertheless, tabs such as indexing are based on permission policies and restricted access. Thus, the uploading of documents in GCTC is available only at certain stages of the extraction (Felices Lago, 2012: 49). The "Processing (statistics)" tab is responsible for the mathematical calculation of candidate terms. The "View" tab is concerned with the filtering of terms considered as false terms carried out through a series of removal options. The "Search" tab facilitates the location of strings of text to retrieve specific contexts in which those linguistic structures are used in any of the stored corpora. The "Search" tool might be seen in Figure 4. Moreover, the "Corpus" tool displays a list of the indexed texts, the total number of tokens included as well as a terminological box comprising a list of false candidates that were discarded by the extractor during the filtering process tackled in the "View" function (Ureña Gómez Moreno, Alameda Hernández & Felices Lago, 2011:305). In addition, the terminological extractor includes the "View" tab, which provides a sorted list of potentially relevant terms and operates with several options, such as discarding a term, definition and connection to the conceptual model (Felices Lago, 2012: 48).

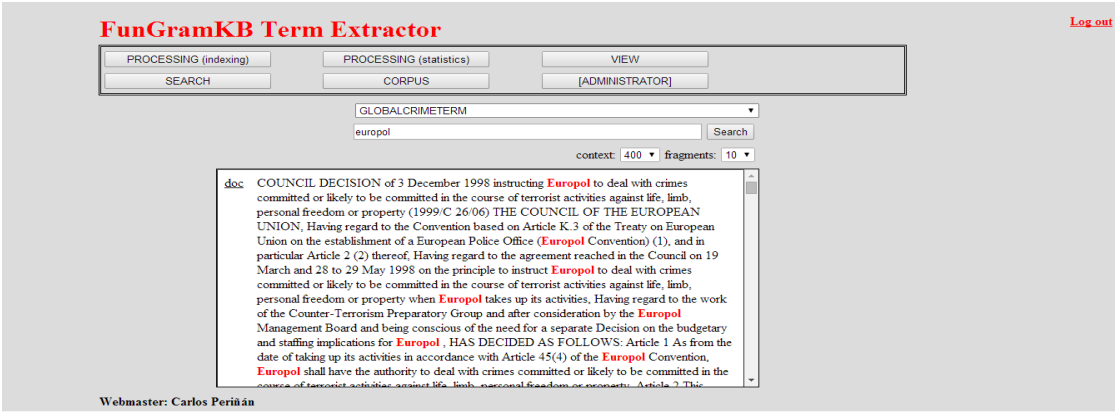

Figure 4. Search tool in FGKBTE

Moreover, if one abides by the distinction drawn between statistically-based extractors and linguistically-based extractors, FunGramKB Term Extractor belongs to the former (Felices Lago, 2012: 47). It is noteworthy to remark here that terminology extraction in FGKBTE is conceived as a two-phase process, consisting of an automatic phase and a manual phase. In the automatic phase, FGKBTE retrieves sets of candidate terms which are identified by the machine as terminological by automatic means. Those terms are retrieved after performing clean-up tasks on the input corpus through a number of filters. This terminological extractor mainly relies on statistics, that is to say on the statistical calculation of the semantic weight of corpus tokens carried through the so-called "term frequency inverse document frequency" (tf-idf). Thus, FGKBTE calculates a tf-idf score for each n-gram in the corpus (Ureña Gómez- Moreno, Alameda Hernández & Felices Lago, 2011: 304). An n-gram is a basic unit of text which coincide with what is understood as a word in an orthographic sense. With regards to the tf-idf, Felices Lago points out that:

> *"This metric, which is expressed with a numerical index, is the result of normalising the occurrence of a word in a document (tf) with the appearance of the same token in the various documents which make up the corpus (idf) (…) The terminological relevance of a candidate and the likelihood that it is in fact a specialised term increases or decreases proportionally with the tf-idf value"* (Felices Lago, 2012: 52).

As an outcome of the mathematical operation the extractor gives three lists of n-grams proposed as candidate terms: unigrams, bigrams and trigrams. The bigrams and the trigrams contain a string of two or three elements respectively separated by a blank space (Felices Lago, 2012: 53). Figures 5, 6 and 7 show the "n-gramatic" candidates obtained for the GCTC:

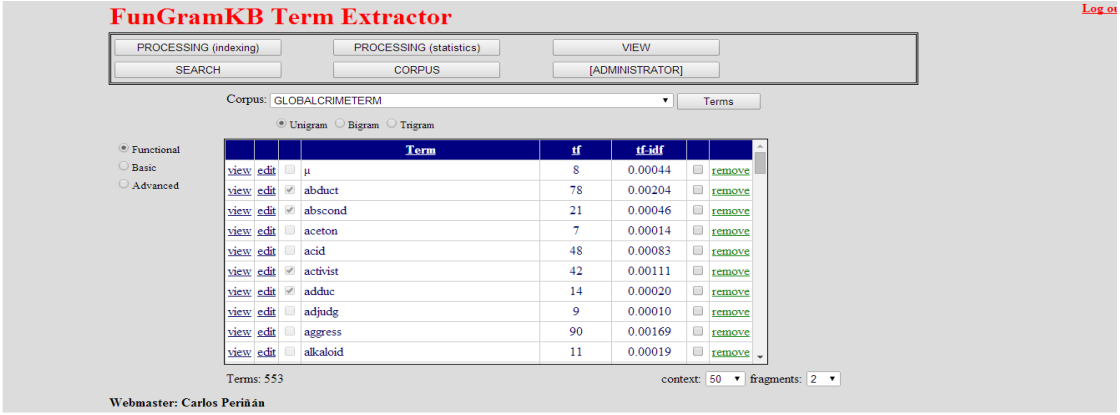

Figure 5. FGKBTE list of candidate unigrams

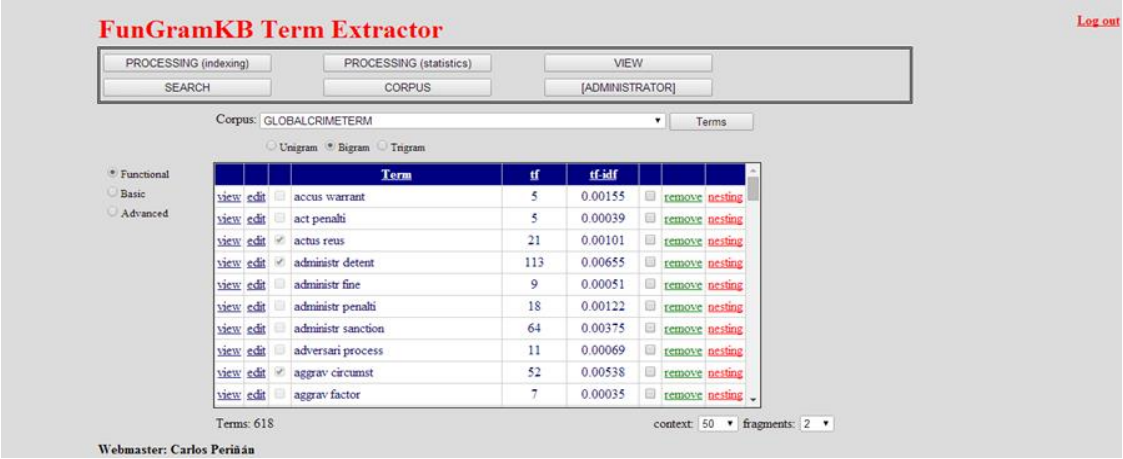

Figure 6. FGKBTE list of candidate bigrams

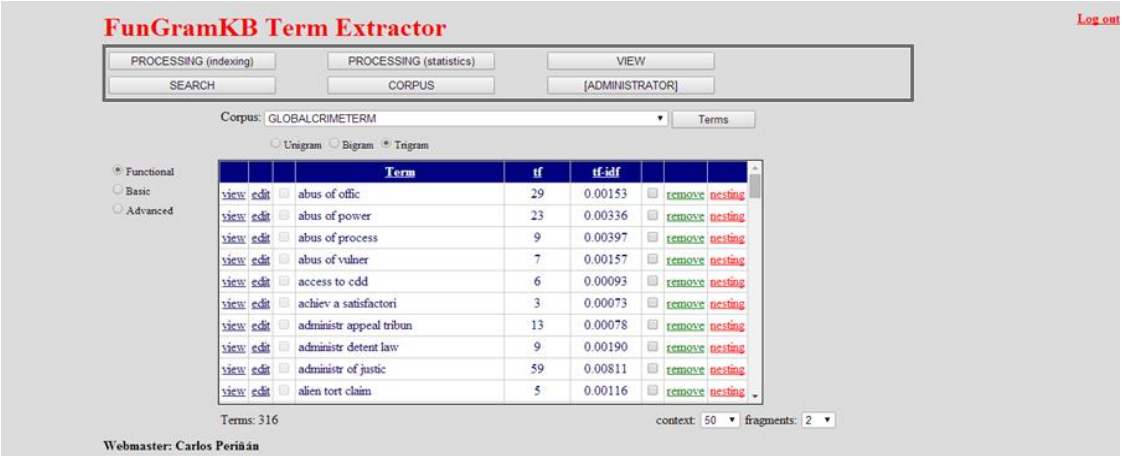

Figure 7. FGKBTE list of candidate trigrams

Thus far, the automatic process leading to the preliminary list of candidate terms has been explained. But in addition to the automatic phase, FGKBTE is complemented by manual editing (Felices Lago & Ureña Gómez-Moreno, 2014: 256-257). In order to come up with the winning terms, manual filtering tasks on the aforementioned candidate set of terms is necessary. In this manual phase, it is the terminologist who eventually decides upon the set of units that are convenient. That means that the terminologist decides the winning terms to build the Satellite Ontology. The manual editing involves discarding the candidate terms culled in the automatic phase as well as recovering the terms that have been discarded by a terminologist in a previous manual editing. This phase is crucial in the process of

subontological creation because the linguistic units that are considered terminological will be kept for a definition in natural language in a later stage (Felices Lago, 2012: 55).

6. FunGramKB Conceptual Module: the Onomasticon

As Periñán Pascual and Arcas Túnez (2010: 2668) explain, the "scheme" model originated in cognitive psychology, and thereafter used in artificial intelligence, is essential to the representation of the knowledge in FunGramKB. Conceptual schemata of FunGramKB are essential to the inference of knowledge during the process of understanding of the language. FunGramKB classifies conceptual schemata along two dimensions: prototypicality and temporality. On the one hand, conceptual representations either store prototypical knowledge, i.e. protostructures or can be used to describe a specific entity or event i.e. bio-structures. On the other hand, knowledge within conceptual schemata can be presented atemporally i.e. microstructures or we can represent temporal frameworks i.e. macrostructures. Figure 8 illustrates the typology of conceptual schemata in FunGramKB.

		TEMPORALITY	
		-	+
PROTOTYPICALITY	+	Proto-microstructure (Meaning postulate)	Proto-macrostructure (Script)
	-	Bio-microstructure (Snapshot)	Bio-macrostructure (Story)

Figure 8. Typology of conceptual schemata in FunGramKB

Within the proto-structures we differentiate between proto-microstructures and proto-macrostructures. When the description of the meaning postulate of a concept is involved it is called proto-microstructure, and it is stored in the Ontology. For instance, the description of the meaning of *court order* or that of a *crime* involves the construction of the proto-structure

of its corresponding concept. Conversely, the description of one or more predications of a prototypical event within a linear temporal framework, that is to say the procedural knowledge of it, is done by means of the proto-macrostructures. In addition, proto-macrostructures are stored in the Cognicon. To add an example of a script, *how to file a complaint* could be used.

Within the bio-structures we distinguish between bio-microstructures and the bio-macrostructures, both of which are integrated within the Onomasticon. The bio-microstructures provide a useful representation for the instances that need to be represented atemporally or synchronically as snapshots, in a similar vein to the description of functions of the President of Eurojust. Conversely, the bio-macrostructures provide a helpful representation for the instances diachronically or in a temporal framework and so as stories, as it is done with the description of the *biography of the President of Eurojust* (Periñán Pascual and Arcas Túnez, 2014: 119). The Cognition Planet might be seen in Figure 9.

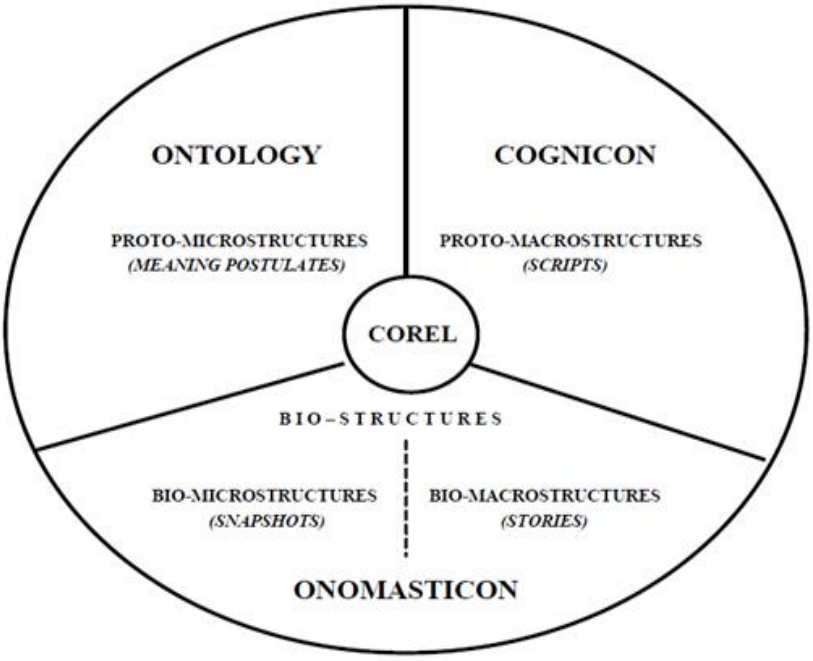

Figure 9. The Cognicon Planet

Figure 10 shows the necessary interrelationship between the three modules of the conceptual level of FunGramKB. In order to ensure the efficiency and success in reasoning tasks of a NLP system, thus allowing it to come to conclusions, the various kinds of knowledge have to be represented through the same formal language. It is necessary to

emulate the components of the human memory for an indefinite period of time. The formal language that represents the three types of conceptual knowledge is COREL (*Conceptual Representation Language*) (Carrión Varela, 2010: 25). Additionally, COREL has the advantage of being machine-readable as well as legible for human beings (Ureña Gómez-Moreno, 2010: 510). Garrido García summarises the features of COREL as follows:

> *"The formal language COREL represents knowledge schemata in FunGramKB. It differentiates two kinds of analytical tool for all representations: (i) a distribution of conceptual meaning across levels that are classified hierarchically and favor the activity inheritance of mechanisms; (ii) a specific notation system in which the interlinguistic representations have to comply with syntax, so that they can be considered as well-formed structures"* (Garrido García, 2011: 24).

The remainder of this section pays special attention to the applications of the Onomasticon. The reason is that this is the conceptual module component which is meant to be developed in this research. The role of the world´s cultural knowledge represented via snapshots and stories in the Onomasticon is essential to establish relationships between the entities and the events which are not apparently associated. For instance, concerning the domain of organised crime and terrorism, the description of criminal cells or individual and collective agents that perpetrate criminal acts can contribute to the discovering and forecast of new profiles of risk taking into account the profiles of people who have similar patterns of criminal behaviour. In this vein, events such as "participation", "belonging", "collaboration" or "buying and selling", etc. referred to any entity could link to any other well-known entity. Furthermore, that link could suggest actions or recommend decision making. The reasoning engine of the knowledge base could be employed to find relations between the places where the crimes have happened and the people under investigation (Ureña Gómez-Moreno, 2010: 513). To sum up, the conceptual encyclopaedic knowledge of the entities and events included in Onomasticon is represented through COREL language once it has been defined according to the DBpedia schemata. The Onomasticon can import the knowledge compiled in digital repositories such as that of Wikipedia infoboxes by means of the structured knowledge base DBpedia, and connect it to the other modules with the aim of increasing the capacity of processing and reasoning of the knowledge base.

7. The DBPedia project

Summarising, according to Lehmann et al.:

> *"The DBpedia community project extracts structured, multilingual knowledge from Wikipedia and makes it freely available using Semantic Web and Linked Data standards. (…) Being a central part of the Linked Open Data (LOD) cloud, it serves as a connection hub for other data sets. For the research community, DBpedia provides a testbed serving real world data spanning many domains and languages. Due to the continuous growth of Wikipedia, DBpedia also provides an increasing added value for data acquisition, re-use and integration tasks within organisations"* (Lehmann et al., 2012: 1).

It is the Onomasticon which uses the mediation of the DBpedia taxonomy to transfer structured information from Wikipedia (Ureña Gómez-Moreno, 2014: 510). The DBpedia project extracts structured information from Wikipedia and generates the DBpedia knowledge base. It aims to make the knowledge accessible in the web (Hahn, 2010: 4). The systematic classification of the entities and their properties in the DBpedia taxonomy is extremely useful for the Onomasticon. In the first instance the ontological tools and models are reusable. Moreover, in the second instance because the huge amount of information might be introduced in the knowledge base by means of semi-automatic processes (Carrión Varela, 2014: 87, 92).

DBpedia taxonomy can be built manually with the aid of Wikipedia infobox templates. Infoboxes are normally used to list the most relevant facts that can be found in an article, structured in the form of relations of attribute-value pairs. As Lehman et al. (2012: 3) indicate "*Infoboxes that appear in a Wikipedia article are based on a template that specifies a list of attributes that can form the infobox*". Infobox templates are usually placed on the upper right-hand side of an article for right-to-left languages (Periñán Pascual & Carrión Varela, 2011:92). Figure 10 shows a Wikipedia info-box.

Figure 10. Wikipedia Info-box

The Wikipedia infobox template system has not evolved with central coordination amongst the different Wikipedia editors and so the attributes or property values are expressed using a wide range of different forms of measurement. Different templates use different names for the same property (e.g. formation or formed). In order to homogenise the attribute names, the DBpedia project maps Wikipedia templates onto an ontology using a custom mapping language (Hahn, 2010:4), meaning that DBpedia is built with the templates which are most commonly used. Besides, on DBpedia the different entities are stored as in a hierarchy of classes. All the classes are gathered under only one category which is a hyperordinate class called "Thing". That hierarchy of classes is a huge net of more than three hundred and twenty classes of entities defined by a total of one thousand six hundred and fifty properties. In general, the entities gathered within the same class or sub-class tend to share the same properties.

After the terminologist has decided the most common properties to populate the ontology and a coherent definition in natural language for the entities is supplied, it is time to develop an ontology schema and mappings from Wikipedia infobox properties to the DBpedia ontology. This process is defined as a Mapping-Based Infobox Extraction by the DBpedia community and it can be done by means of the DBpedia Mappings Wiki. The DBpedia Mappings Wiki is used to map different templates of Wikipedia to the DBpedia ontology. As Lehman et al. describe:

> *"This effort is realized using the DBpedia Mappings Wiki[6], a MediaWiki installation set up to enable users to collaboratively create and edit mappings. These mappings are specified using the DBpedia Mapping Language. The mapping language makes use of MediaWiki templates that define DBpedia ontology classes and properties as well as template/table to ontology mappings. A mapping assigns a type from the DBpedia ontology to the entities that are described by the corresponding infobox. In addition, attributes in the infobox are mapped to properties in the DBpedia ontology*" (Lehmann et al., 2012: 5).

It is noteworthy to highlight that despite the fact that the DBpedia ontology consists of three hundred and twenty classes described by one thousand six hundred and fifty properties already, the DBpedia ontology can be extended by the collaboration of the DBpedia community on the DBpedia Mappings Wiki (Lehmann et al. 2012: 9). It is possible because, to support users in their work, the DBpedia Mappings Wiki offers amongst its tools the Mapping Tool, which "is a graphical user interface that supports users to create and edit mappings" (Lehmann et al. 2012: 6).

Furthermore, besides providing mappings from infoboxes, DBpedia adopts a live synchronization approach. Since Wikipedia encyclopaedic information is continuously revised, DBpedia data has needed to develop a system to keep up-to-date. As a consequence, the DBpedia Live system has been developed. Amongst its major components useful for our purposes it is the Synchronisation Tool. In the words of Lehmann et al., the Synchronisation Tool "allows third parties to keep DBpedia Live mirrors up-to-date by harvesting the produced change sets". In this vein, DBpedia is interlinked with numerous external data sets. External data sets may be interlinked from DBpedia or, conversely, that set-level outgoing links may point at DBpedia resources (Lehmann et al. 2012: 14-15).

In the following sections, after having mapped the Wikipedia infobox properties to those of the DBpedia ontology, it is time for the transference process of structured information from the DBpedia ontology to the Onomasticon. The transference process from DBpedia to the Onomasticon is divided up in three phases. Firstly, COREL structures need to be built to serve as a conceptual receptacle of the encyclopaedic definitions. This phase is

[6] http: //mappings.dbpedia.org

made by means of a specific interface integrated in the FunGramKB editor platform. Secondly, a structured data dump from DBpedia to Onomasticon was completed. Finally, work has to be done to maintain the data of the COREL structures updated since some of the transferred information is not static but may suffer variations over time. Nevertheless, it can be observed that although this methodology is very helpful to acquire the vast amount of information, it does not include the acquisition of unstructured information. In other words, the data which are not organised under a pattern as that of the infoxes. In actual fact, the complex structures of the natural language need to be simplified in order to be translated and represented in COREL (Ureña Gómez-Moreno, 2014: 511, 512).

An illustration of how the FunGramKB community endeavours to work with the mappings of classes and their corresponding properties as well as its transfer to COREL structures by means of the conceptual editor DBpedia Mapper can be seen in Figure 11. In the upper part, the navigation area is divided into two lists: the list on the left hand side which contains the ontology of classes of DBpedia whereas the list on the right hand side shows a list of the properties corresponding to the class classified. In the lower part, the editor is intended to store the COREL structure ascribed to a specific property (Periñán Pascual & Carrión Varela, 2011:95).

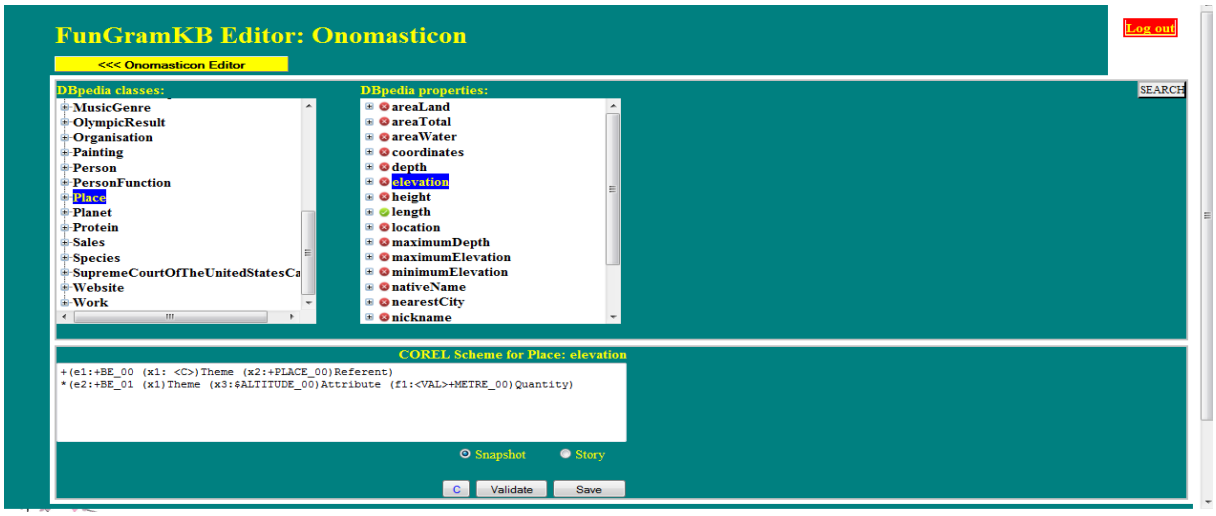

Figure 11. Interface of DBpedia Mapper in the editor of the Onomasticon

A sample of conceptualisation in natural language with its equivalent COREL structure of the organization EUROPOL has already been done by Ureña Gómez-Moreno (2014: 512):

"*Europol's aim is to improve the effectiveness and co-operation between the competent authorities of the member states primarily by sharing and pooling intelligence to prevent and combat serious international organized crime.*
*(e1:+BE_00(x1:%EUROPOL_00)Theme(x2:+ORGANIZATION_00)Referent)
*(e2: +HELP_00 (x1)Theme (x3: %EUROPEAN_UNION_00)Referent (f1: (e3: n +EXIST_00 (x4:+ORGANIZED_CRIME_00)Theme))Purpose)"

II. Methodology

With the aim of fulfilling the objectives presented in the introduction, the following methodology consisting of three phases was planned:

8. Phase 1.- Previous search

This phase is meant to gather information about the most relevant legal instruments and legal institutions which cooperate in the fight against organised crime and terrorism. Once these entities are selected, the next step is to see which ones are included in the sample documents of the GCTC. As it was explained in the theoretical part of this dissertation, terms that are not included in the GCTC yet, cannot be considered as part of the Onomasticon.

However, I looked for the highest number of terminological resources that referred to international and communitarian laws as well as institutions that combat against criminal and terrorist acts in an international as well as in a national setting. Two main sources were considered: academic reference works and international institutions websites. Academic reference works such as encyclopaedias were useful in my task. Besides, international institutions with official websites with free access particularly concerned with terrorism and organised crime issues such as the CCJ, Europol, Eurojust, UN and OSCE were also helpful.

Identifying a list of terms to be included in a specialised ontology as well as in a specialised onomasticon is not a straightforward task. The author´s academic background and professional experience as a jurist has been taken as a reliable methodological support. Actually, "*it is necessary to note that experts and practitioners are the best placed to*

determine and define specialised terms, since they are well acquainted with the jargon used in their professions" (Felices Lago, 2012: 60). Furthermore, the final decision concerning whether a legal instrument or a legal institution were to be included into the Onomasticon, a lexicographical criterion was adopted. The lexicographical criterion is based on the consultation of reliable specialised dictionaries such as online glossaries provided by institutions on their updated websites.

9. Phase 2: Term extraction from FGKBTE

This phase consists of using two tools of FGKBTE, namely "Search" and "Corpus".

9.1. "Search"

Once the relevant terms were selected to be included in the specialised Onomasticon, I sought them by means of the "search" engine to find out if they were included in the GCTC. The linguistic structures of the entities stored in the GCTC are easily located just by entering through the tab called "Search" of the FGKBTE, putting them into the lower toolbar and clicking on the button which is found besides the button entitled "search", as it can be seen in the screen shot in Figure 4 displayed above. In addition, it is worth noting that relevant terms can be found even when the query string is not in the candidate list or has otherwise been removed accidentally or not and thus considered a false candidate at any stage (Felices Lago & Ureña Gómez-Moreno, 2014: 266).

9.2. "Corpus"

The terms that could have been accidentally discarded during the filtering could be recovered with the aid of the option included in the tab "Corpus". The button "Recover a term" restores the selected terms to the original candidate list from which it was discarded (Felices Lago & Ureña Gómez-Moreno, 2014: 266). The deletion of terms may happen because the terminological extraction process in FGKBTE is semi-automatic and so contains a manual editing phase. It may happen because the terminologists make the final decision considering the classification of a term as specialised (Carrión Delgado, 2012: 204).

Thus, many n-grams may have been accidentally removed by another terminologist working on the project who wrongly thought that it did not have to belong to the legal specialised domain, might be confused because of the low tf-idf score. Figure 12 shows the list of terms which has been considered false candidates and so not proposed as candidate terms by the terminologist. It shows a list of false candidates which contains unigrams (incb), bigrams (counter centr) and trigrams (intern labour organ). The terminologist has to consider thoroughly which ones are going to be finally "winning terms" in order to belong to the legal specialised domain. Furthermore, the terminologist does not pay attention here to the tf-idf index of the terms because it is a statistical criterion which does not sometimes match the goal of building up a specialised dictionary (Carrión Delgado, 2012: 212). Moreover, in order to decide if a unigram, bigram or trigram has to be considered as a winning term, the terminologist needs to browse through Google, Wikipedia and specialised glossaries whether or not they are acronyms and initials (Carrión Delgado, 2012: 208).

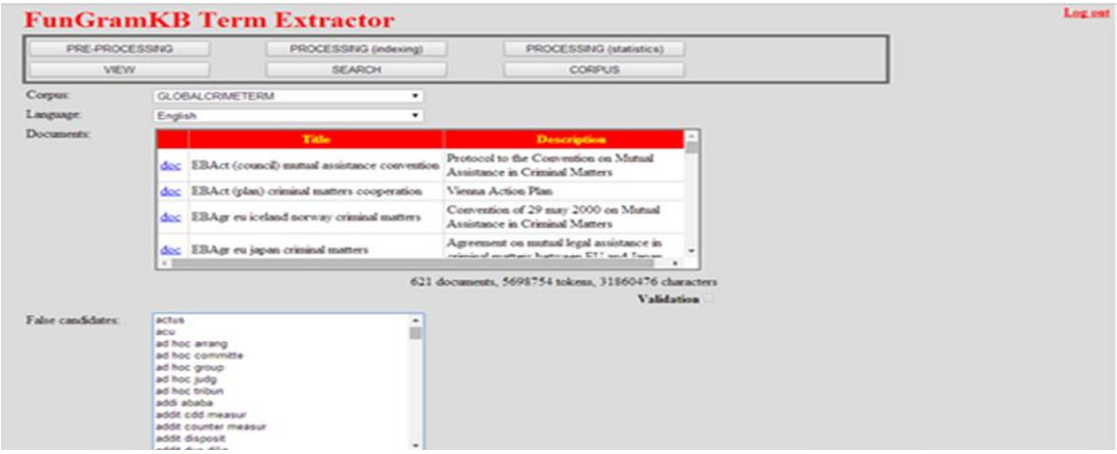

Figure 12. False candidates in the "Corpus"

10. Phase 3: Decisions upon the properties and definitions of the terms in natural language

The last step deals with the terminological aspects concerning the lexicological definitions of candidate terms. In order to help with the mappings from Wikipedia infobox properties to the DBpedia ontology, this dissertation uses the methodological strategy of performing an alignment or mapping between Wikipedia classes and properties and the DBpedia Ontology. Nevertheless, there are classes and properties which cannot be aligned in

DBpedia because similar classes are not found in DBpedia and so those will need to be created. Thus, emulating the DBpedia schemata is necessary.

Thus, to begin with, this research takes a wiki-based approach and finds out the most common property names from the different Wikipedia info-boxes of the main institutions involved in the fight against organised crime as well as the legal instruments available for avoiding those crimes and for pursuing and sanctioning wrongdoers.

An explanation of the properties to be culled needs to be done as follows. The relevance and appropriateness of the most common properties found in the Wiki infoboxes has to be in relation with its reusability for the tasks of NLP in FunGramKB, its grade of generality or prototype of the field, its capacity of expressivity and with technical motives of the used tool, amongst some other reasons that may appear ad hoc in every particular case (Carrión Varela, 2010: 41).

In this regard, amongst the most common properties, those related to the legal features of the legal institutions as well as those of the legal instruments are chosen. That criterion is held due to the assumed high specialisation of the legal domain. The prototypical properties found in Wikipedia, which we decided to map to properties in the DBpedia ontology and so to integrate in the Onomasticon are as follows. In the case of the legal organisations involved in the fight against criminal activities, the typical properties are “legal personality”, “purpose”, “formation”, “headquarters” and “membership”. Conversely, in the case of the area of the legal instruments involved in the fight against criminal activities, the typical properties are “type”, “effective” and “parties”.

In order to map those Wikipedia infobox properties to the ontology schema of DBpedia, a search of similar classes to “legal instrument” and “legal institution” is done in DBpedia Ontology. A search of it has been done by looking at the parameters `legal´ and `law´. Nevertheless, only the classes `legalcase´ and `lawfirm´ have been found, as it can be seen in Figures 13 and 14 respectively.

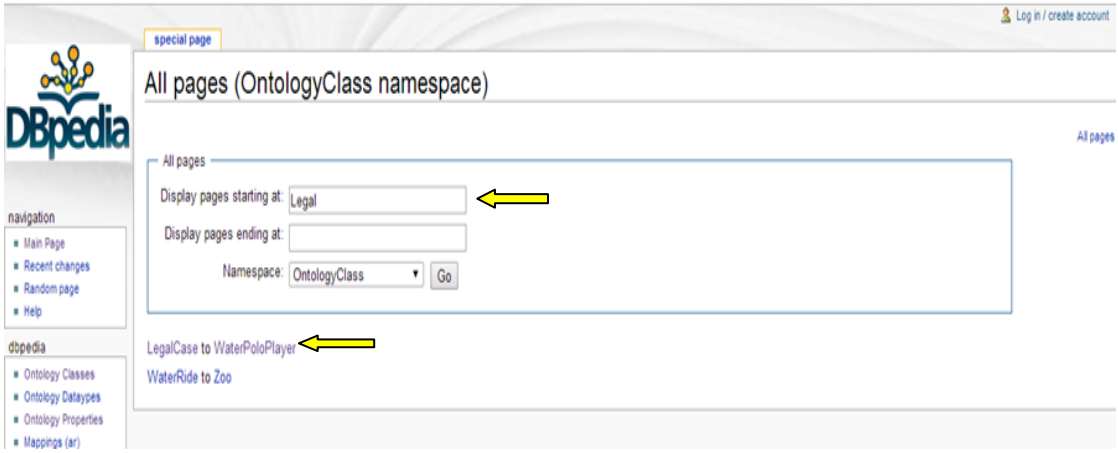

Figure 13. Result of the search "legal" among the DBpedia Ontology classes

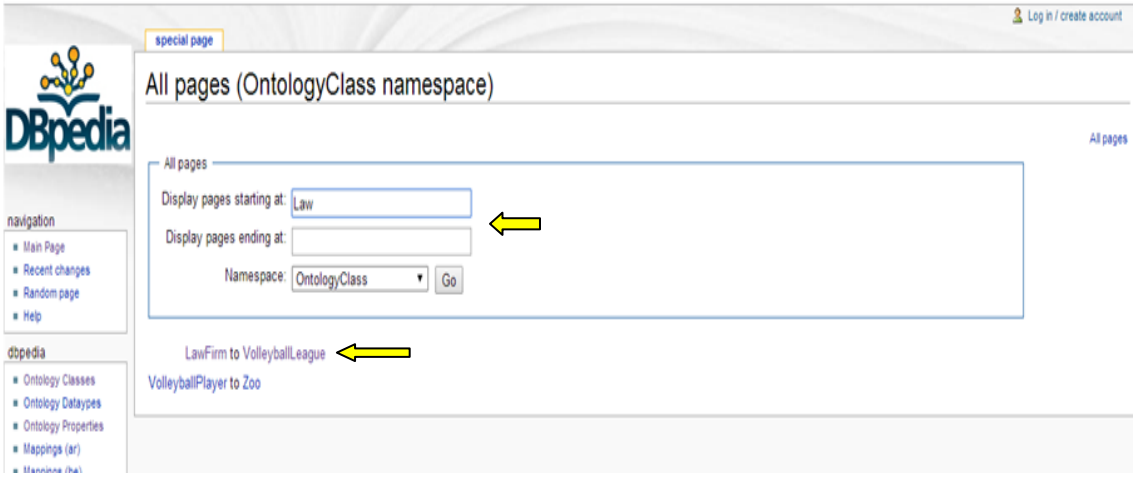

Figure 14. Result of the search "law" among the DBpedia Ontology classes

Just a class with a certain resemblance is found. A similarity may be shown between the class "organisation" included in the DBpedia ontology and the class "legal institution". Furthermore, a search of the subclasses of the class "organisation" is done. Amongst the subclasses which are useful to map "legal institutions" from Wikipedia are "GovernmentAgency" and "Parliament". However, subclasses such as "Judicial Body" are missed. Thus, the subclasses to which the properties are assigned to are not always included in the DBpedia ontology. Despite this fact, a thorough search of the properties of the class "organisation" in DBpedia was done. It gives us an idea of the kind of properties included in the DBpedia Ontology. An extract of the properties´ list can be seen in Figure 15. Furthermore, they were found through the browser of properties in the DBpedia ontology.

Properties on *Organisation*:

Name	Label	Domain	Range	Comment
affiliation (edit)	affiliation	Organisation	Organisation	
age (edit)	age	Agent	*xsd:integer*	
artPatron (edit)	patron (art)	Agent	Artist	An influential, wealthy person who supported an artist, craftsman, a scholar or a noble.. See also
ceo (edit)	chief executive officer	Organisation	Person	
chairperson (edit)	chairperson	Organisation	Person	
childOrganisation (edit)	child organisation	Organisation	Organisation	
discipline (edit)	discipline	Agent	owl:Thing	
endowment (edit)	endowment	Organisation	*Currency*	
extinctionDate (edit)	extinction date	Organisation	*xsd:date*	
extinctionYear (edit)	extinction year	Organisation	*xsd:gYear*	
formationDate (edit)	formation date	Organisation	*xsd:date*	
formationYear (edit)	formation year	Organisation	*xsd:gYear*	
foundationPlace (edit)	foundation place	Organisation	City	
generalCouncil (edit)	general council	Agent	TermOfOffice	
headquarter (edit)	headquarter	Organisation	PopulatedPlace	
hometown (edit)	home town	Agent	Settlement	
honours (edit)	honours	Agent	owl:Thing	
ideology (edit)	ideology	Agent	Ideology	
juniorSeason (edit)	junior season	Agent	owl:Thing	

Figure 15. List of properties of the class "organisation"

From now on, the properties found and not found will be thoroughly examined. In particular, the properties "headquarters", "formationdate" and "membership" belong to the class "organisation" in the DBpedia ontology. Figures 16, 17 and 18 show the properties "headquarter", "foundationdate" and "membership" associated to the class "organisation". Moreover, within the DBpedia Ontology property information, "comments" might be found. "Comments" are meant to indicate a similar property of any other class or an explanation of the property we are dealing with, which turns out to be very useful for the representation of the knowledge in COREL in a later stage (Periñán Pascual, 2011: 94). For instance, Figure 16 shows the similar property "foundingdate" of "formation date".

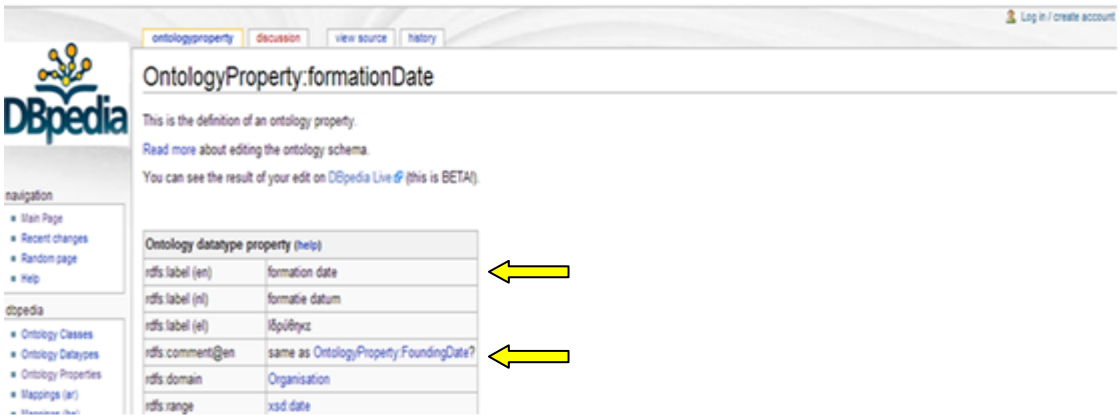

Figure 16. Property "formationdate" similar to the property "foundationdate" of the class "organisation" included in the toolbar "comments"

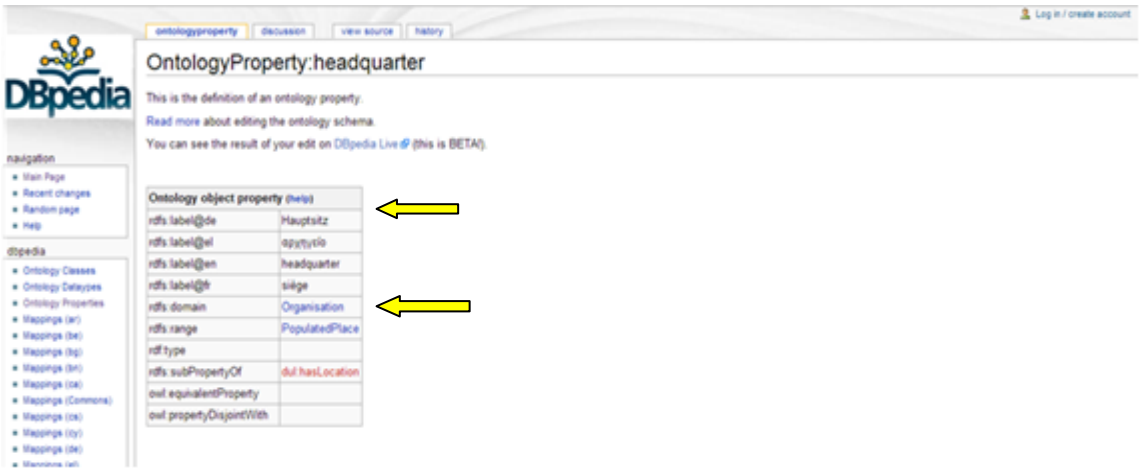

Figure 17. Property "headquarter" of the class "organisation"

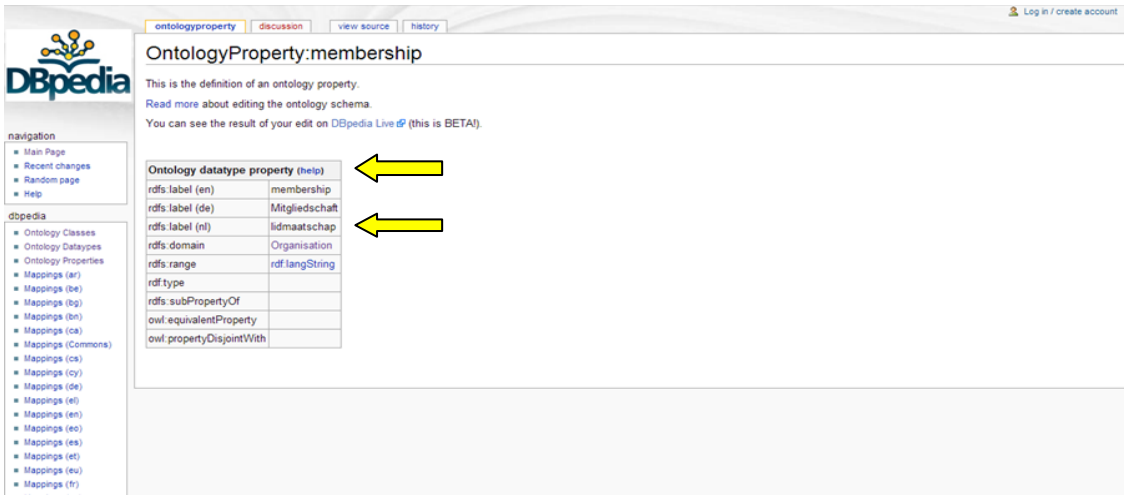

Figure 18. Property "membership" of the class "organisation"

However, although the property "purpose" appears as belonging to the hyperordinate "thing", is not included as a property of the class "organisation", as it can be seen in figure 19. The property "legal personality" is not included amongst the properties of the DBpedia ontology. Only the quite similar property name "legalform" pertains to the list of organisation´s properties. However, the property "legal form" is referred to as a business companies´ legal status, that is to say to this subclass. Figure 20 shows the "legalform" property of the class "organisation".

Figure 19. Property "purpose" of the class "thing"

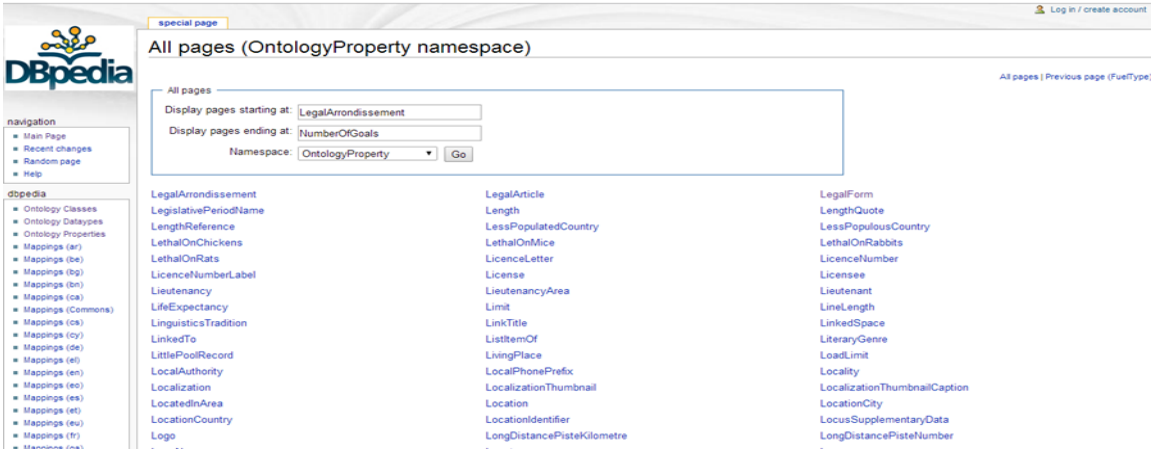

Figure 20. Property "legalform" of the class "organisation"

On the other hand, no class similar to "legal instrument" is found in DBpedia. It demands to be mapped in the DBpedia ontology in order to be included in the Onomasticon. It is convenient because it is a prototypical class of a satellite ontology based on law terminology and so for FunGramKB. Furthermore, it is ontologically practical for the purposes of NLP.

In addition, by getting into the DBpedia ontology you can also realize that the value of each property might be expressed in a different unity of measure and consequently DBpedia may give unequal amount of information of each. Thus, the results of the methodological process may show a different size of definitions for the different entities.

With regards to the definitions of the terms in Natural language, an intensive survey of encyclopaedic information is done through official web pages with free access, Wikipedia articles, journal articles, encyclopaedias of terrorism and of organised crime as well as specialised books on terrorism and organised crime.

IV. Results

11. Table of results

	Organised Crime	**Terrorism**	**Organised Crime and Terrorism**
Legal Institutions	**13**	**9**	**18**
1. International Institutions	**9**	**4**	**11**
1.1. UN	3	4	1
1.2. EU	4	———	3
1.3. Council of Europe	———	———	1
1.4. Other International Institutions	2	———	6
2.National Institutions	**4**	**5**	**7**
Legal Instruments	**13**	**6**	**6**
3.International Instruments	**11**	**4**	**4**
3.1. UN	6	4	1
3.2. EU	———	———	———
3.3. Council of Europe	2	———	2
3.4. Other International Institutions	3	———	1
4. National Legal Instruments	**2**	**2**	**2**

Table 1. Table of results

This section presents the numerical results of the defined terms after applying the methodology which has been described above. To begin with, the quantitative analysis shows that out of the total number of sixty-five terms defined, twenty six deal with organised crime, fifteen deal with terrorism, and twenty-four with organised crime and terrorism at the same time.

As the table below shows, nine out of the twenty-eight terms that referred to organised crime are legal institutions, out of which four are international institutions and five are international institutions. Moreover, thirteen terms are referred to as legal instruments, eleven out of the thirteen are international legal instruments and two are national legal instruments.

Regarding the terms referring to terrorism, nine out of fifteen are legal institutions, four international institutions and five national institutions in particular, whilst six are legal terms, out of which four are international instruments and only two are national.

Concerning the eighteen four terms referring to organised crime and terrorism at the same time, eighteen are institutions, eleven are international and the remaining seven are national. On the other hand, six terms are legal instruments, four are international and two are national.

12. Discussion

In this section, the key results are analysed in the light of the theoretical framework of the GCTC and the research methodology described in the sections 3 and 9.1 respectively. I do that bearing in mind the two main criteria informing the GCTC, namely balance and representativeness. In this regard, the corpus gathered here complies with the criteria of containing a fair representation among all the specific fields since 40% deals with organised crime, 23% deals with terrorism, and the remaining 37% deals with these two topics simultaneously. In addition, a rich variety of texts is assigned to the corpus. It can be observed by looking at the percentages. 37% is referred to international institutions, 25 % concerns national institutions, 29% is about international legal instruments and the remaining 9% is related to national legal instruments. Thus, results of this analysis show that the corpus meets the requirement of representativeness.

13. Glossary

13. 1. Properties and Conceptualisation of the institutions and the legal instruments dealing with the fight against organised crime

Table 2. Institutions of the UN dealing with the fight against organised crime

	Terms	Corel language	Legal personality	Purpose	Formation date	Headquarters	Membership	Description in Natural Language
1.	**International Narcotics Control Board (INCB)**	%INCB_00	UN Judicial body	To implement UN international drug control conventions	1968	New York (USA)	13	The INCB is the independent and quasi-judicial monitoring body for the implementation of the UN international drug control conventions. It was established in 1968 in accordance with the Single Convention on Narcotic Drugs, 1961. It consists of thirteen members elected by the UN Economic and Social Council (ECOSOC henceforth).
2.	**International Labour Organization (ILO)**	%ILO_00	UN Agency	To deal with labour issues	1919	Geneva (Switzerland)	185	The ILO is a UN agency dealing with labour issues, particularly international labour standards and decent work for all. It registers complaints against entities that are violating international rules. One hundred eighty five of the one hundred ninety three UN member states are members of the ILO. It was formed on 1919. The ILO´s headquarters is in Geneva (Switzerland).
3.	**Commission on Narcotic Drugs (CND)**	%CND_00	UN Agency	To decide drug policy-making	1946	New York (USA)	53	The CND is one of the functional commissions of the UN Economic and Social Council (ECOSOC henceforth) and is the central drug policy-making body within the UN system. It was established in 1946. Currently, it consists of fifty three states, serving 4-year terms.

Table 3. Institutions of the EU dealing with the fight against organised crime

	Terms	Corel language	Legal personality	Purpose	Formation date	Headquarters	Membership	Description in Natural Language
4.	**European Anti-fraud Office (OLAF)**	%OLAF_00	EU Agency	To fight fraud affecting the EU budget	1999	Brussels (Belgium)	28	The OLAF is charged by the EU with protecting its financial interests. Its tasks are to fight fraud affecting the EU budget as well as corruption and any other irregular activity within the institutions of the EU. OLAF was created in 1999. Its headquarters is in Brussels (Belgium).Currently, it covers all twenty eight member states of the EU.
5.	**European Police Office (Europol)**	%EUROPOL_00	EU Agency	To combat serious international organized crime	1 July 1999	The Hague (Netherlands)	28	EUROPOL is the EU's law enforcement agency that handles criminal intelligence by sharing and pooling intelligence to prevent and combat serious international organised crime. It became fully operational on 1 July 1999. Its seat is in The Hague (Netherlands). Currently, Europol covers all twenty eight member states of the EU.
6.	**Eurojust**	%EUROJUST_00	EU Agency	To handle cross-border organized crime	28 February 2002	The Hague (Netherlands)	28	Eurojust is an agency of the EU dealing with judicial co-operation in criminal matters. It handles serious cross-border and organised crime by stimulating investigative and prosecutorial co-ordination among agencies of the EU Member States. It was established on 28 February, 2002. Its seat is in The Hague (Netherlands).Eurojust is composed of a College

								formed of twenty eight national members (judges, prosecutors or police officers) from each EU member state.
7.	**Horizontal Drugs Group (HDG)**	%HDG_00	EU Agency	To deal with the fight against drugs	1990	Brussels (Belgium)	28	The HDG can be described as the European Council´s main technical, policy and co-ordination forum to facilitate the joint efforts of the member states and the European Commission in the joint fight against drugs. It has been actively engaged in the fight against drug dealing and use since 1990. Its headquarters is in Brussels (Belgium). It covers all twenty eight member states of the EU.

Table 4. Other international institutions dealing with the fight against organised crime

	Terms	Corel language	Legal personality	Purpose	Formation date	Headquarters	Membership	Description in Natural Language
8.	**Organization for Economic Cooperation and Development (OECD)**	%OECD_00	International organisation	To fight against organised crime	30 September 1061	Paris (France)	34	The OECD is an international economic organisation of 34 countries founded on 30 September 1961. It fights against organised crimes such as corruption. Its headquarters is in Paris (France). It is composed of thirty four countries.
9.	**Inter-American Drug Abuse Control Commission (CICAD)**	%CICAD_00	OAS Commission	To control the production, abuse, and traffic in illicit drugs and related crimes	November 1986	Washington D.C. (USA)	34	The CICAD was established in November 1986 and is composed of thirty four member states of the OAS. Its purpose is to facilitate multilateral cooperation to control the production, abuse, and traffic in illicit drugs and related crimes. Its headquarters is in Washington D.C. (USA).

Table 5. National institutions dealing with the fight against organised crime

	Terms	Corel language	Legal personality	Purpose	Formation date	Headquarters	Description in Natural Language
10.	**Criminal Assets Bureau (CAB)**	%CAB_00	Irish agency	To penalise the criminally acquired assets	1996	Dublin (Ireland)	The CAB is a law enforcement agency in Ireland. Its aims are to identify the criminally acquired assets of persons and to take the appropriate action to deny such people of these assets. The CAB was established in 1996. Its headquarters is in Dublin (Ireland).
11.	**Drug Enforcement Administration (DEA)**	%DEA_00	US agency	To combat drug smuggling	1 July 1973	Arlington, Virginia (USA)	The DEA is a US federal law enforcement agency under the U.S. Department of Justice, tasked with combating drug smuggling and use within the US. It was formed on 1 July, 1973. Its headquarters is in Arlington county, Virginia (USA).
12.	**Financial Intelligence Unit-India (FIU-IND)**	%FIU_IND_00	Indian agency	To strengthen efforts against money laundering	18 November 2004	New Delhi (India)	The FIU-IND was set by the Government of India on 18 November 2004. It is an agency responsible for coordinating and strengthening efforts of national and international intelligence, investigation and enforcement agencies in pursuing the global efforts against money laundering and related crimes. Its headquarters is in New Delhi (India).

13.	**Narcotics Control Bureau (NCB)**	%NCB_00	Indian agency	To fight drug trafficking	17 March 1987	New Delhi (India)	The NCB is the chief law enforcement and intelligence agency of India responsible for fighting drug trafficking and the abuse of illegal substances. It was created on 17 March 1986. Its headquarters is located in New Delhi (India).

Table 6. Legal instruments from the UN dealing with the fight against organised crime

	Terms	Corel language	Purpose	Effective	Parties	Description in Natural Language
14.	**UN Convention against transnational organised crime (UNTOC or Palermo Convention)**	%UNTOC_00	To fight Transnational Organised crime	29 September 2003	179	The UNTOC is a UN treaty against transnational organized crime. It contains elements of the current international law on human trafficking, arms trafficking and money laundering. The Convention came into force on 29 September 2003. Currently, it has been ratified by one hundred seventy nine UN Member States.
15.	**Protocol against the Smuggling of Migrants by Land, Sea and Air (Smuggling Protocol)**	%SMUGGLING_PROTOCOL_00	To fight against the Smuggling of Migrants	28 January 2004	139	The Smuggling Protocol is a UN General Assembly that aims to fight against organized criminal groups that abuse migrants. It is one of the three Palermo Protocols. It came into force on 28 January 2004. As of August 2014, the protocol has been ratified by one hundred thirty nine UN Member States.
16.	**Protocol against the Illicit Manufacturing and Trafficking in Firearms, Their Parts and Components and Ammunition (Firearms Protocol)**	%FIREARMS_PROTOCOL_00	To fight arms trafficking	3 July 2005	110	The Firearms Protocol is an anti-arms trafficking treaty that is supplemental to the Convention against Transnational Organized Crime. It is one of the three Palermo Protocols. It was adopted by the UN General Assembly and it entered into force on 3 July 2005. As of August 2014 it has one hundred ten parties, including one hundred and nine states and the European Union.
17.	**Protocol to Prevent, Suppress and Punish Trafficking in Persons, especially Women and Children (Trafficking Protocol)**	%UN_TIP_PROTOCOL_00	To combat transnational organised crime	25 December 2003	161	The Trafficking Protocol is a protocol to the Convention against Transnational Organised Crime. It is one of the three Palermo protocols. It commits of ratifying states to prevent and combat trafficking in persons, protecting and assisting victims of trafficking. It entered into force on 25 December 2003. It has been ratified by one hundred sixty one UN Member States.
18.	**UN Convention against corruption (UNCAC)**	%UNCAC_00	To criminalise corruption	14 December 2005	171	The UNCAC is a UN international anti-corruption instrument which criminalises certain conducts, strengthens international law enforcement and judicial cooperation, and provides effective legal mechanisms for asset recovery, technical assistance and information exchange. It entered into force on 14 December 2005. Currently, one hundred seventy one UN Member States have ratified it.
19.	**International Convention for Suppression of Counterfeiting Currency**	%INTERNATIONA_ CONVENTION_FOR_ THE_SUPPRESSION_ OF_COUNTERFEITING _CURRENCY_00	To criminalise counterfeiting	22 February 1931	80	The International Convention for the Suppression of Counterfeiting Currency is a UN treaty whereby states agree to criminalise acts of currency counterfeiting. It entered into forced on 22 February 1931. Currently, it has been ratified by eighty member states.

Table 7. Legal instruments of the Council of Europe dealing with the fight against organised crime						
	Terms	**Corel language**	**Purpose**	**Effective**	**Parties**	**Description in Natural Language**
20.	**Convention on Cybercrime (Budapest Convention on Cybercrime or the Budapest Convention**	%BUDAPEST_CONVENTION_00	To address Internet and computer crime	1 July 2004	42	The Budapest Convention is an international treaty seeking to address Internet and computer crime. It was adopted by the Committee of Ministers of the Council of Europe and it entered into force on 1 July 2004. It has been ratified by forty two states.
21.	**Council of Europe Convention on the Protection of Children against Sexual Exploitation and Sexual Abuse**	%COUNCIL_OF_EUROPE_CONVENTION_ON_THE_PROTECTION_OF_CHILDREN_AGAINST_SEXUAL_EXPLOITATION_AND_SEXUAL_ABUSE	To criminalise children sexual exploitation and sexual abuse	1 July 2010	47	The Council of Europe Convention on the Protection of Children against Sexual Exploitation and Sexual Abuse is a multilateral Council of Europe treaty whereby states agree to criminalise sexual abuse and sexual exploitation against children. It entered into force on 1 July 2010. It has been ratified by forty seven member states.

Table 8. Legal Instruments of other International Institutions fighting against organised crime						
	Terms	**Corel language**	**Purpose**	**Effective**	**Parties**	**Description in Natural Language**
22.	**Convention on combating Brivery of Foreign Public Officials in International Business Transactions (OECD Anti-Brivery Convention)**	%OECD_ANTI_BRIVERY_CONVENTION_00	To penalize bribery in international business transactions	15 February 1999	46	The OECD Anti-Brivery Convention penalizes bribery in international business transactions carried out by companies based in the Convention member countries. It came into forced on 15 February 1999. Currently, forty six countries have ratified it (forty OCDE States Parties and six Member States participate in the OEC Working Group in Brivery).
23.	**Inter-American Convention Against Terrorism**	%INTER_AMERICAN_CONVENTION_AGAINST_TERRORISM_00	To fight against transnational organised crime and terrorism	6 July 2003	35	The Inter-American Convention Against Terrorism was adopted by the member countries of the OAS and came into force on 6 July 2003. Amongst its priorities are to elaborate legal tools against transnational organised crime and to improve regional cooperation in the fight against terrorism. It has been ratified by the thirty five member states of the OAS.
24.	**Convention concerning the Prohibition and Immediate Action for the Elimination of the Worst Forms of Child Labour (Worst Forms of Child Labour Convention)**	%WORST_FORMS_OF_CHILD_LABOUR_CONVENTION_00	To combat commercial exploitation of children	19 November 2000	179	The Worst Forms of Child Labour Convention was adopted by the ILO and came into force the 19 November 2000. It fights against the commercial exploitation of children. It has been ratified by one hundred seventy nine ILO Member States.

Table 9. National Laws in the fight against organised crime						
	Terms	**Corel language**	**Purpose**	**Effective**	**Parties**	**Description in Natural Language**
25.	**Anti-Trafficking in Persons Act of 2003, No. 9208**	%ANTI_TRAFFICKING_IN_PERSONS_ACT_OF_2003_R.A_NO_9208_00	To punish human trafficking	12 May 2003	Parliament of Philippines	The Anti-Trafficking in Persons Act of 2003, R.A. No. 9208 was enacted and passed by Congress of the Philippines, Senate of the Philippines and House of Representatives of the Philippines assembled on 12 May, 2003. It institutes policies to eliminate and punish human trafficking.
26.	**Prevention of Money Laundering Act, 2002**	%PREVENTION_OF_MONEY_LAUNDERING_ACT_2002_00	To confiscate properties derived from money-laundering	1 July 2005	Parliament of India	The Prevention of Money Laundering Act, 2002 is an Act of the Parliament of India enacted to prevent money-laundering and to provide for confiscation of property derived from money-laundering which came into force on 1 July 2005.

13.2. Properties and Conceptualisation of the institutions and the legal instruments dealing with the fight against terrorism

Table 10. Institutions of the UN dealing with the fight against terrorism

	Terms	Corel language	Legal personality	Purpose	Formation date	Headquarters	Membership	Description in Natural Language
27.	**Counter-Terrorism Committee (CTC)**	%CTC_00	UN Agency	To bolster the ability of UN Member States to fight terrorism	28 September 2001	New York (USA)	15	The CTC is a subsidiary body of the UN Security Council. The ultimate aim is to bolster the ability of UN Member States to fight terrorism within their borders and across regions. It was formed on 28 September 2001. It is composed of fifteen member states.
28.	**Counter-Terrorism Implementation Task Force (CTITF)**	%CTITF_00	UN Agency	To coordinate counter-terrorism efforts	2005	New York (USA)	34	The CTITF was established by the UN Secretary General in 2005. Its mandate is to enhance coordination and coherence of counter-terrorism efforts of the United Nations system. The Task Force consists of thirty four international entities.
29.	**Counter-Terrorism Action Group (CTAG)**	%CTAG_00	UN Agency	To support the UN Security Council Counter-Terrorism Committee	July 2003	Meet three times a year in different places	G8, European Commission, an a representative from the CTC	The CTAG primarily supports the UN Security Council CTC, by coordinating donations to counter-terrorism capacity building assistance. Amongst its members are the G8, the European Commission, and a representative from the UN Security Council Counter-Terrorism Committee. It was formed in July 2003. It meets three times a year in different places around the Globe.
30.	**United Nations Security Council (UNSC)**	%UNSC_00	UN Agency	To fight against terrorism	17 January 1946	New York (USA)	15	The UNSC is one of the six principal organs of the United Nations and is charged with the maintenance of international peace and security. It is the only UN body with the authority to issue binding resolutions to member states. It held its first session on 17 January 1946. It headquarters is in New York (USA). It is composed of fifteen Member States, five of them which are permanents (China, France, Russia, United Kingdom and United States) and the rest are non-permanent.

Table 11. National institutions dealing with the fight against terrorism

	Terms	Corel language	Legal personality	Purpose	Formation date	Headquarters	Description in Natural Language
31.	**National Counterterrorism Center (NCTC)**	%NCTC_00	US Agency	To fight against terrorism	August 2004	McLean, Virginia (USA)	The NCTC is a US government organization responsible for national and international counterterrorism efforts. It was established in August 2004. Its headquarters is in McLean, Virginia (USA).
32.	**National Counter Terrorism Security Office (NaCTSO)**	%NaCTSO_00	UK Police	To contribute to counter-terrorism	1948	London (UK)	The NaCTSO is a UK police unit and contributes to the UK government's counter terrorism strategy. It was formed in 1948. Its headquarters is in London (UK).
33.	**Counterterrorism and Criminal Exploitation Unit**	%COUNTERTERRORISM_AND_CRIMINAL_EXPLOITATION_UNIT_00	US Agency	To prevent terrorists and other criminals from fraud	1 March 2003	Washington D.C. (USA)	The Counterterrorism and Criminal Exploitation Unit is part of U.S. Immigration and Customs Enforcement´s (ICE) Homeland Security Investigations´ (HSI) National Security Investigations Division, and prevents terrorists and other criminals from exploiting the nation's immigration system through fraud. It was established on 1 March 2003. Its headquarters is in Washington D.C. (US).
34.	**Australian Security Intelligence Organisation**	%AUSTRALIAN_SECURITY_INTELLIGENCE_ORGANISATION_00	Australian Agency	To protect the country and its citizens from terrorism	16 March 1949	Canberra (Australia)	The Australian Security Intelligence Organisation is responsible for the protection of the country and its citizens from terrorism. It was formed in 16 March 1949. Its headquarters is located in Canberra (Australia).

35.	**U.S. Department of Homeland Security (DHS)**	%DHS_00	US Agency	To protect the US from and to respond to terrorist attacks	25 November 2002	Washington, D.C. (USA)	The DHS is a cabinet department of the US federal government with the primary responsibilities of protecting the US from and responding to terrorist attacks. It was established on 25 November 2002. It has its headquarters in Washington, D.C. (USA).

Table 12: Legal Instruments of the UN dealing with the fight against terrorism

	Terms	Corel language	Purpose	Effective	Parties	Description in Natural Language
36.	**Terrorist Convention for the suppression of the financing of terrorism (Terrorist Financing Convention)**	%TERRORIST_ FINANCING_ CONVENTION_00	To criminalise acts of financing terrorist activities	10 April 2002	186	The Terrorist Financing Convention is a UN treaty designed to criminalise acts of financing terrorist activities. It entered into force on 10 April 2002. It has been ratified by one hundred eighty six states, which includes all but eleven UN Member States plus the Cook Islands, the Holy See, and Niue.
37.	**International Convention for the Suppression of Acts of Nuclear Terrorism (Nuclear Terrorism Convention)**	%NUCLEAR_ TERRORISM_CON VENTION_00	To criminalise acts of nuclear terrorism	7 July 2007	95	The Nuclear Terrorism Convention is a UN treaty designed to criminalise acts of nuclear terrorism. It entered into forced on 7 July 2007. It has been ratified by 95 UN member states.
38.	**UN Security Council resolution 1269**	%UN_SECURITY _COUNCIL_ RESOLUTION_ 1269_00	To call upon states to implement anti-terrorist conventions	19 October 1999	15	UN Security Council resolution 1269, adopted unanimously on 19 October 1999, called upon its 15 states to fully implement anti-terrorist conventions.
39.	**UN Security Council Resolution 1373**	UN_SECURITY _COUNCIL_ RESOLUTION_ 1373_00	To fight against terrorism	28 September 2001	193	UN Security Council Resolution 1373, adopted unanimously on 28 September 2001, aimed to hinder terrorist groups by imposing measures to its one hundred ninety three UN Member States as it is to share their intelligence on terrorist groups.

Table 13. National Laws dealing with the fight against terrorism

	Terms	Corel language	Purpose	Effective	Parties	Description in Natural Language
40.	**Canadian Anti-terrorism Act**	CANADIAN_ANTI_ TERRORISM_ACT_ 00	To bolster Canadian government and institutions to fight the terrorism	18 December, 2001	Canada Parliament	The Canadian Anti-terrorism Act gives powers to the Canadian government and to institutions within the Canadian security establishment to respond to the threat of terrorism. It was enacted by the Parliament of Canada and assented on 18 December 2001.
41.	**Homeland Security Act**	HOMELAND_ SECURITY_ACT_00	To help with terrorist attacks	25 November, 2002	107th United States Congress	The Homeland Security Act helps to ensure the response time and preparedness of providers for terrorist attacks. It was enacted by the 107th United States Congress and came into force on November 25, 2002.

13.3. Properties and Conceptualisation of the institutions and the legal instruments dealing with the fight against organised crime and terrorism

Table 14. Institutions of the UN dealing with the fight against organised crime and terrorism

	Terms	Corel language	Legal personality	Purpose	Formation date	Headquarters	Membership	Description in Natural Language
42.	**United Nations Office on Drugs and Crime (UNODC)**	%UNODC_00	UN Agency	To prevent money laundering, organised Crime and terrorism.	1997	Vienna (Austria)	193	The UNODC is a United Nations agency that was established in 1997. Amongst the main themes that it deals with are: crime prevention, drug prevention, human trafficking and migrant smuggling, money laundering, organized crime and terrorism prevention. It is headquartered in Vienna (Austria). It is composed of members elected by all the one hundred ninety three UN Member States.

Table 15. Institutions of the EU dealing with the fight against oganised crime and terrorism

	Terms	Corel language	Legal personality	Purpose	Formation date	Headquarters	Membership	Description in Natural Language
43.	**European Police College (CEPOL)**	%CEPOL_00	EU Agency	To cooperate in the fight against crime	2000	Bramshill, Hampshire (England)	28	CEPOL is an agency of the EU that brings together senior police officers across Europe with the aim of encouraging cross-border co-operation in the fight against crime, maintenance of public security and law and order. It was established by Council Decision 2000/820/JHA in 2000 and its seat is in Bramshill, Hampshire (England). It is made up of representatives from the twenty eight EU Member States
44.	**European Crime Prevention Network (EUCPN)**	%EUCPN_00	EU Agency	To promote crime prevention	May 2001	Brussels (Belgium)	28	The EUCPN was set up in May 2001 by an EU Council Decision to promote crime prevention activity in Member States across the EU. The Network consists of a nominated National Representative from each of the twenty eight EU Member State, a Substitute Representative, and other crime prevention experts including practitioners and academics.
45.	**European Police Chiefs Task Force (EPCTF**	%EPCTF_00	Agency Council of Europe	To co-operate on cross-border crime	1999	Meets twice a year in different places	28	The EPCTF established at the Tampere European Council of 1999, can be described as a top-level forum for interpersonal communication among the police forces of all the twenty eight member states. It meets twice every year to exchange experience, best practices and information on current trends in cross-border crime, all doing this in close co-operation with Europol.

Table 16. Institutions of the Council of Europe dealing with the fight against organised crime and terrorism

	Terms	Corel language	Legal personality	Purpose	Formation date	Headquarters	Membership	Description in Natural Language
46.	**European Committee on Crime Problems (CDPC)**	%CDPC_00	Council of Europe Agency	To coordinate the Council of Europe´s activities in the field of crime prevention and crime control	1958	Strasbourg (France)	47	The CDPC was set up in 1958. It was entrusted by the Committee of Ministers the responsibility for overseeing and coordinating the Council of Europe's activities in the field of crime prevention and crime control. The CDPC meets at the headquarters of the Council of Europe in Strasbourg (France). It is composed of representatives of all the forty seven member states.

Table 17. Other International Institutions dealing with the fight against organised crime and terrorism								
	Terms	**Corel language**	**Legal personality**	**Purpose**	**Formation date**	**Headquarters**	**Membership**	**Description in Natural Language**
47.	**Financial Action Task Force (FATF)**	%FATF_00	International organisation	To combat money laundering and terrorism financing	1989	Paris (France)	36	The FATF is an intergovernmental organization founded in 1989 on the initiative of the G7. The purpose of the FATF is to develop policies to combat money laundering and terrorism financing. The FATF Secretariat is housed at the headquarters of the OECD in Paris (France). The FATF membership is currently made up of thirty-four member jurisdictions and two regional organisations.
48.	**Asia/Pacific Group on Money Laundering (APG or APGML)**	%APGML_00	International organisation	To combat money laundering and the financing of terrorism	1997	Sydney (Australia)	41	The APGML is an international organization that develops and promotes policies to combat money laundering and terrorist financing in the Asia/Pacific region. It was founded in 1997. Its headquarters is in Sydney (Australia). The APG consists of forty one members.
49.	**Caribbean Financial Action Task Force (CFATF)**	%CFTAF_00	International organisation	To deal with money laundering	1990	Trinidad y Tobago	21	The CFTAF is an organisation of states and territories of the Caribbean basin which have agreed to implement common counter-measures against money laundering. It was established in 1990s. The CFTAF consists of twenty one members.
50.	**Organization for Security and Co-operation in Europe (OSCE)**	%OSCE_00	International organisation	To prevent the spread of illegal weapons and to combat terrorism	1973	Vienna (Austria)	57	The OSCE is the world's largest security-oriented intergovernmental organisation. Its mandate includes issues such as arms control and counter-terrorism. It was established on July 1973.It has its headquarters in Vienna (Austria). It is composed of fifty seven participating states.
51.	**Organisation for the Prohibition of Chemical Weapons (OPCW)**	%OPCW_00	International organisation	To verify the compliance of Parties to the Chemical Weapons Convention	29 April 1997	Hague (Netherlands)	190	The OPCW is an intergovernmental organization that promotes and verifies the adherence to the Chemical Weapons Convention. It was formed on 29 April, 1997. It located in The Hague, Netherlands. It is composed of the one hundred ninety member states that have ratified the Convention of Chemical Weapons.
52.	**Middle East and North Africa Financial Action Task Force (MENAFATF)**	%MENATAF_00	International organization	To take measures to combat money laundering and the financing of terrorism	2004	The Kingdom of Bahrain	18	The MENATAF was established in 30th of November 2004. It encourages to take measures with regard to countering money laundering and the financing of terrorism The headquarters is in the Kingdom of Bahrain. It is composed of eighteen member states.

Table 18. National Institutions dealing with the fight against organised crime and terrorism							
	Terms	**Corel language**	**Legal personality**	**Purpose**	**Formation date**	**Headquarters**	**Description in Natural Language**
53.	**CBP**	%CBP_00	US Agency	To prevent terrorism and contraband, illegal drug trade and other contraband	1 March 2003	Washington, D.C.	CBP is the largest federal law enforcement agency of the United States Department of Homeland Security. It prevents terrorists and terrorist weapons from entering the United States but it is also responsible for apprehending individuals attempting to enter the United States illegally including those with a criminal record, stemming the flow of illegal drugs and other contraband. It has its headquarters in Washington, D.C. Its start took place on 1 March, 2003.
54.	**Financial Crimes Enforcement Network**	%FINCEN_00	US Agency	to combat domestic and international	April 25, 1990	Vienna,Virginia (USA)	The FinCEN is a bureau of the US Department of the Treasury that collects and analyzes information about financial transactions in order to combat domestic and international money laundering, terrorist financing, and other financial crimes. FinCEN was

	(FinCEN)			money laundering and terrorist financing,			established on April 25, 1990. Its headquarters is in Vienna, Virginia (USA).
55.	**Criminal Intelligence Service Canada (CISC)**	%CISC_00	Canada Agency	To coordinate and share criminal intelligence amongst member police forces	1970	Ottawa (Canada)	The CISC is an inter-agency organization in Canada designed to coordinate and share criminal intelligence amongst member police forces. Established in 1970, the CISC has a central bureau in Ottawa (Canada).
56.	**Canadian Security Intelligence Service (CSIS)**	%CSIS_00	Canadian Agency	To deal with terrorism and cybercrime	1995	Ottawa (Canada)	The CSIS is Canada´s primary national intelligence service. Amongst its priorities, counter-terrorism and cybercrime are found. Its headquarters is located in Ottawa, Canada. It was formed in 1995.
57.	**Central Intelligence Agency (CIA)**	%CIA_00	US Agency	To deal with organized crime and terrorism	September 18, 1947	Langley, Virginia (USA)	The CIA is one of the principal intelligence-gathering agencies of the United States federal government. It was formed on September 18, 1947 The CIA's headquarters is in Langley, Virginia.
58.	**Federal Bureau Investigation (FBI)**	%FBI_00	US Agency	To protect the US from terrorist attacks and to combat transnational and national criminal organizations	26 July 1908	Washington, D.C. (USA)	The FBI is a governmental agency belonging to the United States Department of Justice that serves as both a federal criminal investigative body and an internal intelligence agency (counterintelligence). Among its priorities are to protect the US from terrorist attacks and to combat transnational and national criminal organizations and enterprises. It is headquartered in Washington, D.C, and it is formed on 26 July, 1908.
59.	**National Investigation Agency (NIA)**	%NIA_00	Indian Agency	To combat terror and organized crime	31 December 2008	New Delhi (India)	The NIA is a federal agency established by the Indian Government to combat terror and organized crime in India. It came into existence on 31 December 2008. Its headquarters is in New Delhi.

Table 19. Legal Instruments of the UN dealing with the fight against organised crime and terrorism

	Terms	**Corel language**	**Purpose**	**Effective**	**Parties**	**Description in Natural Language**
60.	**Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction (Chemical Weapons Convention or CWC)**	%CWC_00	To control the production, stockpiling and use of chemical weapons	19 April 1997	190	The CWC is an arms control treaty which outlaws the production, stockpiling, and use of chemical weapons and their precursors ratified by one hundred ninety UN Member States. It entered into force the 19 April 1997.

Table 20. Legal Instruments of the Council of Europe fighting against organised crime and terrorism						
	Terms	**Corel language**	**Purpose**	**Effective**	**Parties**	**Description in Natural Language**
61.	**European Convention on Mutual Assistance in Criminal Matters**	%EUOPEAN_ CONVENTION _ON_MUTUAL_ ASSISTANCE_ IN_CRIMINAL_ MATTERS_00	To promote assistance in prosecuting criminal suspects	12 June 1962	47	The European Convention on Mutual Assistance in Criminal Matters is a Council of Europe mutual legal assistance treaty. The parties to the Convention agree to offer each other mutual assistance in investigating crimes, procuring evidence, and in prosecuting criminal suspects. The Convention entered into force on 12 June 1962. It has been ratified by all forty seven member states of the Council of Europe.
62.	**Convention on Laundering, Search, Seizure and Confiscation of the Proceeds from Crime and on the Financing of Terrorism (Warsaw Convention or CETS 198)**	%WARSAW_ CONVENTION_00	To prevent and control the money laundering and the financing of terrorism	1 May 2008	47	The Warsaw Convention is a Council of Europe convention which aims to cover both the prevention and the control of money laundering and the financing of terrorism. It entered into force on 1 May 2008. It has been ratified by all forty seven member states of the Council of Europe.

Table 21. Legal Instruments from other International Institutions dealing with the fight against organised crime and terrorism						
	Terms	**Corel language**	**Purpose**	**Effective**	**Parties**	**Description in Natural Language**
63.	**Inter-American Convention Against the Illicit Manufacturing of and Trafficking in Firearms, Ammunition, Explosives, and Other Related Materials (CIFTA)**	%CIFTA_00	To control authorizations of export of Firearms and other related materials	1998	31	The CIFTA is an OAS international firearms control treaty. Amongst its priorities are to emphasise the need for authorizations or licenses of export, imports and transit. It came into force in 1998. The treaty has been ratified by thirty one of the thirty four states in the OAS.

Table 22: National Laws dealing with the fight against organised crime and terrorism						
	Terms	**Corel language**	**Purpose**	**Effective**	**Parties**	**Description in Natural Language**
64.	**Anti-Money Laundering and Countering Financing of Terrorism Act 2009 (AML/CFT Act)**	%AML_CFT_ACT_00	to deter money laundering and terrorism financing.	16 October 2009	Parliament of New Zealand	The AML/CFT Act places obligations on New Zealand's financial institutions and casinos to detect and to deter money laundering and terrorism financing. The Parliament of New Zealand enacted it the 16 October 2009.
65.	**Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act of 2001 (USA PATRIOT Act)**	USA_PATRIOT_ACT_00	To prevent and prosecute terrorism and international money laundering	26 October 2001	United States Congress	The USA Patriot Act is an Act of Congress that was signed into law on October 26, 2001. Amongst its provisions are to authorise measures to enhance the ability of domestic security services to prevent terrorism as well as to facilitate the prevention, detection and prosecution of international money laundering and the financing of terrorism.

V. Conclusions and future avenues of research

Generally speaking, this study has focused on the conceptual module of the FunGramKB, particularly in the Onomasticon component. This main objective of this dissertation has been to contribute to the future building of the Onomasticon and to define the terms in natural language included in it with the aim of fully integrating those in FunGramKB. The onomastical concepts proposed here would be compatible with a legal satellite ontology based on the fight against organised crime and terrorism.

To begin with, it was necessary to describe the context where this research comes from. Firstly, a description of what an ontology is in the field of knowledge engineering was provided. In parallel, the construction of the terminological subontology in natural language about international cooperation and the fight against organised crime and terrorism has been taken as a solid basis for this study. Besides, familiarity with the frame of the organised crime and terrorism offenses are central for the understanding of the outputs. To make the construction of the terminological subontology possible, the tailor-made corpus GCTC and the legal satellite ontology included in FunGramKB are explained. In addition, this dissertation has tried to explain the crucial role of FunGramKB in processing natural languages, mainly in the area of terminology. The extraction of data is done with the aid of the FGKBTE, which has been proved to be very helpful for the semi-automated retrieval of specialised terms. Furthermore, the DBpedia schemata have turned out to be helpful for the building of the Onomasticon.

With regards to the methodology used to facilitate the population of the Onomasticon, three steps were developed in the present dissertation. It started with an account of how the institutions and legal instruments which are adequate for the fight against international terrorism and organised crime offences were found in encyclopaedias and specialised sources. The second step covered the search of the terms included or excluded in the GCTC Satellite Ontology as well as the accidental deletion of candidate terms. Thirdly, the population process of the Onomasticon is performed with the support of the DBpedia project. The great informative value that DBpedia assigned to Wikipedia as well as the utility of organising knowledge within an ontology are also explained. At this stage, certain choices had to be made on the corpus in order to decide the properties related to “legal institutions” and “legal

instruments" entities. Thus, I had to identify which properties were relevant for the topics under scrutiny.

Additionally, the tables of relevant entries show that the corpus meets the requirements of balance and representativeness. In a similar manner, all the components in the GCTC, "organised crime", "terrorism" and "organised crime and terrorism" seem also to be balanced for the purposes in this dissertation. In the same vein, the corpus gathered ad hoc is representative. It is due to the fact it comes from a rich variety of texts containing information about institutions as well as about legal instruments at international or national level.

With regards to the DBpedia project, the DBpedia Mapper is particularly useful to contribute to the population of the Onomasticon. However, it lacks more thematic or content divisions. In this vein, it lacks classes such as "legal instruments". On top of that, concerning the necessary class for the purposes of this research "organization", it lacks its correspondent subclass "judicial body" as well as the properties "purpose" and "legal personality". Thus, the available information of classes and properties is incomplete and so it may be detrimental to the exploitation of that source. On the other hand, possibilities for further improvements include the creation of more mappings in the DBpedia Ontology by means of the DBpedia Mapping Tool. It may hold great potential for future work in the extension of the DBpedia ontology by the community in the DBpedia Mappings Wiki.

Thus far, the contribution of this dissertation to the development of the subontology on international organised crime and terrorism developed here as well as the building of the Onomasticon is determined to provide support to NLP tasks such as easy retrieval of legal information or the solution of problems. The goal is to go beyond the building of a simple repository of terminology and to provide a smart system which is capable of automatically relating written material from several sources included in FunGramKB. Professionals and scholars would benefit from the terms defined here in natural language once they are transduced into the COREL schemata and applied to diverse NLP tasks.

References

Bateman, J.A. (1991). "The theoretical status of ontologies in natural language processing". In Susanne Preuss y Birte Schmitz (eds.). *Text Representation and Domain Modelling - Ideas from Linguistics and AI, KIT-Report 97.* Berlin: Technische Universitaet Berlin, 50-99.

Cynthia Combs, C. and Slann, M. (2007). *Encyclopedia of Terrorism, Revision Edition.* United States of America: Facts on File.

Carrión Delgado, M.G. (2012). *Diseño y desarrollo de un modelo computacional para la representación del conocimiento en el dominio de la cooperación judicial en materia penal.* Universidad Nacional de Educación a Distancia. Ph.D.thesis.

Carrión Varela, M.Ll. (2010). *Representación formal del conocimiento enciclopédico en FunGramKB: el concepto PLACE*. Universidad Nacional de Educación a Distancia. Master thesis.

Carrión Varela, M.Ll. (2014). *Implementación de reglas de proyección conceptuales en el marco de la semántica profunda para la reutilización de bases de conocimiento enciclopédico.* Unpublished PhD. Dissertation. Madrid: Universidad Nacional de Educación a Distancia.

Felices Lago, A., Ureña Gómez-Moreno, P. & Alameda Hernández, A. (2011). "FunGramKB y la adquisición terminológica". *Anglogermánica Online*, 8, 66-86.

Felices Lago, A. (2012). "Creation of a terminological subontology in a multilingual context (English, Italian, Spanish) on criminal law: international cooperation against terrorism and organized crime base don FunGramKB". *Concurso de Acceso al Cuerpo Docente de Catedráticos de Universidad*. Unpublished Research Project. Universidad de Granada.

Felices Lago, A. & Ureña Gómez-Moreno, P. (2012). "Fundamentos metodológicos de la creación subontológica en FunGramKB". *Onomázein*, 26, 49-67.

Felices Lago, A. & Ureña Gómez-Moreno, P. (2014). "FunGramKB term extractor. A tool for building terminological ontologies from specialised corpora". In Nolan, B. and Periñán-Pascual, C. (eds.), *Language Processing and Grammars: The role of functionally oriented computational models*, 251–270.

Garrido García, N. (2010). *Modelling procedural knowledge in a knowledge base: An approach from the perspective of the lexical constructional model*. Universidad de la Rioja. Master thesis.

Gruber, T. R. (1993). "A translation approach to portable ontology specification". *Knowledge Acquisition*, 5 (2), 199-220.

Hahn, R. et al. (2010). "Faceted Wikipedia Search". *13th International Conference,* Berlin, Germany, May 3-5, 2010, 1-11. [http://semantic-web-journal.net/system/files/swj499.pdf]. Accessed on 14 September 2014.

Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., y Bizer, C. (2012). "DBpedia – A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia". *Semantic Web Journal*, 1-29.

Mairal Usón, R. & Periñán Pascual, C. (2009). "The anatomy of the lexicon component within the framework of a conceptual knowledge base". *Revista Española de Lingüística Aplicada,* 1-33.

Mairal Usón, R. & Periñán Pascual, C. (2010). "Role and Referente Grammar and Ontological Engineering". In J.L. Cifuentes et al (eds.). In *Los caminos de la lengua: Estudios en homenaje a Enrique Alcaraz Varó*. Alicante: Universidad de Alicante, 649-65.

Panton, K., Matuszek, C., Lenat, D., Schneider, D., Witbrock, M., Siegel, N. y Shepard, B. (2006). "Common sense reasoning - from Cyc to intelligent assistant". In Yang C. & Abascal, J. (eds.), *Ambient Intelligence in Everyday Life*. Berlin: Springer, 1-31.

Periñán Pascual, C. & Arcas Túnez, F. (2004). "Meaning postulates in a lexico-conceptual knowledge base". *15th Internacional Workshop on Databases and Expert Systems Applications*, IEEE, Los Alamitos (California), 38-42.

Periñán Pascual, C. & Arcas Túnez, F. (2005). "Microconceptual-Knowledge Spreading in FunGramKB". *Proceedings on the 9th IASTED International Conference on Artificial Intelligence and Soft Computing*, ACTA Press, Anaheim-Calgary-Zurich, 239-244.

Periñán Pascual, C. & Arcas Túnez, F. (2007). "Cognitive modules o fan NLP knowledge base for language understanding". *Procesamiento del Lenguaje Natural*, 39, 197-204.

Periñán Pascual, C. & Arcas Túnez, F. (2010). "The Architecture of FunGramKB". *Proceedings of the Seventh conference on International Language Resources and Evaluation*, European Language Resources Association (ELRA), 2667-2674.

Periñán Pascual, C. & Carrión Varela, M.Ll. (2011). "FunGramKB y el conocimiento cultural". *Anglogermánica Online*, 8, 87-105.

Periñán Pascual, C. & Arcas Túnez, F. (2014). "Knowledge engineering in the legal domain: The construction of a FunGramKB Satellite Ontology". *Revista Signos. Estudios de Lingüística*, 47 (84), 113-139.

Ruiz de Mendoza Ibáñez, F.J. y Mairal Usón, R. (2008). "Levels of description and constraining factors in meaning construction: an introduction to the Lexical Constructional Model". *Folia Lingüística*, 42, (3), 355–400.

Sánchez García de Paz, I. (2009). *La criminalidad organizada. Aspectos penales, procesales, administrativos y policiales*. Madrid: Dykinson.

Ureña Gómez-Moreno, P., Alameda Hernández, A. & Felices Lago, A. (2011). "Towards a specialised corpus of organized crime and terrorism". In Carrió, M.L. et al. (eds.) *La investigación y la enseñanza aplicadas a las lenguas de especialidad y a la tecnología*. Valencia: Universitat Politècnica de Valencia, 301-306.

Ureña Gómez-Moreno, P. (2014). "El conocimiento cultural en dominios especializados: Un acercamiento desde la base de conocimiento FunGramKB". In G. Budin & V. Lušicky (eds.), *Languages for Special Purposes in a Multilingual, Transcultural World, Proceedings of the 19th European Symposium on Languages for Special Purposes*, 8-10 July 2013, Vienna, Austria. Vienna: University of Vienna, 509-514.

Valente, A. (2005). "Types and Roles of Legal Ontologies". In V.R.. Benjamins et al. (eds.). *Law and the Semantic Web*. Berlin: Springer, 65-76.