Automatic Detection of Opposition Relations in Legal Texts Using Sentiment Analysis Techniques: A Case Study

Obdulia Pichardo-Lagunas[1], Bella Martinez-Seis[1], Miguel Hidalgo-Reyes[2], Sabino Miranda[3]

[1] Instituto Politécnico Nacional, Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas, Avenida Instituto Politécnico Nacional 2580, Barrio la Laguna Ticomán, Gustavo A. Madero, 07340, Ciudad de México, México;opichardola@ipn.mx, bcmartinez@ipn.mx

[2] Instituto de Ecología A.C. (INECOL), Carretera antigua a Coatepec 351, Col. El Haya, Xalapa, Veracruz. CP 91073, México; miguel.hidalgo@inecol.mx

[3]INFOTEC Centro de Investigación e Innovación en Tecnologías de la Información y Comunicación, Circuito Tecnopolo Sur 112, Fracc. Tecnopolo Pocitos, 20313, Aguascalientes, México; sabino.miranda@infotec.mx

Abstract: The documentation that describes the regulations within a Society, is oriented towards specific areas. This fact does not prevent maintaining concordance in the temporality and transversality of the documents. This work defines the concept of "opposition relations" in legal texts. We identify entities and evaluate the polarity of each paragraph with sentiment analysis techniques. If an entity appears in different paragraphs (articles of law) with opposite polarities, we evaluate the entity's contexts. We look for antonyms between the words that give polarity to the opposite paragraphs. If there is an antonymic relation in words associated with the entity, we have an opposition relation. The described methodology analyzes the relationship of entities in Mexican Environmental Laws, and the study is oriented towards coherence in the legislation for sustainable development. This process was implemented by computational processing, which required the transformation of current Mexican laws, unifying its structure. Eight environmental laws were analyzed, 1920 entities were identified that appear more than once; 44 of them were identified with opposite polarities, due to their context, a detailed analysis of two cases with potential opposite relationships is exemplified.

Keywords: Natural Language Processing; Legal Text; Sentiment Analysis; Opposition Relation

1   Introduction

The generation of laws and regulations at different countries, states, and municipalities allows implementing public policies to benefit the societies they relate. Laws and regulations are written with specific objectives aimed at specific scopes and concerning the different actors involved. For instance, in Mexico, the *Ley de Fondos de Inversión* (Investment Funds Law) is mainly oriented toward the economic sector. In contrast, the *Ley de Hidrocarburos* (Hydrocarbons Law) draws together actors from economics, politics, international, environmental, and transport sectors, among others.

In the last decade, creating public policy legislation that considers the concept of sustainable development has been one of the objectives set by different governments, together with agreements reached with organizations such as the OECD (Organization for Economic Cooperation and Development).

By legislating in this sense, it is intended to put the environment as a priority issue and its protection as a necessary condition for the construction of the development of any country [1]. By directing the actions of federal public administration agencies and entities towards the concept of sustainable development, it is intended to permeate this term throughout the legal system [2] achieving an instrument of truly transversal nature that establishes principles of observable and punishable performance in all activities [3].

However, in many cases, the writing on the legislative documents do not maintain the expected coherence between proposed public policies and the sustainable development goals promised by the State. From this environmental legislation perspective arises the motivation of this work, we make a computational analysis of the laws in order to measure the coherence level between the design of Mexican public policies embodied in environmental legislation and the concept of sustainable development.

The present work proposes a methodology that, using techniques of Natural Language Processing (NLP), identifies the possible inconsistencies in a corpus of laws and regulations from the Mexican normative, based on the different entities identified. This methodology identifies entities and analyzes the paragraphs that contains them. A sentiment analysis is applied to identify the polarity (positive or negative) in the named entities, highlighting opposition relations, also considering the antonymic relations.

We define "opposition relation" as the association that exists between two text sections, from the same or different documents, that are evaluated with opposite polarities while sharing actors and that, in the context of identified entities, contain vocabulary with antonymic relations.

The hypothesis of this work is as follows: The opposition relations identify potentially incoherent elements, because based on the polarity assigned to the context of a previously identified actor (entity) and shared across different sections

of text, it is expected that the classification stays the same. That is, if two texts share an entity, both should be classified with the same polarity and the vocabulary that contextualizes them should not keep antonymic relations.

By developing this type of methodologies, computational tools that help in the validation of the coherence relation between the writing of public policies and sustainable development goals could be generated.

It is difficult to carry out an automatic computational analysis in texts pertaining the legal field, as it is important, initially, to identify characteristics that allow to build computational structures that enable a syntactic analysis. However, given the complexity of the task and the nature of the texts, there are no tools at the moment that allow this task to be carried out properly. Some of the problems that can be identified are:

1. To identify patterns at legal documents that allow to discover knowledge.

2. To identify the semantics of some of the words used in the legal scene, since it sometimes differs from their semantics in a general scope, i.e., how they are commonly used.

3. To build knowledge bases that allow to represent the different concepts within the legal scene.

2   NLP: Overview of Legal Texts

Various applications have been made in the legal framework by using Natural Language Processing (NLP) and Artificial Intelligence (AI) [4]. Standing out among them are the prediction of trials [5], question answering [6], searches [7] and text summarization in various legal frameworks [8]. However, some analyzes represent interesting opportunities to explore, for example, measuring the level of coherence between laws.

According to Zhong et al. [9], two types of law processing methods stand out: on the one hand, methods based on symbols that use an interpreted knowledge or the learned rules [10]; on the other hand, embedded methods that learn on a large scale. The former has a lower efficiency while the latter lacks interpretability, which represented an obstacle in this work due to the validation condition.

The processing of laws with artificial intelligence techniques has been used in multiple works analyzing different countries and languages. For example, in [5] is analyzed the results of various labor lawsuits in different Brazilian regions using different machine learning techniques, identifying the reasons why a trial turns in favor or against. Meanwhile, in [11], the authors classify judgments based on German legislation by using machine learning, classifiers like Support Vector Machines (SVM) or Naive-Bayes, identifying classes as definitions, obligations,

revocations (repeal), among others. They build a pattern-based classifier by the comparison of candidate phrases with standard phrases, with around 90% of accuracy; they point out that when writing German laws, they use legal conventions that allow the interpretation of similar information, using structures such as by X is understood Y where X is a term and Y a definition. In addition, Son et al. [12] present an analysis for laws in the Vietnamese language using deep neural networks.

2.1   Structure of Legal Documents

Among the problems identified by several authors, there is a lack of standardization when writing a law.

Sometimes, the expected pattern is not fitted, or the expressiveness of the words found is inadequate since they could belong to different classes. Palmirani's proposal [13] presents a structure in XML to describe legislative documents under two standards. The general structure includes elements of the normative act, particular elements such as references, and meta-information like references to other laws, notes, or locations.

Ashley and Rissland [14] focus on constructing a computational representation of laws based on legislations from the US and Great Britain. They identify rules and actions from different legislations and the word sets or corpus over time.

The task of recognizing the logical compositions in legal documents ranges from assumptions to sanctions, elements that help to detect a legal document structure. In Mexico, there is no standardization in the structure of laws, either by temporality or by changes in the executive branch, which complicates generating a computational structure.

2.2   Identification of the Coherence Level

From the perspective of identifying the level of coherence in legal documents, in [15], the authors present a project developed by Stanford University in which, by using a software tool, they structure the knowledge that is found within a set of legal texts of the United States. Based on different representations of knowledge extracted using the mentioned tool and labeled manually, among other tasks, they focus on identifying inconsistencies, highlighting the importance of the generation of interpretive manuals due to the difficulty of analyzing legal texts given the interpretation to which they are subject.

In [16] the authors identify the orders that are most closely related by a domain search and organizational structures, giving a measure of similarity. This similarity can be used to measure the coherence between documents.

3   Text Analysis Techniques

The use of computer tools in analyzing texts from any domain is already known; however, the use of the internet has allowed access to large collections of legal texts from various topics and countries. It has increased the scientific community's interest in processing, analyzing, and recovering text information from legal documents. One of the promising research areas is based on acquiring legal knowledge and the synthesis techniques of documents and hypertext structures.

3.1   Named Entity Recognition

Named Entity Recognition consists in locating and classifying lexical units by studying categories such as places, people, organizations, time, and quantities expressions. It is recognized as one of the main tasks of natural language processing (NLP). For example, in the text shown in 1 the identified entities are marked with bold text. It is notorious that entities comprise one or more words that represent the element on which the action takes place.

Table 1

Named entities example in Spanish and its translation in English

ARTÍCULO 6o.- Las atribuciones que esta Ley otorga a la Federación, serán ejercidas por el Poder Ejecutivo Federal a través de la Secretaría y, en su caso, podrán colaborar con ésta las Secretarías de Defensa Nacional y de Marina cuando por la naturaleza y gravedad del problema así lo determine, salvo las que directamente corresponden al Presidente de la República por disposición expresa de la Ley.
ARTICLE 6. The attributions that this Law grants to the Federation, will be exercised by the Federal Executive Power through the Secretariat and, if necessary the Secretariats of National Defense and Navy may collaborate with it when determined by the nature and seriousness of the problem, except those that directly correspond to the President of the Republic by express provision of the Law.

Currently, there are different software tools that automatically allow the identification of entities [17]. Specifically, universities such as Stanford or the Massachusetts Institute of Technology (MIT) have developed tools to identify different types of entities; one of them is the spaCy library [18].

SpaCy is a library developed by MIT that contains pre-trained statistical models and word vectors, allowing the tokenization process in 49 languages. It has convolutional neural networks models for labeling, analysis, and recognition of named entities, integrating deep learning modules to solve these tasks.

The library allows to identify different types of entities such as PERSON (people, even fictitious ones), NORP (nationalities or religious or political groups), FAC Buildings, airports, roads, bridges, etc.), ORG (companies, agencies, institutions,

etc.), GPE (countries, cities, states) and LOC (non GPE Locations like mountain ranges, bodies of water). In the example of the 6th Article of Table II, the entities classified as ORGanization or PERSON are shown.

Table 2

Labeled entities example in Spanish and its translation in English

ARTÍCULO 6o.- Las atribuciones que esta Ley otorga a la Federación(ORG), serán ejercidas por el Poder Ejecutivo Federal(ORG) a través de la Secretaría(ORG) y, en su caso, podrán colaborar con ésta las Secretarías de Defensa Nacional y de Marina(ORG) cuando por la naturaleza y gravedad del problema así lo determine, salvo las que directamente corresponden al Presidente de la República(PERSON) por disposición expresa de la Ley.
ARTICLE 6. The attributions that this Law grants to the Federation(*ORG*), will be exercised by the Federal Executive Power(*ORG*) through the Secretariat(*ORG*) and, if necessary the Secretariats of National Defense and Navy(*ORG*) may collaborate with it when determined by the nature and seriousness of the problem, except those that directly correspond to the President of the Republic(*PERSON*) by express provision of the Law.

3.2   Sentiment Analysis (SA)

The sentiment analysis (SA) is an NLP technique also known as Opinion Mining. The SA aims to perform automatic text classification based on positive or negative connotations of the language used.

Among the tools used to solve this task include the lexicons or dictionaries [19]. The work of Zafra et al. [20] contains a list of words classified in two categories (positive or negative) that will help as a guide for the evaluation of polarity in a text. A software tool used in sentiment analysis is SentiStrength [21] [22].

SentiStrength was initially designed to assess the sentiment in published texts on MySpace [21]. However, it has been used with good results in short text analysis. This tool proposes the use of dictionaries incorporating sets of grammar rules, which provides a double orientation in the identification of the sentiment in the text. SentiStrength provides two values for the parsed text: the first one measures the intensity of the positive sentiment, and the second one the intensity of the negative sentiment. These values range from 1 to 5, also providing the result of the analysis in three formats: Binary (positive/negative), Trinary (positive/negative/neutral) and Simple scale (from -4 to +4) [23].

Negation Treatment in SentiStrength

The negation markers such as "no" (not), "ni" (neither), among others are used employing a list of words by the SentiStrength algorithm, which assigns a polarity to the paragraphs. The negating word list is used to invert following emotion words, skipping any intervening booster words; for more details, see [21] [22].

4 Methodology: Detection of Semantic Opposites in Mexican Environmental Legislation

The methodology described in this document aims to identify the opposition relations existing in a selected set of texts, based on the evaluation of polarity and antonymic relations between them. The stages that make up the solution of this task are shown in the Figure 1.

Figure 1
Described methodology for the identification of opposition

4.1 Text Selection and Entity Identification

The first stage of the methodology for the detection of opposition relations consists of the four tasks shown in Figure 2, going from the selection of laws to the identification of entities, going through the transformation of documents to semi-structured texts and formats.
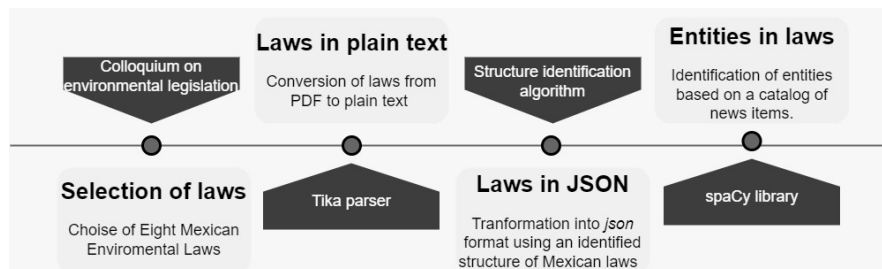

Figure 2
Text selection and entity identification (first step of methodology)

The selection of the set of legislative documents, specifically laws, consists of a set of 8 laws of environmental matter identified as the most significant at the First Interdisciplinary Colloquium for the Analysis of Environmental Legislation [24].
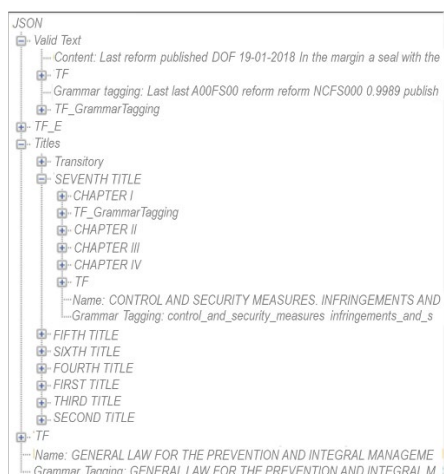
Figure 3

Example of the hierarchical structure identified in a legal document

At this stage, the text was pre-processed and stored in JSON format. The proposed format for document storage allows to maintain the natural hierarchical structure used in the drafting of regulations. Conversion from the source format of the used files was done using the Tika Python library [25].

The Tika library [25] reads the file in pdf format and extracts the metadata of the document and the corresponding content values to later write out all the information in a plain text file.

Next, an algorithm that allows the identification of a generalized hierarchical structure in Mexican laws was designed; this algorithm transforms the plain text to a semi-structured text in JSON format, customized for legislation. Figure 3 shows an example of the hierarchical structure identified in a legal document.

The entity recognition in the texts was carried out using the spaCy library [26]. This library has a catalog of entities based on identified entities in news notes. It is important to mention that the catalog used by spaCy is of common knowledge, meaning that it does not count with vocabulary specialized in legislative matter.

4.2   Sentiment Analysis of the Entity's Contexts

This stage aims to identify the polarity of the words around the detected entities as shown in Figure 4. Specifically, the polarity of the paragraphs in which the entities occurred is obtained to detect pairs of paragraphs that could have any opposition relation concerning the entities. Once the entities described in Section 4.1 have been identified, these are replaced in the generated text file assigning a sequential number to each of the entities. Each entity is replaced by the word entity plus the corresponding sequential number; for example, the entity *"Aguas Nacionales"*

(National Waters) is replaced in the documents by the term "entity45". This substitution is made to prevent in this stage that the tool used recognizes any word that belongs to the entity as a term that could indicate polarity (positive or negative).
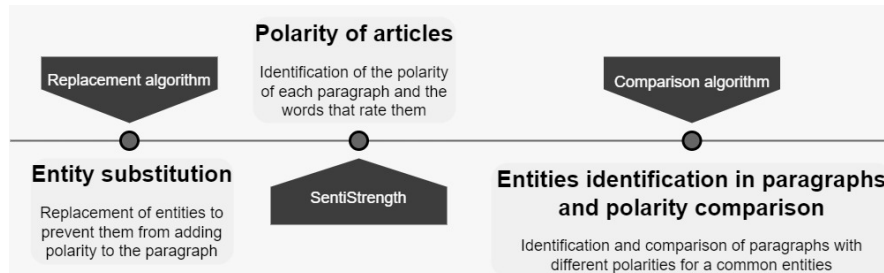

Figure 4

Process for the detection of the sentiments associated with an entity, according to the paragraphs in which the latter is located (second methodology stage)

Once the entities are substituted, each one of the paragraphs is sent to the SentiStrength tool [21], to identify each paragraph's polarity and word's polarity that indicate any charge, both positive and negative, considering the entities identified in Section 4.1 and the polarities of the paragraphs in which they are found.

Next, for a particular entity, a Python script is used to compare the polarities identifying those paragraphs that show a distinct polarity between them. That is, positive and negative polarities, an example is shown in Section 5.3.

4.3   Antonym Identification

At this stage, given the entities that show different polarities in different paragraphs in the same law or two different laws, each word from a particular paragraph that indicates any sentiment, positive or negative, is compared with the words from another paragraph that indicates the opposite sentiment, negative or positive, to identify antonyms. The identification is performed using a dictionary of antonyms based on the Wordreference platform [27], an example of this case is shown in Table 8, in which the word *"explotación"* (exploitation) is found in the first column and the second column its antonyms identified in another paragraph.

Finally, by identifying different words that indicate a polarity of different paragraphs by a relation of antonyms, these paragraphs are identified as having an opposition relationship for a particular entity.

4.4   Coherence Validation

In this stage, a list of the entities that appear in at least two separate paragraphs is obtained. The list contains the number of times that each entity appears for each polarity, that is, the number of positive, negative, and neutral paragraphs in which

each entity appears. For the entities that appear in paragraphs with different polarities, an analysis of antonyms is carried out within the words that give the meaning charge to polarity analysis. It allows to identify potential opposite relations that would generate low levels of coherence between laws.

The coherence validation is carried out by the hand of public policy experts, lawyers, and legislators. In this way, they would be able to delve into the context of entities and determine the level of coherence, using this tool as support. In other contexts, there are labeled corpora related to the terms polarity, but in the legislative and environmental context there are no developed semantic resources for legal domain.

5   Analysis of Results in Mexican Environmental Laws

This section shows some of the results obtained by applying the presented methodology on the test corpus, described in Section 4, recalling that an opposition relation is detected when antonyms between two different paragraphs of two different laws are detected.

5.1   Laws Documents

This work is developed in the environmental context, and in this particular case, eight environmental laws listed in Table 3 were analyzed.

Table 3

Document's Names employed in the experiment in Spanish and its translation in English

Mexican Environmental Laws
Ley General del Equilibrio Ecológico y la Protección al Ambiente (General Law of Ecological Balance and Environmental Protection)
Ley de Aguas Nacionales, reformada en el 2020 (National Waters Law, reformed in 2020)
Ley de Bioseguridad de Organismos Genéticamente Modificados (Biosafety Law for Genetically Modified Organisms)
Ley General de Desarrollo Forestal Sustentable (General Law of Sustainable Forest Development)
Ley General de Pesca y Acuacultura Sustentable (General Law on Sustainable Fisheries and Aquaculture)
Ley General de Vida Silvestre (General Wildlife Law)
Ley General para la Prevención y Gestión Integral de los Residuos (General Law for the Prevention and Comprehensive Management of Waste)

5.2   Identification of Potential Semantic Opposites in Environmental Laws

The methodology described in Section 4.1 obtains the list of all entities with more than one appearance to compare them. This list identifies how many times an entity appears in paragraphs with positive, negative, and neutral polarity; and which laws they appeared. An example of this list is shown in Table 4.

Table 4

Example of some detected entities in Spanish and its translation in English. It is shown the number of occurrences and the polarity associated with the paragraphs in which they appear

Entity	Repetitions	Constant Polarity	Positive Appearances	Negative Appearances	Neutral Appearances
Nueva ley (new law)	2	Neutral	0	0	4
Diario Oficial de la Federación (Official Journal of the Federation*)*	8	-	14	1	1
Ley de Comercio Exterior(Foreign Trade Law)	3	-	2	4	0
Estados Unidos Mexicanos (Mexico)	8	Neutral	0	0	16
Ley (Law)	7	Positive	14	0	0
Vicente Fox Quesada (name of former president)	3	Neutral	0	0	6
Presidente de los Estados Unidos Mexicanos (President of Mexico)	5	-	2	0	8
Aguas Nacionales (National Waters)	2	-	3	1	0
El servicio de aseguramiento (The assurance service)	1	Negative	0	2	0
Foro Consultivo Científico y Tecnológico (Scientific and Technological Consultative Forum)	1	-	1	1	0
Comisión Federal de Electricidad (Federal electricity commission)	2	-	3	1	0

Ley General de Pesca y Acuacultura Sustentable (General Law on Sustainable Fisheries and Aquaculture)	4	-	3	1	4
Secretarías (Secretaries)	3	-	4	0	2

In the study of the eight environmental laws, 1920 entities were identified with at least two appearances. The distribution of these entities is shown in Fig. 5. 1240 entities appear in paragraphs with constant polarity. It means that all its occurrences are positive, negative, or neutral; for example, the entity "*Estados Unidos Mexicanos*" (Mexico, full name of Mexico Country) appears 16 times in the eight laws, and all their appearances are in paragraphs with neutral polarity.

Out of the 1920 entities, only 680 do not have a constant polarity. In this sense, two levels of coherence are identified. The first refers to entities in paragraphs with neutral occurrences combined with positive or negative occurrences; in such a manner that the level of coherence could still be present; this is the case of 636 entities out of the 680. However, for the second level identified, there are entities that have occurrences in both positive and negative contexts, which could present potential opposites relations, such is the case of the 44 entities shown in red in Fig. 5. This low percentage is consistent with the fact that all the laws analyzed in this work belong to the same topic.
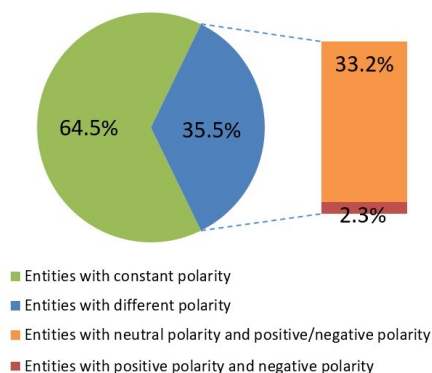

Figure 5

Polarity distribution of the 1920 entities with at least two occurrences, in the eight environmental laws

The entities that are in paragraphs with positive and negative polarity are studied in the antonym identification stage (Section 4.3) and in the coherence validation stage (Section 4.4). From those last stages, most of the entities, like "Economía" (Economy), are in paragraphs with different polarities, but there are no antonyms detected in the words that give polarity to the paragraphs; then, an opposition relation is not detected. Nevertheless, few entities present potential opposition relations, the entities that stand out are exemplified below within three of the laws.

5.3 Opposition Markings Identified for the Entity "*Aguas Nacionales*" (*National Waters)*

After analyzing the law: *"Ley General de Equilibrio Ecológico y Protección al Ambiente", LGEEPA* (General Law of Ecological Balance and Environmental Protection) and the law: "*Ley de Aguas Nacionales*", *LAN* (National Waters Law), it was identified that they share the use of some entities such as "Aguas Nacionales" (National Waters) whose paragraphs of occurrence are shown in Table 5 and 6 respectively. Based on the entity "Aguas Nacionales" (National Waters) and by analyzing the vocabulary found in the context of this entity, the existence of an opposition relation between both texts was detected.

Table 5

Paragraph of LGEEPA with positive charge for "aguas nacionales" (national waters) in Spanish and its translation in English

Los lineamientos para la realización de las acciones de preservación, restauración y aprovechamiento sustentable de los recursos naturales dentro de las áreas naturales protegidas, para su administración y vigilancia, así como para la elaboración de las reglas administrativas a que se sujetarán las actividades dentro del área respectiva, conforme a lo dispuesto en ésta y otras leyes aplicables; Las medidas que el entity41 podrá imponer para la preservación y protección de las áreas naturales protegidas, serán únicamente las que se establecen, según las materias respectivas, en la presente Ley, las entity116, de <u>aguas nacionales</u>, de entity149, entity58, y las demás que resulten aplicables
The guidelines for carrying out actions for the preservation, restoration and sustainable use of natural resources within protected natural areas, for their administration and surveillance, as well as for the elaboration of administrative rules to which activities within the area will be subject respective, in accordance with the provisions of this and other applicable laws; The measures that entity41 may impose for the preservation and protection of protected natural areas will only be those established, according to the respective matters, in this Law, entity116, <u>national waters</u>, entity149, entity58, and other that are applicable.

Table 6

Paragraph from the LAN law with negative charge for the entity in Spanish and its translation in English

"entity319": Acto emitido por el entity91 por causas de utilidad pública o interés público, mediante la declaratoria correspondiente, para extinguir: a. entity320 o asignaciones para la explotación, uso o aprovechamiento de entity321, de sus bienes públicos inherentes, ob. entity322, equipar, operar, conservar, mantener, rehabilitar y ampliar infraestructura hidráulica federal y la prestación de los servicios respectivos; XLVI. ', 'entity493 por causas de utilidad pública o interés público, declaratorias de rescate, en materia de concesiones para la explotación, uso o aprovechamiento de aguas nacionales, de sus bienes públicos inherentes, en los términos establecidos en la entity265; V. entity493 por causas de utilidad pública o

interés público, declaratorias de rescate de concesiones otorgadas por "la entity45", para construir, equipar, operar, conservar, mantener, rehabilitar y ampliar infraestructura hidráulica federal y la prestación de los servicios respectivos, mediante pago de la indemnización que pudiere corresponder; VI
"entity319 ": Act issued by the entity91 for reasons of public utility or public interest, through the corresponding declaration, to extinguish: a. entity320 or assignments for the exploitation, use or exploitation of entity321, of its inherent public goods, or b. entity322, equip, operate, conserve, maintain, rehabilitate and expand federal hydraulic infrastructure and the provision of the respective services; XLVI. ',' entity493 for reasons of public utility or public interest, declarations of rescue in the matter of concessions for the exploitation, use or use of <u>national waters</u> , of their inherent public goods, in the terms established in the entity265; V. entity493 for reasons of public utility or public interest, declarations of rescue of concessions granted by "the entity45", to build, equip, operate , conserve, maintain, rehabilitate and expand federal hydraulic infrastructure and the provision of services respective, by payment of the compensation that may correspond; VI ..

As it can be seen in Table 7, there are two sets of words for each of the paragraphs in which the entity "Aguas Nacionales" (National Waters) occurs, one of them made out from positive words and another from negative words. These sets of words were constructed from the results of the SentiStrength tool [21].

Table 7

Words that indicate the polarity at the inconsistency in Spanish and its translation in English

Law	Words	
	Positives	Negatives
LGEEPA	*realización*[+3] (realization)	
	preservación[+2] (preservation)	
	restauración[+2] (restoration)	
	protegidas[+2] (protected)	
	dispuesto[+2] (arranged)	
	medidas[+2] (measures)	
	protección[+2] (protection)	
	únicamente[+2] (only)	
LAN	*establecidos*[+3] (established)	
		explotación[-3] (explotation)
		servicios[-2] (services)

By using the WordReference dictionary of antonyms [19], it is noted that the word "explotación" (explotation) identified in the context of the entity "Aguas Nacionales" (National Waters), in the paragraph shown in Table 6 for the *LAN* law, has a relationship of antonymy with different words found in the paragraph of the *LGEEPA* law (Table 5) for the same entity; these words are shown in Table 8.

Table 8

Identified antonyms from the words that indicate the polarity in Spanish and its translation in English

Word (LAN)	Antonyms (LGEEPA)
explotación (exploitation)	*Realización*(realization)
	Preservación (preservation)
	protegidas (protected)
	protección (protection)

According to the methodology developed in the present work, it is concluded that the entity "Aguas Nacionales" (National Waters) presents a low level of coherence, given that when performing the sentiment analysis in the paragraphs in which this entity is mentioned in the *LGEEPA* and *LAN* laws it was identified that some of the words, that give polarity to the paragraph and that are related to this entity, are antonyms.

5.4 Opposite Polarity without Opposition Relation: Entity "Áreas" (Areas)

It was identified that the entity "Areas" appears in the LGEEPA and in the law "Ley General de Vida Silvestre" *LGVS* (General Wildlife Law); their paragraphs of appearance are shown in Tables 9 and 10 respectively. Based on the entity "Áreas" (areas), opposite polarities in the texts were obtained; nevertheless, when analyzing the vocabulary found in the context of this entity, there was no opposition relation obtained.

As it can be seen in Table 11, there are two sets of words for each one of the paragraphs in which the entity "Áreas" (Areas) occurs, one of them of positive words and the other of negative words. These sets of words were constructed from the results of the polarity analysis. Depending on the occurrences of these words a polarity was assigned to the paragraph, resulting in "Áreas" (Areas) associated with positive polarity for the *LGEEPA* law and negative polarity for *LGVS* law.

Table 9

Paragraph from the *LGEEPA* with positive charge for the entity "Áreas" (Areas) in Spanish and its translation in English

Se consideran áreas naturales protegidas: I. entity173; II. Se deroga. III. entity146 nacionales; IV. entity138 naturales; V. Se deroga. VI. áreas de protección de recursos naturales; VII. áreas de protección de flora y fauna; VIII. entity179; IX. entity147, así como las demás categorías que establezcan las legislaciones locales; X. entity224, así como las demás categorías que establezcan las legislaciones locales, y XI. áreas destinadas voluntariamente a la conservación. Para efectos de lo establecido en el presente Capítulo, son de competencia de la entity56 las áreas naturales protegidas comprendidas en las fracciones I a VIII y XI anteriormente señaladas. entity125 y del entity36, en los términos que señale la legislación local en la materia, podrán establecer parques, reservas

*estatales y demás categorías de manejo que establezca la legislación local
en la materia, ya sea que reúnan alguna de las características señaladas
en las fracciones I a VIII y XI del presente artículo o que tengan
características propias de acuerdo a las particularidades de cada entidad
federativa. . .*

The following are considered protected natural areas: I. entity173; II. It is
repealed. III. entity146 nationals; IV. natural entity138; V. It is repealed.
VI. Natural resource protection areas; VII. Flora and fauna protection areas;
VIII. entity179; IX. entity147, as well as the other categories established by
local laws; X. entity224, as well as the other categories established by local
legislation, and XI. Areas voluntarily designated for conservation. For the
purposes of what is established in this Chapter, the natural protected areas
included in sections I to VIII and XI mentioned above are the responsibility
of the entity56. entity125 and entity36, under the terms indicated by local
legislation on the matter, may establish parks, state reserves and other
management categories established by local legislation on the matter,
whether they meet any of the characteristics indicated in sections I to VIII
and XI of this article or that have their own characteristics according to the
particularities of each federative entity.

Table 10

Paragraph from the *LGVS* with negative charge for the entity "Áreas" (Areas) in Spanish and its
translation in English

*... entity15 podrá establecer, mediante acuerdo entity48, hábitats críticos
para la conservación de la vida silvestre, cuando se trate de: a) <u>áreas</u>
específicas dentro de la superficie en la cual se distribuya una especie o
población en riesgo al momento de ser listada, en las cuales se desarrollen
procesos biológicos esenciales para su conservación. b) <u>áreas</u> específicas
que debido a los procesos de deterioro han disminuido drásticamente su
superficie, pero que aún albergan una significativa concentración de
biodiversidad. c) <u>áreas</u> específicas en las que existe un ecosistema en riesgo
de desaparecer, si siguen actuando los factores que lo han llevado a reducir
su superficie histórica. d) <u>áreas</u> específicas en las que se desarrollen
procesos biológicos esenciales, y existan especies sensibles a riesgos
específicos, como cierto tipo de contaminación, ya sea física, química o
acústica, o riesgo de colisiones con vehículos terrestres o acuáticos, que
puedan llevar a afectar las poblaciones.*

... entity15 may establish, through an entity48 agreement, critical habitats
for the conservation of wildlife, in the case of: a) specific <u>areas</u> within the
surface in which a species or population at risk is distributed at the time of
listing, in the which essential biological processes are developed for their
conservation. b) specific <u>areas</u> that due to deterioration processes have
drastically decreased their surface area, but still harbor a significant
concentration of biodiversity. c) specific <u>areas</u> in which there is an
ecosystem at risk of disappearing, if the factors that have led it to reduce its
historical surface continue to act. d) specific <u>areas</u> in which essential
biological processes are developed, and there are species sensitive to
specific risks, such as certain types of pollution, whether physical, chemical
or acoustic, or risk of collisions with land or water vehicles, which may lead
to damage populations..

Table 11

Words that indicate polarity in the context of the entity "Áreas" (Areas) in Spanish and its translation in English

Law	Words	
	Positives	Negatives
LGEEPA	*protegidas*[+2](protected)	
	protección[+2] (protection)	
	voluntariamente[+3] (voluntarily)	
	competencia[+2] (competition)	
	acuerdo[+2] (agreement)	
LGVS	*protección*[+2] (protection)	*críticos*[-2] (critical)
	acuerdo[+2] (agreement)	*procesos*[-2] (processes)
	significativa[+2] (significant)	*riesgo*[-2] (risks)
	sensibles[3] (sensitive)	*deterioro*[-2] (deterioration)
		riesgos[-2] (risks)
		contaminación[-2] (pollution)

According to the results, in the *LGEEPA* law, the entity is in a context of protected areas, while in the *LGVS*, it is found in the context of critical habitats, so they get opposite polarities. However, at the antonym detection stage of the vocabulary obtained from the context of both paragraphs, there is no antonym relation in any pair of words that were found in the paragraphs from the *LGEEPA* and *LGVS* laws for this entity. Based on our methodology developed in the present work, it is concluded that the entity "Aguas Nacionales" (National Waters) has opposite polarities, but it is coherent by not presenting opposition relations.

5.5   Analysis of Results

The case study reflected in this document was obtained from the analysis of a set of eight laws, all of the ecological matter. It can be seen that even in documents focused on a particular subject, there can be identified opposition relations. By the analysis embodied in the Section 5.2, in Table 4, it is observed that the identified entities in different paragraphs present the same polarity in most cases. Only 2.3% of the entities are in paragraphs with opposite polarities. Those entities are considered for a deeper study because of the relation of those entities to their context. The example of the *"Aguas nacionales"* (National Waters) entity in Section 5.3 shows how the methodology, based on antonym relations in the context vocabulary, achieves the identification of a potential opposition relation. The *"Áreas"* (Areas) example (Section 5.4) highlights the importance of all the steps from this methodology (identify entities, obtain polarity and identify antonym relations) to mark an opposition relation since the polarity tested in the sentiment analysis is not enough for the marking. The consistency validation is carried out by experts in public policies, lawyers and legislators.

Conclusions

In this research, the concept of "opposition relations" in legal texts was defined, starting from antonymic relations in the vocabulary associated with entities whose paragraphs present opposite polarities when implementing sentiment analysis techniques.

The developed methodology consists of 4 stages, ranging from the computational transformation of texts to the coherence validation; going through two core stages: the identification of polarity in identified entities and the analysis of antonymic relations in the context. This methodology was implemented with computational algorithms and applied to the environmental scope of Mexican laws, but it is applicable in other areas.

The sentiment analysis tasks and the identification of entities were made with generic order tools. The polarity is evaluated using a dictionary, and for the marking of entities, a list of entities identified in news was used. Therefore, it is important that with the collaboration of different professionals, such as political scientists, lawyers and/or politicians, language resources could be generated, that allow characterization of the different concepts and relations that are involved in the diverse texts in the legal domain.

Acknowledgements

References

[1]   H. Koff and C. Maganda: The EU and the human right to water and sanitation: Normative coherence as the key to transformative development. The European Journal of Development Research, 2016, Vol. 28, No. 1, pp. 91-110

[2]   M. Rovalo Otero: La transversalidad del Derecho Ambiental: Un paradigma necesario en el Siglo XXI [The transversality of Environmental Law: A necessary paradigm in the 21st Century]. Política y Gestión Ambiental [Environmental Policy and Management], 2016, pp. 49-52

[3]   Antena Radio: Diversidad Ambiental, Sección Medio Ambiente, ¿Qué puedo hacer yo? [Environmental Diversity, Environment Section, What can I do?] with Francisco Calderón Córdova / IMER - Horizonte 107.9 FM, - 1220 AM y - Radio México Internacional, 29/01/2018 [Online] [Last access: 02/02/2021] Available: http://diversidadambiental.org/medios/nota610.html

[4]   E. L. Rissland, K. D. Ashley and R. Loui: AI and Law: A fruitful synergy. Artificial Intelligence, 2003, Vol. 150, No. 1-2, pp. 1-15

[5]   R. Barros, A. Peres, F. Lorenzi and L. K. Wives: Case law analysis with machine learning in brazilian court. International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, 2018, Vols. 1 de 2 Springer, Cham, pp. 857- 868

[6]   R. Giri, Y. Porwal, P. Chadha and R. Kaushal: Approaches for information retrieval in legal documents. 2017 Tenth International Conference on Contemporary Computing (IC3), 2017, Vol. IEEE, pp. 1-6

[7]   Y. Chen, Y. Liu and W. Ho: A text mining approach to assist the general public in the retrieval of legal documents. Journal of the American Society for Information Science and Technology, 2013, Vol. 64, No. 2, pp. 280-290

[8]   Y. Ma, P. Zhang and J. Ma: An ontology driven knowledge block summarization approach for Chinese judgment document classification. IEEE Access, 2018, Vol. 6, pp. 71327-71338

[9]   H. Zhong, C. Xiao, C. Tu, T. Zhang, Z. Liu and M. Sun: How Does NLP Benefit Legal System: A Summary of Legal Artificial, 2020, arXiv preprint arXiv:2004.12158

[10] M. Dragoni, S. Villata, W. Rizzi and G. Governatori: Combining NLP Approaches for Rule Extraction from Legal Documents. De 1st Workshop on MIning and RIasoning with Legal texts (MIREL 2016), 2016

[11] E. d. Maat, K. Krabben and R. Winkels: Machine Learning versus Knowledge Based Classification of Legal Texts. JURIX, 2010, pp. 87-96

[12] N. T. Son, N. T. P. Duyen, H. B. Quoc and N. L. Minh: Recognizing logical parts in Vietnamese legal texts using conditional random fields. The 2015 IEEE RIVF International Conference on Computing Communication Technologies-Research, Innovation, and Vision for Future (RIVF), IEEE, 2015

[13] M. Palmirani and R. Brighi: Metadata for the legal domain. 14th International Workshop on Database and Expert Systems Applications, 2003, Proceedings, 2003, Vol. IEEE, pp. 553-558

[14] K. D. Ashley and E. L. Rissland: Law, learning and representation. Artificial Intelligence, 2003, Vol. 150, Nos. 1-2, pp. 17-58

[15] K. H. Law, G. Lau, S. Kerrigan and J. A. Ekstrom: REGNET: Regulatory information management, compliance and analysis. Government Information Quarterly, 2014, Vol. 31, pp. S37-S48

[16] G. Lau, K. H. Law and G. Wiederhold: A Relatedness Analysis of Government Regulations using Domain, Knowledge and Structural Organization. Information Retrieval, 2006, Vol. 9

[17]　X. Schmitt, S. Kubler, J. Robert, M. Papadakis and Y. LeTraon: A replicable comparison study of NER software: StanfordNLP, NLTK, OpenNLP, SpaCy, Gate. 2019 Sixth International Conference on Social Networks Analysis, Management and Security, 2019 (SNAMS) IEEE

[18]　ExplosionAI: Spacy, industrial-strength natural language processing. [Online], [Last access: 01/04/2021] Available: https://spacy.io/

[19]　M. Taboada, J. Brooke, M. Tofiloski, K. Voll and M. Stede: Lexicon-based methods for sentiment analysis. Computational linguistics, 2011, Vol. 37, No. 2, pp. 267-307

[20]　S. M. Jiménez Zafra, E. Martínez Cámara, M. T. Martín Valdivia and M. D. Molina González: Tratamiento de la Negación en el Análisis de Opiniones en Espanol [Treatment of Negation in Spanish Opinion Analysis], Procesamiento del Lenguaje Natural [Natural Language Processing] 2015, No. 54, pp. 37-44

[21]　M. Thelwall, K. Buckley, G. Paltoglou, D. Cai and A. Kappas: Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology, 2010, Vol. 61, No. 12, pp. 2544-2558

[22]　M. Thelwall: Heart and soul: Sentiment strength detection in the social web with sentistrength. Cyberemotions: Collective emotions in cyberspace, 2014

[23]　T. Baviera: Técnicas para el análisis del sentimiento en Twitter: Aprendizaje Automático Supervisado y SentiStrength [Techniques for Sentiment Analysis on Twitter: Supervised Machine Learning and SentiStrength], Dígitos, 2017, Vol. 1, No. 3, pp. 33-50

[24]　UPIITA-IPN: Primer Coloquio Interdisciplinario para el Análisis de Legislación Ambiental [First Interdisciplinary Colloquium for the Analysis of Environmental Legislation], UPIITA-IPN, 13-14 08 2019 [Online] [Last access: 15/03/2021. Available: https://www.upiita.ipn.mx/novedades/coloquio-legislacionambiental

[25]　C. Mattmann, Zitting J.: Tika in Action. Greenwich, CT., 2011

[26]　J. D. Choi, J. Tetreault and A. Stent: It depends: Dependency parser comparison using a web-based evaluation tool. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 2015, (Volume 1: Long Papers)

[27]　B. Landesman: WordReference.com, Reference Reviews, 2002

[28]　H. Dubossarsky, E. Grossman and D. Weinshall: Outta Control: Laws of Semantic Change and Inherent Biases in Word Representation Models. Proceedings of the 2017 conference on empirical methods in natural language processing, 2017, pp. 1136-1145