



---

---

**INSTITUTO POLITÉCNICO NACIONAL**

CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN

**Minería de textos para búsqueda de  
fundamentos jurídicos en sentencias**

T E S I S

QUE PARA OBTENER EL GRADO DE  
MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN

PRESENTA:

LIC. FRANCISCO GABRIEL SANDOVAL FLORES

DIRECTORES DE TESIS: DR. ALEXANDER GELBUKH  
DR. FRANCISCO HIRAM CALVO CASTRO

ENERO 2016





# INSTITUTO POLITÉCNICO NACIONAL SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

## ACTA DE REVISIÓN DE TESIS

En la Ciudad de México, D.F. siendo las 10:00 horas del día 04 del mes de diciembre de 2015 se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación del:

### **Centro de Investigación en Computación**

para examinar la tesis titulada:

**“Minería de textos para búsqueda de fundamentos jurídicos en sentencias”**

Presentada por el alumno:

**SANDOVAL**  
Apellido paterno

**FLORES**  
Apellido materno

**FRANCISCO GABRIEL**  
Nombre(s)

Con registro: 


B	1	3	0	1	2	8
---	---	---	---	---	---	---

aspirante de: **MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN**

Después de intercambiar opiniones los miembros de la Comisión manifestaron **APROBAR LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

### LA COMISIÓN REVISORA Directores de Tesis

  
Dr. Francisco Hiram Calvo Castro

  
Dr. Alexander Gelbukh

  
Dr. Sergio Suárez Guerra

  
Dr. Grigori Sidorov

  
Dra. Olga Kolesnikova

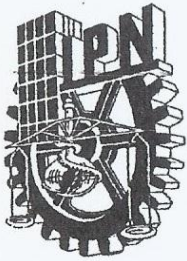
  
Dr. Ildar Batyrshin

PRESIDENTE DEL COLEGIO DE PROFESORES

  
Dr. Luis Alfonso Villa Vargas



ESTADOS UNIDOS MEXICANOS  
INSTITUTO POLITÉCNICO NACIONAL  
SECRETARÍA DE INVESTIGACIÓN  
Y POSGRADO  
CENTRO DE INVESTIGACIÓN  
EN COMPUTACIÓN



**INSTITUTO POLITÉCNICO NACIONAL**  
**SECRETARÍA DE INVESTIGACIÓN Y POSGRADO**

**CARTA CESIÓN DE DERECHOS**

En la Ciudad de México D. F. el día 15 del mes de **diciembre** del año 2015, el (la) que suscribe **C. Francisco Gabriel Sandoval Flores** alumno (a) del Programa de **Maestría en Ciencias de la Computación** con número de registro **B130128**, adscrito al **Centro de Investigación en Computación del Instituto Politécnico Nacional**, manifiesta que es autor (a) intelectual del presente trabajo de Tesis bajo la dirección de **Dr. Alexander Gelbukh** y **Dr. Francisco Hiram Calvo Castro** y cede los derechos del trabajo intitulado **“Minería de textos para búsqueda de fundamentos jurídicos en sentencias”**, al **Instituto Politécnico Nacional** para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección **fgab\_1004@hotmail.com**. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

---

Francisco Gabriel Sandoval Flores

## Resumen

El Derecho tiene una importancia central en la sociedad; su estudio se relaciona con prácticamente todas las áreas del conocimiento; desde el ámbito de las Ciencias de la Computación se han realizado múltiples aportaciones a esta disciplina principalmente en las áreas conocidas como *informática jurídica documental, informática jurídica de gestión e informática jurídica decisional*.

Nuestro trabajo se ubica en esta última y abordamos el tema considerando que la aplicación del derecho es una tarea interpretativa que consiste en subsumir hechos jurídicamente relevantes en el contenido abstracto de una norma jurídica. Desde nuestra perspectiva esta labor se ha vuelto sumamente compleja debido a múltiples factores, entre ellos: la abrumadora cantidad de leyes existentes, las modificaciones que de forma permanente se realizan al sistema legal, los constantes cambios sociales, etc.

La forma en que los juristas han sobrellevado estas dificultades ha sido la especialización en el estudio de las distintas ramas jurídicas; sin embargo, esto ha limitado el análisis integral del sistema jurídico y genera dificultades al abordar problemas que involucran diversas áreas legales.

Por otra parte, un elemento que caracteriza al ámbito legal es que la mayor parte de la información que en él se genera se encuentra en forma de texto, constituyendo las sentencias judiciales un componente sustancial de esta información.

Considerando este factor, utilizamos un corpus de decisiones judiciales y a través de la minería de textos desarrollamos diferentes métodos que permiten descubrir las relaciones existentes entre los casos que han sido resueltos y las normas que en ellos fueron aplicadas. La minería de textos es un área de investigación dirigida a la extracción de patrones y conocimiento en grandes colecciones de textos.

Los métodos que proponemos hacen posible la búsqueda de fundamentos jurídicos aplicables a situaciones concretas independientemente del área legal en que puedan ser clasificados, esto facilita el estudio y análisis cabal de los asuntos jurídicos en cualquier rama del derecho.

## Abstract

Law has a great significance in society. Its study is related with almost every area of knowledge. Specifically, the area of Computer Science has made multiple contributions to this discipline, most of them through the automatic systems that facilitate the management of legal information.

In this work we consider that law enforcement is an interpretive labor, which consists of subsuming a situation within the abstract content of legal norms, and this has become a complex task, because legal systems are composed of an overwhelming number of rules, and these rules are always changing.

To address these problems, lawyers have separated the study of law in diverse fields; however, this has constrained the complete analysis of the legal system and also creates problems when a case is related to different legal areas at the same time.

On the other hand, most of the information that is generated on the legal field is in text form, and judicial decisions are a significant portion of this information. Taking this fact, we used text mining to develop different methods in order to discover relations between legal cases and applicable rules using a corpus of legal judgments. Text mining is dedicated mainly to discover interesting patterns and knowledge that are present in text collections.

The methods that we propose are useful to retrieve the rules that apply for a given specific case, independently of the legal areas to which they belong. We believe that these methods can be extremely helpful in the analysis and resolution of legal issues.

*para Juan y Carmen,  
por su amor y dedicación*

## **Agradecimientos**

*A mi esposa y a mi hija,  
por felicidad que me han dado*

*A mis hermanos,  
por su motivación y por acompañarme  
en esta etapa de cambios y experiencias*

*A mis asesores Dr. Alexander Gelbukh y Dr. Hiram Calvo,  
por su confianza y apoyo*

*A mis amigos,  
por los momentos compartidos y  
los consejos que me brindaron*

*Al Instituto Politécnico Nacional,  
y al Centro de Investigación en Computación,  
por ser un espacio abierto a la investigación que  
permite el desarrollo de la ciencia y las ideas*

*Al CONACyT,  
por las facilidades otorgadas para  
el desarrollo de este trabajo*

*“El juicio, el verdadero, el justo juicio, el juicio que no falla  
está solamente en las manos de Dios.”*

***Francesco Carnelutti***



## Contenido

1. Introducción .....	1
1.1. Objetivos .....	3
1.2. Aportaciones .....	4
2. Estado del arte .....	5
2.1. Sistemas basados en conocimiento .....	6
2.1.1 Expertus .....	7
2.1.2. ARPO-2.....	8
2.1.3. KBS .....	8
2.1.4. Sistema para detectar conflictos en ordenamientos jurídicos distintos.....	9
2.1.5. LOIS .....	9
2.2. Sistemas basados en probabilidad .....	10
2.2.1. Sistema para determinar la relevancia de las pruebas .....	10
2.2.2. Sistema para predecir la resolución de un problema legal.....	11
2.3. Sistemas que utilizan métodos de clasificación .....	12
2.3.1. Sistemas de clasificación de normas jurídicas.....	12
2.3.2. Sistemas para la detección de argumentos en sentencias .....	14
2.4. Sistemas que utilizan métodos de recuperación de información .....	15
2.4.1. Sistemas para la recuperación de normas jurídicas relacionadas .....	16
2.4.2. Sistema para la extracción de argumentos en normas jurídicas .....	17
2.4.3. Sistema para generación de resúmenes de sentencias judiciales .....	18
3. Marco teórico.....	19
3.1. Consideraciones teórico-jurídicas .....	19
3.1.1. La función del derecho .....	19
3.1.2. Visión constructivista del derecho .....	21
3.1.2.1. La jurisprudencia analítica de H. L. A. Hart.....	22
3.1.2.2. Elementos de la teoría del derecho de Ronald Dworkin.....	23

3.1.3. Visión realista del derecho .....	25
3.1.3.1. Indeterminación de las normas jurídicas .....	26
3.1.3.2. Crítica a la discrecionalidad .....	28
3.2. Elementos de lingüística jurídica.....	30
3.2.1. Relaciones formales entre normas jurídicas .....	31
3.2.2. Relaciones semánticas entre normas jurídicas .....	33
3.2.3. Relaciones pragmáticas entre normas jurídicas.....	35
3.3 Resoluciones judiciales.....	39
3.3.1. Requisitos formales de las sentencias judiciales.....	40
3.3.2. Estructura de las sentencias judiciales .....	40
3.3.3. Requisitos de fondo de las sentencias .....	41
3.4 Procesamiento del Lenguaje Natural .....	41
3.4.1. Extracción de Información.....	42
3.4.2. Recuperación de información .....	45
3.4.2.1. Modelo de espacio vectorial .....	46
3.5 Minería de texto.....	49
3.5.1. Clasificación de documentos.....	50
3.5.2. Funciones de distancia y semejanza .....	52
4. Estructura del trabajo propuesto.....	55
4.1 Recopilación de sentencias judiciales .....	55
4.2. Procesamiento.....	57
4.2.1 Reconocimiento de jurisprudencias y leyes .....	58
4.2.2. Reconocimiento de conceptos legales .....	63
4.2.3. Reconocimiento de artículos.....	64
4.3. Métodos desarrollados .....	68
4.3.1. Método 1 .....	69
4.3.2. Método 2 .....	70
4.3.3. Método 3 .....	71
4.3.4. Método 4 .....	73

5. Resultados y análisis.....	75
5.1. Descripción del conjunto de datos para evaluación .....	75
5.2. Resultados experimentales aplicando el método 1 .....	77
5.3. Resultados experimentales aplicando el método 2 .....	87
5.5. Comparación de resultados (métodos 1, 2 y 3) .....	100
5.6. Resultados experimentales aplicando el método 4 .....	101
6. Conclusiones .....	105
Referencias.....	107

## Índice de tablas y figuras

Tabla 1. Clases y evaluación de los resultados obtenidos por el sistema NIR.....	13
Tabla 2. Ejemplo de información estructurada. ....	42
Tabla 3. Promedio de leyes y artículos en relación con los elementos que integran un considerando. ....	76
Tabla 4. <i>precision</i> y <i>recall</i> obtenidos en la recuperación de leyes con n=2 (método 1).....	78
Tabla 5. <i>precision</i> y <i>recall</i> obtenidos en la recuperación de leyes con n=5 (método 1).....	79
Tabla 6. <i>precision</i> y <i>recall</i> obtenidos en la recuperación de leyes con n=10 (método 1).....	80
Tabla 7. <i>precision</i> y <i>recall</i> obtenidos en la recuperación de artículos con n=2 (método 1). ....	83
Tabla 8. <i>precision</i> y <i>recall</i> obtenidos en la recuperación de artículos con n=5 (método 1). ....	84
Tabla 9. <i>precision</i> y <i>recall</i> obtenidos en la recuperación de artículos con n=10 (método 1). ....	85
Tabla 10. <i>precision</i> y <i>recall</i> obtenidos en la recuperación de leyes con n=2 (método 2).....	88
Tabla 11. <i>precision</i> y <i>recall</i> obtenidos en la recuperación de leyes con n=5 (método 2).....	89
Tabla 12. <i>precision</i> y <i>recall</i> obtenidos en la recuperación de leyes con n=10 (método 2).....	90
Tabla 13. <i>precision</i> y <i>recall</i> obtenidos en la recuperación de artículos con n=2 (método 2). ....	93
Tabla 14. <i>precision</i> y <i>recall</i> obtenidos en la recuperación de artículos con n=5 (método 2). ....	94
Tabla 15. <i>precision</i> y <i>recall</i> obtenidos en la recuperación de artículos con n=10 (método 2). ....	95
Tabla 16. <i>precision</i> y <i>recall</i> obtenidos con 10 grupos (método 3).....	99
Tabla 17. Comparación de la medida <i>precision</i> (método 4).....	102
Tabla 18. Comparación de la medida <i>recall</i> (método 4).....	103

Figura 1. Tipología de las reglas de acuerdo a jurisprudencia analítica de Hart. ....	23
Figura 2. Esquema de un sistema de extracción de información. ....	43
Figura 3. Esquema de un sistema de recuperación de información.....	46
Figura 4. Formas de representar un vector. ....	47
Figura 5. Representación de características en un vector.....	48
Figura 6. Distancia Euclidiana entre dos puntos.....	52
Figura 7. Distancia Manhattan entre dos puntos. ....	53
Figura 8. Representación de semejanza coseno entre dos vectores.....	54
Figura 9. Trie para la localización de secuencias de palabras.....	59
Figura 10. Artículos 74 fracción III y 73 fracción X párrafo segundo de la ley de amparo. ....	66
Figura 11. Artículo 42 bis de la ley federal de los trabajadores al servicio del estado.....	66
Figura 12. Agrupamiento de términos legales. ....	72
Figura 13. Crecimiento de la función sigmoide con valores de -1 a 1. ....	74
Figura 14. Formato en el que son ordenados los resultados.....	77
Figura 15. Comparación de la medida <i>precision</i> en la recuperación de leyes (método 1).....	81
Figura 16. Comparación de la medida <i>recall</i> en la recuperación de leyes (método 1). ....	82
Figura 17. Comparación de la medida <i>precision</i> en la recuperación de artículos (método 1).....	86
Figura 18. Comparación de la medida <i>recall</i> en la recuperación de artículos (método 1).....	87
Figura 19. Comparación de la medida <i>precision</i> en la recuperación de leyes (método 2).....	91
Figura 20. Comparación de la medida <i>recall</i> en la recuperación de leyes (método 2). ....	92
Figura 21. Comparación de la medida <i>precision</i> en la recuperación de artículos (método 2).....	96
Figura 22. Comparación de la medida <i>recall</i> en la recuperación de artículos (método 2).....	96
Figura 23. Comparación de las medidas <i>precision</i> y <i>recall</i> (método 3). ....	98
Figura 24. Comparación de la medida <i>precision</i> (métodos 1, 2, 3). ....	100
Figura 25. Comparación de la medida <i>recall</i> (métodos 1, 2, 3). ....	101
Figura 26. <i>precision</i> y <i>recall</i> obtenidos variando cardinalidad de L (método 4).....	103
Figura 27. <i>precision</i> y <i>recall</i> obtenidos variando el número de artículos recuperados (método 4).....	104

## Glosario

**Agrupamiento por densidad.** Conjunto de métodos para agrupar objetos basado en la noción de densidad, esto es, dado un dato  $d$ , un radio de vecindad  $r$ , y un umbral  $u$ , se considera que  $d$  se encuentra en una *zona de alta densidad* si el número de datos con distancia menor o igual a  $r$  respecto a  $d$  es mayor o igual a  $u$ ; Se agregan elementos a un grupo en tanto estos se encuentren en un *zona de alta densidad*.

**Análisis pragmático.** Método que busca estudiar la intención de una palabra u oración.

**Análisis semántico.** Procedimiento para estudiar el significado de una oración o palabra.

**Análisis sintáctico.** Método para identificar la función de las palabras en una oración.

**Chunker.** Técnica para realizar un *análisis sintáctico superficial*, esto es, identificar la función sólo de ciertas palabras en una oración.

**Clasificación no supervisada.** También conocido como *agrupamiento*, es un conjunto de métodos para identificar y agrupar objetos similares utilizando las características propias de los objetos.

**Clasificación supervisada.** Clasificación en sentido estricto, es un conjunto de métodos en los que a partir de una muestra de datos a los que les ha sido asignada una clase (muestra de supervisión) se identifican las características y los valores que distinguen a los elementos de cada clase respecto de otras, con el fin de que sean utilizadas para determinar de manera automática la clase a la que pertenecen objetos que se encuentran fuera de la muestra de supervisión.

**Congruencia.** Característica que debe tener toda resolución judicial, y que consiste en una correspondencia entre lo que se solicita al tribunal y lo que este resuelve.

**Considerandos.** Parte las resoluciones judiciales en la que se realiza el análisis jurídico mediante el cual se llega a una determinación.

**Corpus.** Conjunto de textos.

**Distancia de Hamming.** Número de bits que tienen que modificarse para transformar una cadena binaria en otra.

**Exhaustividad.** Característica que debe tener toda resolución judicial, consistente en analizar todas y cada una de las cuestiones planteadas al órgano jurisdiccional.

**Expresiones regulares.** Secuencia de palabras o caracteres que forma un patrón de búsqueda.

**Fundamentación.** Propiedad que debe poseer todo acto de autoridad y que consiste en indicar el precepto legal sobre el que se basa su actuación.

**Grafo no conexo.** Un grafo es no conexo cuando existe al menos un par de vértices que no se encuentra conectado por un camino.

**Grafo no dirigido.** Grafo en el que las aristas representan una relación simétrica entre cada par de vértices y por lo cual el darles una dirección no aporta información.

**Gramática Libre de Contexto.** Es una gramática formal  $G = (T, N, P, S)$  en donde T es un conjunto finito de símbolos terminales, N es un conjunto finito símbolos no terminales, P es un conjunto finito de producciones de la forma  $T \rightarrow (T \cup N)^*$  y S es el símbolo inicial que pertenece a N.

**Informática jurídica.** Conjunto de métodos computacionales que solucionan en todo o en parte un problema jurídico.

**Lematización.** Proceso para estandarizar palabras que consiste en asignar a cada palabra un lema, el cual, es seleccionado mediante convenio como representante de todas las formas flexionadas de una palabra.

**Ley adjetiva.** Ley que estable las formas y los procedimientos para hacer efectivos los derechos públicos subjetivos.

**Ley sustantiva.** Ley que establece derechos, obligaciones y/o sanciones a los sujetos de derecho.

**Máquina de soporte vectorial.** Método de clasificación supervisada que se enfoca en encontrar un hiperplano que divida un espacio vectorial a fin de distinguir entre diferentes clases.

**Motivación.** Obligación de todo órgano jurisdiccional de expresar los razonamientos que llevan a la autoridad a aplicar los fundamentos jurídicos al caso concreto.

**n-gramas.** Subsecuencia de n palabras de una oración.

**POS tagging.** *Part of speech tagging o, etiquetado de categorías gramaticales, es una técnica que consiste en determinar la categoría gramatical de cada palabra presente en una oración.*

**Precision.** Medida de evaluación que en el área de recuperación de información se calcula dividiendo el número de elementos relevantes recuperados entre el número de elementos recuperados.

**Rankeo.** Método para ordenar un conjunto de elementos, usualmente asignando las primeras posiciones a los objetos mejor evaluados.

**Recall.** Medida de evaluación que en el área de recuperación de información se calcula dividiendo el número de elementos relevantes recuperados entre el número de elementos relevantes.

**Reconocimiento de Entidades Nombradas.** Método para detectar de forma automática una palabra o secuencia de palabras que se identifican con una persona física o moral, un lugar, una cantidad, una fecha, etc.

**Stop-words.** Conjunto de palabras que por encontrarse presente en prácticamente todos los textos, se considera que no aportan información para diferenciar entre un texto y otro.

**String.** Tipo de dato mediante el cual se pueden representar secuencias de caracteres.

**Tokenización.** Método para dividir secuencias de palabras.



# Introducción

En toda sociedad el derecho juega un papel trascendente para las personas, pues cada acto que llevamos a cabo puede tener una connotación jurídica: realizar una compra, caminar por la calle, incluso en algunos casos la forma de vestir<sup>1</sup> tiene una cierta regulación; de esta manera podemos decir que las personas convivimos permanentemente con normas jurídicas. La importancia del estudio del derecho radica precisamente en el impacto que este tiene sobre la sociedad.

A su vez, el estudio del derecho es una de las cuestiones más complicadas de atender; en primer lugar, trabaja con conceptos no definidos. Hoy en día no tenemos una definición formal de justicia, como tampoco tenemos una definición de derecho. El análisis del alcance de una norma jurídica no es menos difícil de realizar; hasta junio de 2013 el derecho a la información se encontraba estipulado en el artículo sexto constitucional de la siguiente manera: ***el derecho a la información será garantizado por el estado***, hoy en día tiene una regulación más detallada; sin embargo, la cantidad de páginas escritas y de análisis realizados sobre estas 10 palabras es realmente considerable. Por lo que se refiere al estudio del ordenamiento jurídico, podemos afirmar que se ha convertido en un aspecto imposible de atender para cualquier jurista o institución, pues la cantidad de normas jurídicas y de modificaciones al sistema legal es realmente abrumadora.

En cuanto a su aplicación, el derecho pretende corregir los comportamientos no deseados en la sociedad a través de la solución a problemas jurídicos. En este punto cabe realizar la clásica pregunta ¿existe una forma correcta de resolver los asuntos jurídicos? Robert Alexy plantea cinco idealizaciones para alcanzar este objetivo: tiempo ilimitado, información ilimitada, claridad lingüística y conceptual ilimitada, capacidad y disposición ilimitada para el cambio de roles y carencia de prejuicios ilimitada [1]; en realidad todo juez tiene una capacidad limitada para entender las normas jurídicas y sus alcances en un caso concreto y esta limitación se ve acrecentada por los continuos cambios legales y sociales. En cuanto al tiempo, el acceso a una justicia pronta es una exigencia que todo juzgador debe atender, por lo cual el tiempo que se tiene para resolver un asunto es limitado. Cabe mencionar que parte de las demandas sociales que culminaron en las reformas constitucionales de junio de 2008<sup>2</sup>, encuentran su origen en estos problemas.

---

1 Reglamento de la toga universitaria de la UNAM.

2 Ésta ha sido una de las reformas más importantes a nuestro sistema jurídico, aunque también una de las más

Finalmente, los ciudadanos en general no tenemos forma de predecir las decisiones que se toman en materia jurídica ni sus consecuencias, específicamente no podemos predecir los cambios que se presentarán en el ordenamiento jurídico o la determinación que tomará un juez en un asunto.

Frente a estos problemas, desde la perspectiva computacional existen dos posturas: la postura suave supone que la complejidad de estos problemas puede reducirse a través de métodos computacionales ayudando a los operadores jurídicos (legisladores, jueces, abogados y doctrinarios) a entender de una mejor manera los problemas con los que tratan y tomar mejores decisiones; la postura fuerte considera que los problemas jurídicos pueden ser tratados directamente a partir de sistemas computacionales. En nuestro caso asumimos una postura débil; inclusive se destaca que los problemas que mencionamos se encuentran más allá de la capacidad actual de las computadoras.

---

costosas; el presupuesto autorizado a la SETEC para la implementación de la reforma en materia de justicia penal de 2008 a 2012 fue de más de 1, 600 millones de pesos; además, mediante el fondo de apoyo para infraestructura y seguridad se aprobó un presupuesto de 12, 000 millones de pesos para infraestructura y equipamiento en seguridad pública y sistema de justicia en las entidades federativas [información consultable en el *libro blanco de la Secretaría Técnica del Consejo de Coordinación para la implementación del Sistema de Justicia Penal de la Gestión 2009-2012*]. Para este año, fueron destinados más de 1, 000 millones al programa presupuestario “Otorgamiento de subsidios para la implementación de la reforma al sistema de justicia penal” [información consultable en el *acuerdo por el que se establecen las Políticas para la obtención y aplicación de los recursos destinados a la implementación de la reforma del Sistema de Justicia Penal a favor de las entidades federativas para el ejercicio fiscal 2015*]. Se estima que el presupuesto necesario para aplicar en todos los estados el nuevo sistema de justicia es de aproximadamente 30,000 millones de pesos para 2015 y 2016 [<http://www.excelsior.com.mx/nacional/2014/09/03/979759>].

## 1.1. Objetivos

Nuestro trabajo se centra en el problema de la aplicación del derecho, específicamente, en la búsqueda de fundamentos jurídicos para resolver una situación concreta. Consideramos que realizar esta tarea de manera correcta consume la mayor parte del tiempo que destinan los profesionistas del derecho a resolver cuestiones legales; se han realizado estudios sobre la manera en la que los abogados dan solución a los problemas jurídicos [36, 54] y se ha encontrado que en general para llegar a una solución los abogados tienden a recurrir en primer lugar a colegas, a compañeros y a jueces; en lugar de recurrir directamente a la legislación o a los precedentes y en nuestro caso a la jurisprudencia. Cabe puntualizar que a partir de la reforma constitucional en materia de derechos humanos de junio de 2011, en el futuro se podría plantear la operatividad judicial del precedente en México.

Lo anterior se debe a que mucho del razonamiento legal es analógico, Cass R. Sunstein ha señalado por ejemplo, que en el *common law* la tarea se centra en buscar buenas analogías [51] y es considerablemente más fácil encontrar un caso análogo preguntando a un conocido que directamente rastrearlo. De hecho en nuestro país esta tarea se ve mucho más compleja debido a la falta de organización y sistematización de las sentencias emitidas por los órganos jurisdiccionales; sin embargo, haciendo referencia a los operadores jurídicos en lo individual, es una práctica común resolver asuntos similares de forma similar, aunque en general es difícil encontrar criterios unívocos respecto a un problema.

Otra forma de localizar los fundamentos jurídicos para un caso en concreto es a través de los motores de búsqueda en internet, o de los sistemas actualmente desarrollados para este fin (como por ejemplo compila u otras herramientas editadas por la SCJN) sin embargo, en estos casos se presenta una dificultad adicional: el buscar en términos apropiados, así como tomar en cuenta las características de las fuentes que buscamos.

Considerando lo anterior, sería deseable la existencia de un sistema de búsqueda basado en casos, a partir del cual fuera posible localizar de forma inmediata los fundamentos jurídicos que se han aplicado en situaciones similares; dicho sistema contribuiría a satisfacer la necesidad de una justicia clara y expedita. El objetivo central de este trabajo se enfoca precisamente en desarrollar una metodología para hacer esto posible.

Para lograr este fin, en primer lugar es necesario representar las situaciones jurídicas de forma tal que puedan ser manejadas computacionalmente; además, requerimos una forma de detectar cuándo dos (o más) situaciones jurídicas son similares y finalmente ubicar qué fundamentos jurídicos son aplicables a estas situaciones.

Para alcanzar estos objetivos nosotros proponemos un método en el cual no consideramos el contenido de las normas jurídicas, detectamos las **relaciones pragmáticas** que entre ellas existen así como los casos en que estas son aplicables utilizando un **corpus de decisiones judiciales**.

## 1.2. Aportaciones

Las principales aportaciones de este trabajo son las siguientes:

- Corpus recopilado con más de 1.8 millones de sentencias judiciales.
- Método para representar situaciones jurídicas.
- Corpus etiquetado con más de 7.9 millones de *considerandos*.
- Tres métodos para localizar fundamentos jurídicos a partir de situaciones o hechos concretos.
- Un método para la detección e identificación de relaciones entre fundamentos jurídicos.

## Estado del arte

La aplicación de las Ciencias de la Computación al Derecho es un tema que se ha estudiado desde dos ángulos: uno científico y otro social. Desde el punto de vista científico, la investigación se ha centrado en descubrir y aplicar metodologías que nos permitan una mejor comprensión del sistema jurídico; es decir, se pretende responder ¿qué métodos nos ayudan a entender el sistema jurídico? y ¿cómo podemos aplicarlos?. Desde la óptica social el análisis se ha enfocado en la regulación de las nuevas tecnologías, el área encargada de contestar esta pregunta es el *derecho informático* o *derecho de la informática*.

Desde la perspectiva jurídica, el estudio de la aplicación de las Ciencias de la Computación al Derecho ha sido planteado a través de la *informática jurídica*. Julio Téllez define esta área como una técnica interdisciplinaria que tiene por objeto el estudio e investigación de los conocimientos de la informática general, aplicables a la recuperación de información jurídica [42]. Se habla además de tipos de informática jurídica, a saber, Informática jurídica documental, informática jurídica de gestión e informática jurídica decisional [21].

La informática jurídica documental se refiere a la aplicación de técnicas informáticas a documentos jurídicos en aspectos como el análisis, archivo y recuperación de información contenida en la legislación, jurisprudencia, doctrina o cualquier otro documento con contenido jurídico relevante [42]. Por su parte, la informática jurídica de gestión tiene como propósito apoyar el desenvolvimiento de las tareas de carácter rutinario en el ámbito legal [21], como por ejemplo el seguimiento de trámites. Finalmente la informática jurídica decisional abarca una gran variedad de esfuerzos y proyectos que intentan obtener de las aplicaciones de la informática al derecho, algoritmos que resuelvan problemas jurídicos o al menos auxilien a hacerlo [42].

Con respecto a esta clasificación de la informática jurídica consideramos que la misma no es adecuada para clasificar los sistemas que han sido propuestos para el tratamiento de problemas en el área jurídica, esto se debe a las siguientes consideraciones:

En primer lugar porque todos los sistemas propuestos tienen como fin el auxiliar en la resolución de problemas jurídicos, y por tal motivo pueden ser calificados como parte de la informática jurídica decisional, aún y cuando su principal función sea por ejemplo, la recuperación de información, que de acuerdo con esta clasificación tendría que ubicarse dentro de la informática jurídica documental.

Por otra parte, si bien es posible hablar de informática jurídica de gestión, y en ella podríamos encontrar desarrollos como el Sistema Integral de Seguimiento de Expedientes (SISE) utilizado en el Poder Judicial de la Federación, dicha área actualmente no tiene una alta productividad como línea de investigación en las Ciencias de la Computación; esto se debe principalmente a que su actividad puede considerarse como parte de la informática jurídica documental, así como de la informática jurídica decisional, por tanto, nos parece más adecuado agrupar los sistemas computacionales aplicados al derecho de acuerdo con los métodos que han utilizado para alcanzar sus objetivos. De esta forma expondremos las siguientes clases de sistemas computacionales aplicados al derecho: Sistemas basados en conocimiento, Sistemas basados en probabilidad, Sistemas que utilizan métodos de clasificación y Sistemas que utilizan métodos de recuperación de información.

Antes de proceder a la exposición de estos sistemas, deseamos aclarar que para alcanzar los objetivos propuestos en el presente trabajo hacemos uso de los siguientes métodos computacionales: clasificación no supervisada (agrupamiento) reconocimiento de entidades nombradas (NER) modelo de espacio vectorial (VSM) extracción de información y recuperación de información, por tal motivo la exposición de dichos temas se ubica en el capítulo 3.

## **2.1. Sistemas basados en conocimiento**

Hemos decidido comenzar explicando los sistemas basados en conocimiento, dado que fueron algunos de los primeros métodos considerados inteligentes aplicados al derecho. Estos sistemas encuentran su fundamento en la siguiente reflexión “para resolver un problema en la práctica, es necesario saber de antemano la correspondiente respuesta” [44]. Básicamente estos tipos de sistemas utilizan el conocimiento específico de un dominio, para facilitar el desarrollo de *etapas largas del razonamiento*, lo cual hace posible resolver casos recurrentes en dominios de conocimiento restringido [44].

En cuanto a su desarrollo, una vez que se ha decidido que cierto tipo de problemas tiene una solución factible a partir de estos sistemas, usualmente se recurre a personas expertas en tratar dichos problemas, a fin de generar una base de conocimiento, este conocimiento es sistematizado y formalizado a través de un conjunto de *reglas* que son usadas por el sistema para resolver los problemas originalmente planteados. Debido a ello este tipo de sistemas son conocidos también como *sistemas expertos*. Ahora procedemos a describir algunos de ellos:

### 2.1.1 Expertius

Sistema desarrollado en el Instituto de Investigaciones Jurídicas (IIJ) de la UNAM en el año 2008 [9]. Está enfocado a resolver problemas en el área del derecho familiar, específicamente en el estado de Tabasco, y su objetivo es *aconsejar* respecto a la determinación de si una persona se encuentra obligada o no al otorgamiento de una *pensión alimenticia*, así como respecto al monto de la misma. Es un sistema dirigido a jueces y personal de juzgados y consta de tres módulos:

- **Tutorial.** En este módulo se orienta al usuario sobre cómo proceder en las diferentes etapas de un juicio. Su función principal es proporcionar al usuario métodos apropiados para solucionar problemas que son característicos de las tareas cognitivas en el marco de un proceso judicial. Básicamente en este módulo el usuario selecciona la etapa del procedimiento en la que se encuentra y la tarea específica en la que necesita ayuda. El sistema es capaz de indicar si la etapa tiene un efecto directo o indirecto sobre la resolución legal que se determine, por ejemplo, la admisión de una prueba tendría un efecto directo en la resolución, y no así el acuerdo que indica que el periodo probatorio ha concluido, al considerarse una cuestión de mero trámite (su efecto sería indirecto).
- **Módulo de inferencia.** En este módulo se evalúa la evidencia presentada por las partes, basándose en pesos heurísticos que el usuario debe proporcionar para las diferentes evidencias; está encargado de evaluar la evidencia y determinar que proposiciones han sido derrotadas y cuales prevalecen a través del proceso, considerando los pesos heurísticos que fueron asignados.
- **Módulo financiero.** Provee asistencia al usuario sobre cómo determinar el monto de la pensión. Toma en consideración las necesidades reales del demandante y la capacidad económica del demandado. Las reglas de este módulo fueron creadas a partir de un estudio sociológico.

Para el desarrollo de este módulo se analizaron 433 casos a partir de los cuales se construyeron “plantillas” que representan escenarios que contienen formas estandarizadas de resolver problemas, y se identificaron casos atípicos que se definen como escenarios que no encajan en las formas estandarizadas.

## 2.1.2. ARPO-2

Este sistema fue presentado en el año de 1996 por Jesús Cardeñosa y Pilar Lasala [11], y considera el hecho de que cuando un abogado busca una solución a un problema jurídico, su enfoque aglutina dos vertientes: la solución del caso, y los fundamentos de su decisión.

Se encuentra dirigido a jueces y abogados, a los primeros los auxilia en la justificación de sus decisiones y a los segundos a predecir sobre cuál puede ser la determinación que el juez adopte. Su dominio se encuentra restringido al incumplimiento de contratos de construcción. Para obtener la base de conocimientos del sistema, se acudió a clases de la Facultad de Derecho de la Universidad de Zaragoza en España y se obtuvo la ayuda de un magistrado de la Corte de Justicia de *Aragón*, también en ese país.

Este sistema determina en primer lugar si existe un contrato de construcción; una vez establecido esto, muestra todas las posibles acciones a seguir, o dada una acción deseada, indica si es posible llevar a cabo la misma. Finalmente realiza ciertas preguntas al usuario, y a partir de la información que se le proporciona llega a una decisión y proporciona la documentación que fundamenta la misma. El sistema puede llegar a alguna de las siguientes decisiones: No existe contrato de construcción, existe incumplimiento de contrato (no se realizó la construcción) o se realizó la construcción pero se presentaron daños o vicios ocultos.

## 2.1.3. KBS

Sistema expuesto por Julia Barragán [11], y diseñado para determinar el ámbito de aplicabilidad de una ley penal venezolana a un caso determinado. Este sistema toma en consideración los siguientes aspectos:

- Los juristas no aplican la ley de manera única.
- Una decisión judicial se toma bajo condiciones de incertidumbre; éstas se pueden presentar al agregar nuevas consideraciones al hecho, o al manipular de manera distinta los elementos normativos (refinando o *reformulando* el contenido de una norma).

Es definido como un sistema basado en reglas de derecho, que captura la experiencia de los magistrados, así como las pautas y estructuras de inferencia empleadas por los juristas [3].



Para concluir con la exposición referente a los *sistemas expertos*, únicamente mencionamos los sistemas que se han desarrollado en algunos países con el fin de mostrar la importancia que ha llegado a tener este enfoque computacional: CREDOC en Bélgica, DATA LEX en Dinamarca, JURISDATA en Francia, JURIS en Alemania, LEXIX y WESTLAW en EU.

#### **2.1.4. Sistema para detectar conflictos en ordenamientos jurídicos distintos**

A diferencia de los sistemas mencionados, el sistema presentado por *Tringting Li, et al.* [31], proporciona un *marco de trabajo* en el que, dados dos escenarios específicos, se detecta de forma automática si los escenarios pueden dar lugar a diferentes soluciones jurídicas; esto es, determina si un comportamiento puede ser interpretado de forma distinta en diferentes jurisdicciones, y si el mismo puede llevarnos a resultados divergentes (que pueden ser favorables o perjudiciales). El sistema encuentra su justificación en el hecho de que dos órganos jurídicos de distinta jurisdicción pueden ser competentes para resolver un mismo asunto. En estos casos, resulta de gran importancia identificar las disparidades que existen entre las diferentes leyes aplicables, así como las posibles consecuencias de las mismas.

Para ilustrar el funcionamiento del proyecto, dos ordenamientos jurídicos de países distintos (Italia e Inglaterra) fueron *alineados* y, aunque por el momento sólo se detectan discrepancias respecto a prohibiciones y permisiones, el sistema fue capaz de identificar las disparidades en leyes de carácter civil para un caso específico. Ahora bien, el sistema requiere para su funcionamiento que los textos legales se encuentren alineados semánticamente y, por tanto, asume que existe una *ontología* que permite realizar de forma automática esta tarea. Si bien aún no se cuenta con este recurso, el trabajo resulta relevante pues por un lado, una vez que el sistema genera un marco comparativo (reglas) los usuarios pueden proporcionar casos para su análisis automático; por otro, identificar las discrepancias entre dos o más ordenamientos jurídicos es una tarea bastante ardua considerada como una disciplina dentro del derecho (el derecho comparado).

#### **2.1.5. LOIS**

LOIS (Lexical Ontologies for Legal Information Sharing) es un proyecto para la representación del conocimiento [53], su finalidad es la de construir un lexicón semántico capaz de soportar múltiples lenguajes, con la finalidad de que sea usado como fuente de conocimiento semántico y facilitar la recuperación de información en múltiples lenguajes.

La ontología está compuesta por alrededor de 35,000 conceptos en 5 lenguajes diferentes (Inglés, Alemán, Checo, Portugués, e Italiano). Los conceptos primarios han sido clasificados de acuerdo a la doctrina y a las descripciones de los expertos, agregando los términos legales que tienen una definición en la legislación. Cuando se detecta la correspondencia entre dos términos provenientes de diferentes lenguajes, automáticamente se ligan con un identificador único (*interlingua term*). El esquema se compone de palabras y conjuntos de sentidos de las palabras (*synsets*) al que se agregan además traducciones de expertos, y sentidos de palabras creados manualmente. Esta ontología se creó dentro del dominio de los derechos de los consumidores, y se espera que facilite las búsquedas tanto en uno como en múltiples lenguajes, así como comparar diferentes sistemas legales.

## **2.2. Sistemas basados en probabilidad**

Este tipo de desarrollos aplicados en el campo del derecho han estado centrados en predecir la forma en que un asunto será resuelto. Consideramos oportuno presentar este enfoque pues como veremos en la sección 3.1.1 el derecho pretende en la medida de lo posible, prever los efectos jurídicos y sociales, de las acciones que se llevan a cabo de forma cotidiana.

### **2.2.1. Sistema para determinar la relevancia de las pruebas**

En el trabajo desarrollado por T. Sjoerd, et al. [52], se introduce el concepto de *grafo de soporte* como un nivel intermedio entre las redes bayesianas y los modelos de argumentación. Las redes bayesianas son una representación de las dependencias existentes entre un conjunto de variables aleatorias, y pueden ser explotadas para determinar la independencia entre las variables y minimizar el número de las mismas, por ejemplo, en un asunto penal en el que un sujeto A fuerza a un sujeto B a lesionar a un sujeto C, si quisiéramos minimizar las causas que provocaron las lesiones del sujeto C, podríamos decir que estas fueron provocadas por el sujeto A, aún y cuando el sujeto B forme parte de la cadena causal. De hecho desde la teoría del delito podríamos decir que se presenta una *vis absoluta*, esto es, una fuerza física irresistible proveniente de un ser humano y, por tanto, sólo podríamos considerar al sujeto A responsable del hecho delictivo.

El concepto de *grafo de soporte* es introducido como una herramienta para explicar las redes bayesianas pues se estima que, aunque una red bayesiana es computacionalmente una representación eficiente de la distribución probabilística, resulta poco intuitiva para representar el soporte argumentativo, y por otra parte, los modelos argumentativos jurídicos pueden ser representados a través de un grafo.

Finalmente, en el trabajo descrito se considera que este puede facilitar el acercamiento entre el conocimiento legal y el conocimiento de expertos de otras áreas al facilitar la interpretación de las pruebas, además de ser una herramienta a partir de la cual se pueden construir diferentes propuestas argumentativas para un mismo caso.

### **2.2.2. Sistema para predecir la resolución de un problema legal**

Este trabajo fue presentado por Sabahi, et al. [45]. Su objetivo es asistir a los jueces en la determinación de sus fallos, y tiene su fundamento en la relación entre las reglas de la probabilidad y las reglas de la lógica. En este trabajo se utilizó lógica difusa para entender y modelar problemas imprecisos y toma como base la manera general en la que juzgan los jueces, asumiendo que *“la sabiduría del grupo converge en resultados verdaderos”*. Se toman en cuenta además los siguientes aspectos:

- Los fundamentos de una decisión judicial, se presentan bajo condiciones de incertidumbre, al existir una falta de certeza del juez sobre la evidencia, ambigüedades en el lenguaje, y un cierto grado de ignorancia.
- Los procesos judiciales se basan en la persuasión y, por tanto, no siempre se llega a los mismos resultados.

Para el desarrollo del sistema se tomaron un total de 73 casos, de los cuales 58 fueron usados para entrenamiento y 15 para pruebas. Previamente se realizó un análisis de los casos y se identificaron los factores más relevantes para la determinación de una decisión judicial en materia penal, a saber: mentiras, confesiones, declaraciones de los coacusados, evidencia, opiniones de expertos, evidencia documental, evidencia demostrativa, testigos, frecuencia criminal e historial de arrestos. Para saber qué tanto influyen estos aspectos se recabó la opinión de 35 abogados de distintas áreas quienes calificaron la importancia de estos factores en una escala de 0 a 10.

Se aplicó el enfoque descrito a casos judiciales y se compararon los resultados con las decisiones de los jueces, obteniéndose resultados razonables. Ya que el sistema determina si el acusado es o no es culpable, se considera que puede ser de utilidad para evitar apelaciones innecesarias y reducir el tiempo en el que los casos son decididos de forma definitiva.

## **2.3. Sistemas que utilizan métodos de clasificación**

La clasificación es uno de los métodos más utilizados en el área de la computación, esto se debe principalmente a que muchas cuestiones pueden ser modeladas como problemas de clasificación, incluso la toma de decisiones (condena/absuelve, otorga/niega). Como se verá más adelante, para utilizar estos métodos de forma exitosa, es necesario hacer una cuidadosa extracción de los rasgos, así como seleccionar adecuadamente el algoritmo de clasificación considerando el conjunto de datos con el cual se va a trabajar, así como las características de los mismos.

### **2.3.1. Sistemas de clasificación de normas jurídicas**

En primer término hacemos referencia al módulo de clasificación automática de normas jurídicas del sistema *Norm in Rete* (NIR) [5]. En este sistema se utilizó un modelo de espacio vectorial (VSM) para la representación de las normas. En esta representación las palabras fueron consideradas como dimensiones del espacio y las normas jurídicas como vectores en dicho espacio; el algoritmo utilizado para clasificar mostró un mejor desempeño cuando fueron asignados valores binarios a los vectores, esto es, cuando se asignó 1 en la dimensión que representaba la palabra  $n$  si la palabra se encontraba presente en la norma y 0 si la palabra no se encontraba en la norma. Se utilizó además un modelo de clasificación supervisada empleando una Máquina de Soporte Vectorial (SVM). Para evaluar el clasificador fueron utilizadas 582 normas que pueden ser clasificadas en alguna de las once clases mostradas en la tabla 1. En la misma tabla se muestran las medidas *precision* y *recall* que fueron obtenidas al evaluar el clasificador.

	<b>Tipo de norma</b>	<i>precision</i>	<i>recall</i>
<b>Clase 1</b>	normas de revocación	0.945	0.986
<b>Clase 2</b>	normas que contienen una definición	0.875	0.700
<b>Clase 3</b>	normas que delegan	1.0	1.0
<b>Clase 4</b>	normas cuya regulación específica se dará a través de actos administrativos	1.0	1.0
<b>Clase 5</b>	normas de prohibición	0.455	0.385
<b>Clase 6</b>	normas de reserva	0.615	0.444
<b>Clase 7</b>	normas de inserción	0.967	0.975
<b>Clase 8</b>	normas de obligación	0.885	0.915
<b>Clase 9</b>	normas de permisión	0.385	0.333
<b>Clase 10</b>	normas penales	0.960	0.992
<b>Clase 11</b>	normas de sustitución	0.964	0.973

**Tabla 1. Clases y evaluación de los resultados obtenidos por el clasificador automático de normas jurídicas del sistema NIR.**

Como se aprecia en la tabla, este clasificador muestra resultados superiores a 0.9 tanto en *precision* como en *recall*. De manera similar, en un trabajo realizado en el *Leibniz Center for Law* de la Universidad de Amsterdam [35] se logró clasificar de forma correcta el 94% de un total de 530 normas jurídicas. Para la realización de este último trabajo fueron utilizadas 20 leyes, en las cuales se identificaron 81 estructuras que se utilizaron para clasificar las normas en 14 clases distintas (definiciones, derechos, obligaciones, penas, disposiciones transitorias, ...) una condición necesaria para utilizar este método de clasificación, es que la *estructura de la legislación* se encuentre previamente marcada por algún lenguaje de etiquetado de textos como XML, en los experimentos presentados se ocupó el sistema MetaLex<sup>3</sup> para el marcado automático.

Un tercer trabajo realizado con este enfoque [19], intentó clasificar dos tipos de normas: derechos y obligaciones, a partir de una Gramática Libre de Contexto (CFG). Para obtener la gramática, un conjunto de oraciones (contenidas en las normas jurídicas) fue etiquetada manualmente. Bajo el supuesto de que los

---

<sup>3</sup> <http://www.metalex.eu/>

elementos de las normas siguen patrones específicos, la clasificación se basa en las características sintácticas y semánticas de las normas jurídicas. El trabajo tuvo éxito para identificar los derechos y sólo algunas obligaciones, y se considera que podría ayudar a detectar las inconsistencias existentes en un marco legal.

Otro enfoque que deseamos incluir en esta sección, fue realizado en el contexto del *common law*, específicamente en Australia, el cual, forma parte de un proyecto para la generación automática de resúmenes en el área legal. En este trabajo se intentó clasificar las referencias que realizan respecto a los precedentes.

En el derecho anglosajón el precedente es la fuente más importante del derecho, y tiene su base en la doctrina del *stare decisis* (*strare decisis et non quieta movere*, es decir, estar a lo resuelto y no perturbar lo que ya está firme) la cual es una regla que atribuye fuerza jurídica a las sentencias dictadas y obliga a su observancia en todos los casos futuros que sean similares. Sin embargo, un tribunal puede desvincularse de un precedente solamente cuando se establecen diferencias (*distinction*) entre el caso examinado y el precedente [23]. El trabajo realizado por Filippo Galgani y Achim Hoffmann [20] consistió en clasificar si un precedente citado terminó siendo aplicado o no (al establecerse una distinción) para ello se construyó un analizador sintáctico con el que se procesaron los reportes de sentencias de la Corte Federal Australiana, y considerando la manera como fueron citados los precedentes, se determinó si éstos habían sido aplicados o no. Para ello se utilizó una metodología de adquisición del conocimiento (específicamente *Ripple Down Rules*). Con esta metodología se alcanzó una exactitud (*accuracy*) del 88%. En este trabajo resaltó el hecho de que dos abogados pueden no estar de acuerdo en la clasificación del precedente.

### **2.3.2. Sistemas para la detección de argumentos en sentencias**

La identificación de argumentos permite a quien toma una decisión seguir más de cerca el proceso de razonamiento, y alcanzar mejores decisiones [46].

En el trabajo que procederemos a exponer [38] la detección de argumentos se trató como un problema de clasificación, en donde se intenta reducir la cantidad de información existente en las decisiones judiciales. En efecto, debido a los requisitos de fondo y de forma que deben cumplir las sentencias, estas cuentan con una gran cantidad de información con respecto a un asunto, sin embargo, sólo una pequeña parte de esa información se utiliza realmente al determinación de la resolución.

El trabajo se avocó a detectar si una oración es una premisa, una conclusión o una oración que no contiene argumentos. Para ello las oraciones fueron consideradas como ventanas de  $n$  elementos. Tomando en cuenta que el contexto de una oración es importante para determinar si ésta contiene un argumento, se utilizó una lista de 286 palabras o secuencias de palabras que indican la presencia de un argumento (frases como “esto es porque”, “es por esto que”).

Para realizar esta tarea se ocupó un *corpus* (conjunto de documentos) de 54 documentos de la Corte Europea de Derechos Humanos (25 casos y 29 autos de admisión) cada uno en promedio con 145 oraciones y cada oración con 49 *tokens* en promedio, obteniéndose un total de 12,904 oraciones. Cada oración se representó como un vector de palabras y fueron manualmente etiquetadas para posteriormente entrenar un clasificador. En total se encontró que 10,133 (78%) oraciones no contenían argumentos; de las restantes 2771, 416 fueron etiquetadas como conclusiones y 2355 como premisas. Para la realización de pruebas utilizó un clasificador bayesiano y un clasificador de máxima entropía alcanzándose 82% de exactitud (*accuracy*).

Un proyecto similar (ARGUMENTUM) se encuentra en desarrollo en el Instituto de informática jurídica de la universidad del Sarre (Alemania) [27]; en éste, se utiliza un corpus de 60 decisiones judiciales del Tribunal Constitucional Federal Alemán, y se emplea minería de texto para la identificación, análisis y recomendación de estructuras argumentativas. Su objetivo es clasificar cada oración como una premisa o un explicación utilizando para ello una SVM como método de clasificación y un VSM para la representación de los datos (aplicando la medida TF-IDF) aunque también pretende experimentar con una Gramática Libre de Contexto (CFG) generada a partir de un Etiquetado de Categorías Gramaticales (*POS tagging*) y Reconocimiento de Entidades Nombradas (NER).

## **2.4. Sistemas que utilizan métodos de recuperación de información**

Los sistemas que hemos revisado hasta ahora han utilizado técnicas cuyo estudio se ubica principalmente en el área de la Inteligencia Artificial. La recuperación de información textual, si bien utiliza estas técnicas, se encuentra ubicada en el área del Procesamiento del Lenguaje Natural. Tomando en cuenta que gran parte de la información generada en el ámbito legal se encuentra en forma de texto, existen diversos trabajos que se han desarrollado desde esta óptica. Exponemos aquí algunos de ellos.

### 2.4.1. Sistemas para la recuperación de normas jurídicas relacionadas

Para esta sección, el trabajo más relevante que hemos encontrado fue realizado en Centro de Investigación en Computación del IPN (CIC-IPN) por Monroy A., Calvo H., Gelbukh A., et al. [39]. Su objetivo es ayudar en el proceso de consulta legal, y recuperación de información de una pregunta formulada en lenguaje natural. En el proyecto se consideró un aspecto muy importante de los textos legales, el hecho de que se encuentran estrechamente relacionados entre ellos, a partir de las referencias que existen entre los artículos.

Considerando esta característica se construyó un *grafo no dirigido*, en el que los nodos representan artículos (8,987) y las aristas representan tanto las referencias entre ellos, como su grado de similitud. La similitud de los artículos se obtuvo representando los artículos en un VSM y utilizando una función de semejanza coseno. La importancia de los nodos fue obtenida aplicando el algoritmo *page rank*.

La recuperación de información se realiza a partir de una pregunta formulada en lenguaje natural, esta pregunta es agregada al grafo y en combinación con el algoritmo Dijkstra (algoritmo para obtener el camino más corto) se obtiene una lista de artículos *rankeados*. De esta lista se toman los primeros n artículos y considera que la respuesta a la pregunta es correcta, si la misma contiene todos los artículos relevantes (estos han sido identificados previamente). Si contiene únicamente un subconjunto de éstos la respuesta se considera parcialmente correcta y, finalmente, si no contiene ninguno de los artículos relevantes, la respuesta se considera incorrecta. El sistema se evaluó con 40 preguntas de las cuales 36 fueron respondidas de manera correcta y 4 de forma parcial.

Otro trabajo que toma en consideración las referencias entre artículos fue realizado en el *Leibniz Center for Law* de la Universidad de Amsterdam [55]. En este trabajo se intentan determinar automáticamente los contextos relevantes de un artículo, y con ello proporcionar información adicional a los usuarios con respecto al mismo. De manera similar al trabajo previamente reseñado, se analizaron 6 leyes, y se construyó un grafo en el que los nodos representan a los artículos y las aristas representan las citas de un artículo hacia otro. De esta manera fueron identificados los nodos que contienen más referencias; se considera que estos pueden tener mayor relevancia sobre otros al intentar determinar el contexto de un artículo dado.

El tercer trabajo con este enfoque fue localizado en el módulo para la detección de referencias del sistema NIR [18], sin embargo, los detalles de su implementación no están disponibles de forma abierta.



## 2.4.2. Sistema para la extracción de argumentos en normas jurídicas

En esta sección describimos el módulo para reconocimiento automático de elementos estructurales de una norma jurídica del sistema NIR [5]. El propósito de este módulo consiste en detectar los roles semánticos en una norma. Se toma como punto de partida el texto de una disposición jurídica, así como la clase que corresponde a dicha disposición (determinada de forma automática por el clasificador reseñado en la sección 2.3.1). El sistema se encarga de asignar las etiquetas semánticas que le corresponden a los elementos de la disposición como son: destinatario de la norma, verbo rector de la norma y disposición jurídica en concreto.

En la etapa de pre-procesamiento fueron etiquetadas las categorías gramaticales de las palabras (*POS tagging*) y posteriormente se hizo un análisis sintáctico superficial (*chunks*). A partir de este análisis sintáctico se localizan elementos relevantes de las normas tales como verbos, sujetos, oraciones y frases. Una vez localizados estos elementos se asignó una plantilla a la norma, dependiendo de la clase de norma jurídica que le correspondía. Finalmente los espacios en la plantilla se cubrieron con los elementos identificados en el análisis sintáctico.

Para la evaluación de este método fueron utilizadas 473 normas jurídicas, de las cuales en el 82.09% fueron identificadas de forma correcta todos los roles semánticos contenidos en las normas. En el 15.35% se identificó correctamente sólo un subconjunto del total de roles semánticos de la norma, y sólo en el 2.56% restante el sistema falló completamente, pues no logró identificar ningún rol semántico de manera correcta.

En este capítulo, hemos citado en diversas ocasiones en sistema *Norm in Rete*<sup>4</sup> (NIR, esto es, legislación en la red) al considerarlo como un sistema integral, hacemos referencia en esta sección a los cuatro módulos que lo componen:

- **Cross-Reference Parser** (Analizador para la detección de referencias). Módulo diseñado para detectar las referencias entre normas jurídicas [18].
- **Structure Parser** (Analizador de estructuras). Módulo diseñado para convertir el contenido de una legislación en formato XML [18].

---

4 Sistema desarrollado en Italia, propuesto por el Ministerio de Justicia y el Centro Nacional para las Tecnologías de la Información en la Administración Pública. En el desarrollo del sistema participaron además el Instituto de Lingüística Computacional del Consejo Nacional de Investigación Italiano, el Departamento de Sistemas e Informática de la Universidad de Florencia y el Instituto de Teoría y Técnicas de la información Jurídica.

- **Provision Automatic Classifier** (Clasificador automático de normas jurídicas). Módulo para clasificar una norma jurídica entre once clases distintas [18].
- **Provision Arguments Extractor** (Extractor de argumentos de normas jurídicas). Módulo para reconocimiento automático de elementos estructurales de una norma jurídica [50, 4].

### 2.4.3. Sistema para generación de resúmenes de sentencias judiciales

El método desarrollado por Mi-Young Kim, et al. [30], tuvo como objetivo la extracción de resúmenes de sentencias judiciales, seleccionando de forma automática un conjunto de oraciones representativas, sin que fuera necesario indicar la tasa de compresión. Para ello se construyó un grafo dirigido no conexo, en el que cada subgrafo es conexo e indica un conjunto de oraciones que comparten un tema dentro de un documento. Los nodos del grafo representan oraciones, y las aristas representan la probabilidad de que una oración se encuentre dentro de otra; por ello, los pesos de las aristas son asimétricos. Dado que el grafo es no conexo, se afirma que se puede asegurar la diversidad de las oraciones seleccionadas, escogiendo sólo ciertas oraciones de los subgrafos conexos.

Para la generación del resumen, se analiza en primer lugar la sección de los resolutivos de la sentencia; ya seleccionadas las palabras y las oraciones más frecuentes en los resolutivos, se procede a buscar en el grafo las oraciones con mayor cohesión. Una vez que se han seleccionado todas estas oraciones se procede a complementar el resumen buscando en los subgrafos restantes las oraciones con mayor cohesión.

## Marco teórico

El objetivo central de este trabajo es facilitar la búsqueda de fundamentos jurídicos en la resolución de problemas legales. El realizar esta tarea a partir de métodos computacionales implica el acercamiento de tres áreas del conocimiento: Ciencias de la computación, Lingüística y Derecho. En este capítulo, presentamos un panorama general respecto a estas disciplinas.

### 3.1. Consideraciones teórico-jurídicas

Para resolver adecuadamente un problema de recuperación de información en un dominio específico, debemos entender en primer lugar las necesidades particulares del dominio. Para ello, debemos profundizar en dos aspectos: la función que se desempeña, y la forma en que se realiza. Por tanto, en esta sección responderemos, por una parte, a la pregunta ¿Cuál es la función del derecho? por otra ¿Cómo es que el derecho realiza su función? Para esta última pregunta presentamos dos enfoques: el constructivista y el realista; decidimos presentar estas dos visiones del derecho a efecto de mostrar la complejidad y la indeterminación que existe en el ámbito jurídico. Hacemos la aclaración de que este trabajo se realizó desde una perspectiva constructivista, sin embargo, el enfoque realista nos ayuda a visualizar las limitaciones del mismo.

#### 3.1.1. La función del derecho

El derecho tiene papel preponderante para todos los seres humanos, lo encontramos presente en todas las facetas de nuestra vida, incluso desde antes del nacimiento (*nasciturus*<sup>5</sup>) y hasta después de la muerte (*de cuius*). Siguiendo a Recánsens “El derecho no es lo más importante ni lo más alto que hay en la vida; pero sin él

---

5 Artículo 22 del código civil federal. La capacidad jurídica de las personas físicas se adquiere por el nacimiento y se pierde por la muerte; pero desde el momento en que un individuo es concebido, entra bajo la protección de la ley y se le tiene por nacido para los efectos declarados en el presente Código.

el hombre no puede vivir” [47]. En este sentido, se ha aceptado que el derecho es una forma de control social, puesto que está encaminado a regular la conducta del hombre en sociedad. Esta forma de describir al derecho intenta responder a la pregunta ¿Cuál es la función del derecho? considerando una perspectiva muy limitada respecto al cómo (a través del ordenamiento legal) pues el derecho es más que un conjunto de normas jurídicas.

El derecho positivo, se gesta en virtud de una urgencia de certeza y seguridad [47], esto es, la necesidad de saber a qué atenernos, el conocer a qué estamos obligados, y saber qué conductas se encuentran permitidas y cuáles se consideran prohibidas. Denominamos a esto certeza de la acción y expectativas jurídicas; lo ilustramos con una situación muy simple:

*Una persona realiza un contrato de compraventa con una inmobiliaria sobre un inmueble, específicamente, un departamento que aún se encuentra en construcción, y con respecto al cual se pacta su entrega una vez que se hayan terminado las maniobras de construcción que se realizan al mismo; por su parte, el comprador se compromete a pagar el inmueble de acuerdo a un calendario de pagos.*

Para realizar esta acción (adquirir el departamento) el comprador no necesita saber qué cosas se encuentran dentro del comercio y cuáles no, tampoco que su conducta se encuentra regulada por el código civil, ni que el contrato requiere una forma específica; lo que sabe cualquier persona en general es que se pueden comprar departamentos, y que nadie va a la cárcel por ese sólo hecho (certeza de la acción). Lo que ambas partes si tienen claro, es lo que esperan recibir (expectativas); la inmobiliaria, espera recibir *su dinero* y, la persona espera recibir *su departamento*.

Con respecto a la certeza de la acción, Flavio López considera que *“la certeza del derecho es certeza de la acción”* [32]; esto es, asegurar la calificación de los comportamientos posibles. En efecto, el convivir en sociedad nos da a todos, un conocimiento mínimo de cómo son calificadas las acciones que en ella se realizan (legales/ilegales o en términos de Luhmann: derecho/no derecho); no es lo mismo comprar un departamento que comprar un arma de fuego, y de hecho, a partir de la posibilidad de una sanción se puede reconocer si la expectativa va en el sentido del derecho o en contra del mismo [33].

Continuando con nuestro ejemplo, supongamos que *el comprador, de acuerdo con el calendario establecido, ha realizado el pago total del departamento, sin embargo, la inmobiliaria no hace entrega del mismo en las condiciones estipuladas en el contrato*. En este caso, la persona puede defender su aspiración de certeza e invocar la defensa del derecho, para que su aspiración, no se vea defraudada [32].

En este sentido, el derecho tiene como función el asegurar nuestras expectativas (hacer que la inmobiliaria entregue el departamento en las condiciones pactadas). Para ello, cuenta con mecanismos (como el *proceso*) que permiten hacer más probable el cumplimiento de una expectativa. Esto implica que el derecho debe ser previsible en la medida de lo posible, y aunque esto no garantiza que todas las expectativas puedan quedar satisfechas nos permite un mayor grado de confianza con respecto a las conductas que realizamos [33].

Ahora bien, una expectativa puede no ser alcanzada aunque esté amparada por el derecho; en tal caso, el derecho debe ofrecer al menos sustitutos del cumplimiento exigido, pues como se ha llegado a mencionar “el derecho no puede estar continuamente diciendo: Usted tiene razón, pero desgraciadamente no podemos ayudarle” [33].

Siguiendo con el ejemplo, podemos imaginar que *la inmobiliaria no tiene forma alguna de cumplir con el contrato*; tenemos entonces, que incluso, en estos casos se deben asegurar las expectativas a quienes están en su derecho, transformando la expectativa en castigo.

Luhmann agrega un nivel de complejidad en este tema al señalar “el sistema jurídico desde la perspectiva de la totalidad, opera con base en la seguridad de la expectativa normativa de la expectativa normativa” esto es, tener la certeza de que a quien le corresponde aplicar la norma, efectivamente va a aplicar la norma. Con relación a este punto haremos algunas consideraciones en la sección 3.1.3.

Quizá pudiera parecer extraño el que no nos hayamos referido en esta parte a la solución de conflictos; consideramos que tal situación va más allá de la función del derecho y asumimos la posición de Luhmann quien las considera *prestaciones* del derecho, aunque reconocemos posturas como la de Alejandro Nieto, en el sentido de que el objetivo último del derecho es la solución de conflictos [40].

### **3.1.2. Visión constructivista del derecho**

El derecho realiza su función a través del sistema jurídico<sup>6</sup>, éste se compone en general por un conjunto de normas y los mecanismos para aplicarlas a situaciones concretas. Expondremos aquí dos visiones clásicas en la teoría del derecho que analizan estos aspectos.

---

<sup>6</sup> Nos interesa resaltar el hecho de que el Derecho realiza su función a través del propio Derecho.

### 3.1.2.1. La jurisprudencia analítica de H. L. A. Hart

Herbert Lionel Adolphus Hart fue uno de los filósofos más importantes del siglo XX. Su trabajo más trascendente lo encontramos en su obra *The concept of law* [25]. En ella se señala en primer lugar, que la descripción lingüística del derecho mediante una noción simple de regla<sup>7</sup>, supone una pérdida de semántica sin la cual no podemos elucidar las formas más elementales del derecho.

A través de una noción simple del derecho no podemos responder a las preguntas ¿por qué el derecho es válido? o ¿cuál es la diferencia entre una obligación moral y una obligación jurídica?. El aspecto que aquí nos interesa de la teoría de Hart es el relativo a ¿Qué son las reglas y en qué medida el derecho es una cuestión de reglas?.

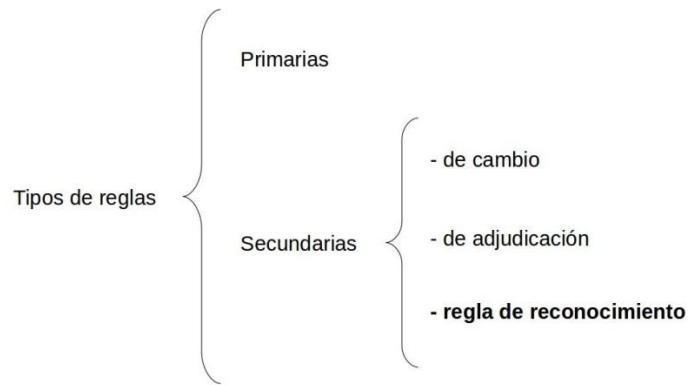
Hart distingue entre dos tipos de reglas: reglas primarias y reglas secundarias. Las reglas primarias imponen deberes positivos (deberes de hacer) o negativos (deberes de no hacer) a los individuos; las reglas secundarias confieren potestades públicas o privadas y versan sobre la manera en que las reglas primarias pueden ser violadas o cumplidas, agregadas o eliminadas.

Las reglas secundarias se dividen a su vez en reglas de cambio, reglas de adjudicación y son complementadas por la regla de reconocimiento. Una regla de cambio faculta a una persona o grupo de personas a introducir nuevas reglas primarias y a dejar sin efecto las anteriores. Por su parte, una regla de adjudicación, identifica a los individuos que pueden juzgar, define procedimientos a seguir, y confiere un estatus especial a las declaraciones judiciales.

En cuanto a la regla de reconocimiento, es una “regla para la identificación incontrovertible de las reglas primarias de obligación” [25], su existencia es una cuestión de hecho que consiste en la aceptación práctica de parámetros para la identificación de normas en un sistema legal.

---

<sup>7</sup> Una de las definiciones de regla que da la real academia española es: *Aquello que ha de cumplirse por estar así convenido por una colectividad.*



**Figura 1. Tipología de las reglas de acuerdo a jurisprudencia analítica de Hart.**

En este modelo, las normas tienen una *textura abierta*, en primer lugar por su generalidad, es decir, por estar dirigidas a clases de personas o a clases de cosas; en segundo, porque permanecen vigentes por largos periodos y se aplican a situaciones no previstas en el momento de su creación; se reconoce también la unicidad de la carga semántica de las normas, es decir, las normas son entendidas como un todo.

Para Hart, dentro del sistema normativo existe un núcleo semántico duro, el de los *casos fáciles*, en el que el significado de la regla es claro, y existe un acuerdo general de aplicabilidad en el que la identificación de las normas que regulan alguna situación en concreto parece hacerse de forma automática; el problema hermenéutico radica sólo en la forma de su aplicación. Existe además una “zona de penumbra”, los *casos difíciles* en el que el significado es débil (cuando un hecho no se puede subsumir claramente en una norma jurídica) y debe ser establecido con base en las acciones resultantes de los actos de habla emitidos por los tribunales al resolver casos difíciles (precedentes). En estos casos, el ámbito de discrecionalidad que deja el lenguaje puede ser muy amplio, de modo que, si bien las determinaciones pueden no ser arbitrarias o irracionales, son en realidad una cuestión de elección.

### **3.1.2.2. Elementos de la teoría del derecho de Ronald Dworkin**

Otro de los juristas más estudiados es precisamente Ronald Dworkin, y uno de sus trabajos que más han sido analizados versa sobre la crítica a la regla de reconocimiento de Hart, Dworkin señala que no todas las normas jurídicas son reglas precisas, pues, junto a ellas encontramos los principios y las políticas. Las normas para Dworkin son reglas de conducta precisas, cuya aplicación determina por sí sola el resultado de un caso

concreto; los principios, son estándares que deben ser observados por una exigencia de justicia, equidad o alguna otra dimensión de la moralidad y, por tanto, no determinan de forma inmediata el desenlace de un caso, dado que su aplicación requiere usualmente de la ponderación de principios encontrados; las políticas son modelos que proponen objetivos que, de ser alcanzados, mejoran algún rasgo económico, político o social de la comunidad [15].

En la sección anterior, señalamos que para Hart casos difíciles son una cuestión de elección (tesis de la discrecionalidad) Dworkin confronta esta teoría de la adjudicación (utilizando para ello al juez Herbert) con la tesis de la respuesta correcta (personificada por el juez Hércules) enfatizando que los jueces no tienen discrecionalidad para crear el derecho pues siempre existirán los principios que servirán para resolver.

La adjudicación para Dworkin es una tarea gobernada por la virtud, siendo ésta un compromiso de las autoridades públicas de tratar a los particulares de manera consistente con los principios de moralidad política plasmados en las instituciones. El derecho como integridad, se refiere precisamente a que los casos similares, sean tratados de manera similar, y evitar las distinciones caprichosas.

Para este autor, la práctica jurídica si bien es interpretativa, el proceso de interpretación tiene el propósito de construir el verdadero sentido de las reglas. Este proceso está estructurado de la siguiente forma:

- **Fase pre-interpretativa.** Identificación de las reglas y principios presentes en la práctica jurídica. Responde a la pregunta ¿Qué normas son aplicables a un caso concreto?.
- **Fase interpretativa.** Justificación moral y política. Responde a la pregunta ¿por qué son aplicables las normas identificadas al caso concreto?.
- **Fase post- interpretativa.** Formulación de adaptaciones a la práctica jurídica con el fin de acercar las reglas y los principios a los requerimientos de la justificación. Responde implícitamente a tres preguntas ¿qué diferencias hay entre el asunto a resolver y lo ya establecido?, ¿por qué a pesar de esas diferencias el asunto es similar a lo ya establecido? y si el asunto es similar a lo ya establecido, ¿por qué le son aplicables las mismas reglas?.

Nuestro trabajo se centra precisamente en la fase pre-interpretativa, es decir, en identificar (de manera automática) las normas jurídicas que son aplicables a un caso concreto, pues como señala Alejandro Nieto “el conocimiento jurídico práctico es consecuentemente la búsqueda y hallazgo de la solución que ya está en la



ley” [40], aunque tomamos en cuenta sus planteamientos con respecto a que en la aplicación del derecho no basta con hacer un ejercicio de rastreo, sino que es preciso desarrollar una actividad intelectual que consiste primero en identificar los hechos jurídicamente relevantes y luego subsumirlos en un concepto abstracto tipificado en el ordenamiento, nos inclinamos a pensar que subsunción también puede realizarse de forma automática; no obstante, en este trabajo nos limitamos precisamente a realizar este “ejercicio de rastreo”.

### **3.1.3. Visión realista del derecho**

De cierto modo, evocar en este trabajo el debate Hart-Dworkin es resaltar que el derecho no es algo definido, y claramente estos autores lo tenían presente. Al inicio de su obra, Hart señala: “pocas preguntas referentes a la sociedad humana han sido formuladas con tanta persistencia y respondidas por pensadores serios de maneras tan diversas, extrañas y aún paradójicas, como la pregunta ¿qué es el derecho?”; Alejandro Nieto emite una opinión similar “los juristas no se han puesto de acuerdo en el objeto de sus preocupaciones” [40]; de hecho, en su crítica a la razón pura Kant llega a mencionar: “Noch suchen die Juristen eine Definition zu ihrem Begriff vom Recht” es decir, “todavía buscan los juristas una definición para su concepto del derecho”. Por su parte, tanto Dworkin como por Frank citan en sus trabajos la siguiente frase de Learned Hand (quien fue juez de los Estados Unidos) “Debo decir que si fuera un litigante temería a un pleito [legal] por sobre casi todas las demás cosas, salvo la enfermedad y la muerte” [14, 28].

En el contexto de este trabajo, el que el derecho no tenga una definición sólo nos muestra la complejidad a la que nos enfrentamos al estudiar algo que no está determinado y, por tanto, decidimos incluir algunos de los cuestionamientos más fuertes que se han planteado con respecto al ordenamiento jurídico y la manera en que resuelven los jueces; éstos, han sido la base para abordar el problema de la recuperación de normas jurídicas de una manera diferente, más completa, y más cercana la realidad; además han sido de utilidad para identificar las limitaciones de nuestro trabajo.

Nosotros identificamos dos etapas del realismo. La primera, centrada en una crítica a la discrecionalidad judicial que culminó con el establecimiento del principio de legalidad; la segunda, actualmente vigente, abarca distintos aspectos entre ellos, la indeterminación de las normas jurídicas, la falsedad de la idea de una autonomía total del derecho, la manipulación política, etc. No omitimos mencionar que no estudiamos estas etapas de forma separada, pues al ser temas de una amplitud considerable, requieren de un análisis detallado que va más allá de los alcances de este trabajo, por lo cual, nos enfocamos únicamente a dos aspectos: la indeterminación de las normas jurídicas y crítica a la discrecionalidad judicial.

### 3.1.3.1. Indeterminación de las normas jurídicas

El primer factor que dificulta el análisis normativo es la excesiva cantidad de leyes que existen. Luigi Ferrajoli ha denominado a este fenómeno *inflación legislativa* y al respecto hace la siguiente observación:

*“las leyes, en todos los ordenamientos avanzados, se cuentan ahora ya por decenas de miles y están formuladas en un lenguaje cada vez más oscuro y tortuoso, dando lugar a veces a intrincados enredos y laberintos normativos” [17].*

Esta no es una cuestión menor, y lo mostramos con algunos números. Para realizar nuestro trabajo recopilamos un total de **5,147** ordenamientos jurídicos, sin que esta colección constituya todas las leyes y reglamentos que se encuentran “vigentes” en México. De hecho, reconocimos 135 ordenamientos que están fuera de nuestro catálogo. Además, la cantidad de artículos es variable de una ley a otra, pues aunque hay legislaciones con no más de 30 artículos y hay otras bastante extensas con más de 1000 artículos (como la Ley Federal del Trabajo) esta es una cantidad de normas imposible de manejar para cualquier abogado.

El problema se complica aún más cuando consideramos la dinámica del proceso legislativo; por ejemplo, sólo en la LXII legislatura fueron aprobadas 42 nuevas leyes, 35 reformas constitucionales, 419 reformas a leyes existentes, haciendo un total de 496 modificaciones<sup>8</sup> al ordenamiento jurídico. Números similares se presentaron en la LXI legislatura con 473 modificaciones al sistema legal.

A esto debemos agregar que los miembros del parlamento casi nunca logran expresarse con claridad en los textos que producen [40]. En efecto, la completa falta de técnica legislativa para establecer reglas de sustitución de una norma a otra, dificulta aún más las tareas de interpretación, así encontramos fórmulas nocivas como: *“Se derogan todas las disposiciones que se opongan al presente Decreto”*. Esta remisión la podemos encontrar en diversos artículos transitorios en distintas leyes; en este caso, sólo nos referimos al artículo cuarto transitorio del decreto publicado en el diario oficial de la federación el 15 de octubre de 2012, mediante el cual se reformaron diversas disposiciones constitucionales. Este tipo de disposiciones generan preguntas que parecen no tener una respuesta clara, por ejemplo, ¿qué se entiende por todas las disposiciones, se refiere a todas las normas vigentes o sólo a aquéllas relacionadas con la materia?, ¿qué sucede con las disposiciones afectadas de manera indirecta?, si la norma que se promulga, resulta ser inconstitucional, ¿todas las disposiciones que se oponen siguen vigentes?, ¿o debemos considerar que es voluntad del legislador dejarlas fuera del ordenamiento jurídico?.

---

<sup>8</sup> Informes estadísticos de la cámara de diputados consultables en [http://www.diputados.gob.mx/sistema\\_legislativo.html](http://www.diputados.gob.mx/sistema_legislativo.html)

En retrospectiva, la fase pre-interpretativa que describe Dworkin y que se refiere a la identificación de reglas y principios, se vuelve sumamente compleja. En esta línea, nos atrevemos a afirmar que no hay jurista en el país que conozca todo el ordenamiento jurídico. Para reducir esta complejidad jurídica, nosotros optamos por un método computacional: la recuperación de información.

Existe además otra crítica al sistema normativo que es importante resaltar y, aunque no fue considerado en nuestro trabajo, nos ayuda a entender los alcances y limitaciones del mismo. Esta crítica se refiere al contenido normativo: Duncan Kennedy, en su crítica a la adjudicación [29] apunta dos aspectos, por un lado, el grado en que las normas reproducen sistemas jerárquicos, otorgando poder a unos grupos en detrimento de otros; por otro, el grado en el que el sistema normativo, contiene lagunas y ambigüedades.

Para Kennedy, el discurso de la coherencia y neutralidad del derecho hace parecer justo el funcionamiento de un sistema jurídico que tiende a favorecer los intereses de quienes tienen un mayor poder económico para defenderlos. Agrega además, que existen principios, reglas e instituciones basadas en ideologías contradictorias, como el individualismo (basado en la responsabilidad civil subjetiva, en donde quien causa un daño, sólo se ve obligado a repararlo si actuó con culpa) y el altruismo radical (basado en la responsabilidad civil objetiva, en donde quien causa un daño, debe repararlo, independientemente de si actúa o no con culpa) cuya contradicción se ve reflejada en el ordenamiento. De esta forma, el derecho en general y decisiones judiciales en particular, son una extensión de la política.

Estas contradicciones las podemos verificar en casos específicos en los que dos derechos fundamentales se contraponen, *v. gr.*, derecho a la información frente al derecho a la intimidad (¿los ciudadanos tenemos derecho a conocer el estado de salud del presidente, o esta información se encuentra protegida por su derecho a la intimidad?). Aunque para resolver estas “*aparentes contraposiciones*” efectivamente se puede realizar un ejercicio de ponderación, resulta sustancial el observar estas dualidades.

Los cambios de criterio respecto a ciertos temas reflejan también la presencia de ideologías. Por ejemplo, en 2008 la Suprema Corte de Justicia de la Nación, al resolver la acción de *inconstitucionalidad* 146/2007 determinó la constitucionalidad de las reformas al código penal y la ley de salud (ambos ordenamientos para el distrito federal) que permiten interrumpir el embarazo durante las primeras 12 semanas de gestación. Por el contrario, en 2011, en las acciones de inconstitucionalidad 11/2009 y 62/2009 al no alcanzarse una mayoría de 8 votos, fueron consideradas como constitucionalmente válidas las reformas a las legislaciones de los estados de Baja California, y San Luis Potosí, que protegen la vida desde el momento de la concepción y que en consecuencia criminalizan el aborto. Aunque la integración de la corte cambió de 2008 a 2011, la votación se

mantuvo en el mismo sentido. El voto decisivo fue el de la ministra Luna Ramos, que estuvo presente en ambas decisiones y que cambió su criterio. Resaltó también la opinión que externaron diversas organizaciones y actores políticos, en aras de influir en la decisión de la corte, principalmente el activismo del presidente Felipe Calderón. Lo que llama la atención es que bajo un mismo marco constitucional, tenemos estados en los que interrumpir el embarazo es un delito y otros en los que se encuentra legalmente permitido.

### 3.1.3.2. Crítica a la discrecionalidad

La primera crítica que se realizó a la discrecionalidad se presentó durante el establecimiento del principio de legalidad, quien realiza este análisis es el reconocido jurista Luigi Ferrajoli [17]. Él señala que en derecho pre-moderno no existía un sistema unitario y formalizado de fuentes positivas, las normas eran establecidas a través de la jurisprudencia y la doctrina; y dependían no de su forma de producción, sino, de la intrínseca racionalidad de sus contenidos. El paso de este derecho, al estado legislativo de derecho (y posteriormente al estado constitucional de derecho) se da a partir del establecimiento del principio de legalidad como regla de reconocimiento del derecho válido, naciendo con él las garantías de certeza, seguridad e igualdad ante la ley. Se prefirió en este caso la racionalidad de la ley, frente a lo que Hobbes llamó la sabiduría de los jueces desordenados.

Otra gran crítica a la discrecionalidad de los jueces vino desde del ámbito penal, con el establecimiento del principio *nullum crimen sine lege*. En su tratado de los delitos y la penas Beccaria afirma *“sólo las leyes pueden decretar las penas y los delitos; y esta autoridad debe residir únicamente en el legislador”* [8]; ahora bien, para Beccaria los jueces no tienen autoridad de interpretar la ley, sólo de aplicarla mediante un silogismo perfecto; en otro caso abren la puerta a la incertidumbre, y se refiere a esto de la siguiente manera *“El espíritu de la ley sería, pues la que resulta de la buena o mala lógica de un juez, de su buena o mala digestión; dependería de la violencia de sus pasiones, de la flaqueza del que sufre, de la relaciones que tuviese con el ofendido y de todas aquellas pequeñas fuerzas que cambian las apariencias de los objetos en el ánimo fluctuante del hombre”*.

Hoy en día es aceptado (aunque no de manera general) que un juez puede sacrificar la justicia en beneficio de la ley, y viceversa, justificando su decisión de la forma que le parezca más correcta, apoyándose para ello en los principios del conocimiento jurídico [40]. Con respecto a este punto Kennedy elabora un ensayo [13] en el que describe cómo un juez (a quien presenta como el tercer juez, los primeros dos son el juez Herbert y el Juez Hércules ideados por Dworkin) conforma su decisión, en un asunto en el que se presenta un conflicto entre lo

que la ley exige y la sentencia a la que el juez quiere llegar (esta puede atender a la justicia pero también puede obedecer a otras motivaciones como un soborno, o un arreglo). En tal caso, el juez puede simplemente aplicar la ley, o desplegar una estrategia argumentativa compleja que culmine con la construcción de un razonamiento jurídico convincente que sustente satisfactoriamente la sentencia anhelada.

De hecho, este autor señala que, el juez se encuentra en un dilema siempre que percibe una brecha entre lo que considera justo y lo que dicta el derecho, y es completamente consciente de esta situación, hasta el punto de su ocultación. Desde otra perspectiva, referente a la incertidumbre del derecho Jerome Frank (quien fue juez de apelación en los Estados Unidos) también hace alusión a la ocultación de las razones que tienen los jueces para dictar sus sentencias señalando:

*“... otros consideran embustero a un testigo que echa su cabeza hacia atrás, o que se frota las manos, o que desvía la mirada rápidamente, o que ruboriza,... Si ese juez deja constancia en el expediente de que para decidir el caso ha usado una regla empírica tan absurda es probable que su fallo sea revocado. Pero es raro que un juez deje semejante constancia en el expediente...”.*

Por tanto, para Frank, la predicción de las sentencias envuelve en todo o en parte una “profecía de las creencias” (irrealizable) de los jueces de primera instancia o de los jurados, sobre los hechos [28].

Regresando al ensayo elaborado por Kennedy, él afirma que en su trabajo cotidiano el juez experimenta una sensación de libertad para crear derecho y una sensación de restricción causada por la obligación de aplicar los materiales jurídicos existentes. En este sentido el juez es completamente libre de decir lo que desee (confrontando la tesis de Dworking en la que el juez está atado por el derecho) y de hecho apunta que, la visión de la sentencia a la que el juez quiere llegar es previa a la búsqueda de los elementos jurídicos que servirán para sustentarla; sin embargo, el juez puede sentirse atado por el resultado que la comunidad jurídica o la opinión pública espera de su caso.

Kennedy también confronta la tesis de Hart, de los casos fáciles y difíciles, según Hart, en los primeros, al existir una regla precisa que indique cómo ha de resolverse una situación, el juez se encuentra completamente atado, y sólo en los casos difíciles, en los que no existe una regla que contemple los hechos, puede haber lugar a interpretaciones. En contraste, Kennedy sostiene que la facilidad o dificultad depende de la experiencia concreta del juez, y de hecho, un caso puede ser difícil aunque exista una regla clara que le sea aplicable, si el juez considera que el resultado al que conduce dicha regla es injusto. De igual forma, un caso puede ser fácil precisamente porque existe una laguna, o una ambigüedad que le permiten al juez emitir la sentencia que desea.

## 3.2. Elementos de lingüística jurídica

El derecho es un fenómeno de comunicación cuya natural expresión es el lenguaje. Visto el derecho de esta forma, podemos designar como ámbito jurídico al conjunto de procesos comunicativos que tienen como referente un ordenamiento jurídico, el cual en principio podemos afirmar que está conformado por un conjunto de textos (la constitución, los tratados, las leyes, las sentencias de los jueces) que se constituyen a partir de procesos de comunicación [43].

Un proceso de comunicación que ocupa un papel preponderante en el ámbito jurídico es el proceso judicial; éste, es eminentemente comunicativo, desde su inicio (presentación de la demanda o consignación ante juez) hasta su fin (notificación de la sentencia); por medio de comunicación procesal se entiende el vínculo, forma, o procedimiento, por el cual se transmiten ideas y conceptos dentro de la dinámica del proceso y para la consecución de sus fines [22].

En el caso de los ordenamientos jurídicos, estos procesos de comunicación se dan en el interior de los parlamentos y son el resultado de la capacidad que estos tienen para tomar decisiones que vinculen colectivamente [34]. Una vez publicados, los ordenamientos jurídicos suponen un mensaje en sí, que contiene nuevas reglas dirigidas a la sociedad en general (percibidas como nuevas reglas de conducta); a los órganos de gobierno (entendidas como nuevas reglas que deberán observarse al emitir actos de autoridad); al propio parlamento (nuevas reglas que deben considerarse en la elaboración de leyes futuras); y al poder judicial (vistas como nuevas reglas que deben aplicarse a casos concretos).

Nosotros deseamos referirnos a este último proceso de comunicación, en el que la ley es entendida como un mensaje, que para ser entendido, debe ser interpretado; respecto a este punto, Bobbio hace las siguientes consideraciones:

*“las tres fases de desarrollo de la jurisprudencia, siempre que por jurisprudencia se entienda justamente el análisis lingüístico que tiene por objeto a las proposiciones normativas de un determinado ordenamiento jurídico [son]. a) El lenguaje del legislador no es necesariamente riguroso: la primera tarea del jurista es hacerlo más riguroso; b) el lenguaje del legislador no es necesariamente completo: la segunda tarea del jurista es completarlo lo más posible; c) el lenguaje del legislador no está necesariamente ordenado: la tercera tarea del jurista es reducirlo a sistema” [7].*

Tradicionalmente se dice que interpretar es “desentrañar el sentido de la norma”. Desde nuestra perspectiva la interpretación de una norma, no puede realizarse de forma aislada pues al constituir una entidad lingüística, su

interpretación depende del contexto; así, identificamos tres tipos de contextos: contexto subjetivo, contexto objetivo y contexto normativo.

El contexto subjetivo va a depender de la persona que interpreta la norma y de su ideología. El contexto objetivo se encuentra dado por el caso concreto con respecto al cual se realiza la interpretación, aunque reconocemos que en un escenario judicial, los hechos que se juzgan pueden distar de lo que ocurrió en realidad. Finalmente el contexto normativo se conforma por todas aquellas reglas que el intérprete debe considerar al decidir si una norma es aplicable o no a un caso concreto.

Este contexto normativo lo entendemos en el sentido planteado por Joseph Raz al tratar el tema de la identidad y membresía de los sistemas jurídicos momentáneos, esto es, “Un conjunto de enunciados normativos es una descripción completa de un sistema jurídico momentáneo si, y sólo si cada uno de los enunciados de este conjunto describe el mismo sistema momentáneo (o parte de él) como todos los otros, y todo enunciado normativo que describe el mismo sistema jurídico momentáneo (o parte de él) es implicado por este conjunto” con lo cual, puede decirse que un sistema jurídico momentáneo es aquel que contiene todas las disposiciones jurídicas reconocidas por un órgano jurídico aplicador [41].

En el contexto normativo, nosotros distinguimos entre tres tipos de relación entre las normas: formal, semántica y pragmática.

### **3.2.1. Relaciones formales entre normas jurídicas**

Las relaciones formales entre las normas jurídicas nacen a partir de la estructura del ordenamiento jurídico. Para describirlas tomamos en cuenta la siguiente consideración: todo sistema jurídico prohíbe el uso de la fuerza en ciertas condiciones y, a su vez permite el uso de la fuerza en ciertas otras circunstancias; en todo sistema jurídico *todas* las disposiciones jurídicas guardan relaciones internas con las disposiciones jurídicas que prohíben el uso de la fuerza o con las disposiciones jurídicas que permiten el uso de la fuerza [41].

En este sentido, cada una de las normas se encuentra relacionada con todas las normas del sistema jurídico. Este tipo de relación se manifiesta con mayor claridad cuando una norma hace referencia a otra. Para mostrar esta situación citamos el artículo 130 del código fiscal de la federación.

**Artículo 130.-** *En el recurso de revocación se admitirá toda clase de pruebas, excepto la testimonial y la de confesión de las autoridades mediante absolucón de posiciones. No se considerará comprendida en esta prohibición la petición de informes a las autoridades fiscales, respecto de hechos que consten en sus expedientes o de documentos agregados a ellos.*

*Las pruebas supervenientes podrán presentarse siempre que no se haya dictado la resolucón del recurso.*

*Cuando el recurrente anuncie que exhibirá las pruebas **en los términos de lo previsto por el último párrafo del artículo 123 de este Código**, tendrá un plazo de quince días para presentarlas, contado a partir del día siguiente al de dicho anuncio.*

*La autoridad que conozca del recurso, para un mejor conocimiento de los hechos controvertidos, podrá acordar la exhibición de cualquier documento que tenga relación con los mismos, así como ordenar la práctica de cualquier diligencia.*

*Harán prueba plena la confesión expresa del recurrente, las presunciones legales que no admitan prueba en contrario, así como los hechos legalmente afirmados por autoridad en documentos públicos, incluyendo los digitales; pero, si en los documentos públicos citados se contienen declaraciones de verdad o manifestaciones de hechos de particulares, los documentos sólo prueban plenamente que, ante la autoridad que los expidió, se hicieron tales declaraciones o manifestaciones, pero no prueban la verdad de lo declarado o manifestado.*

*Cuando se trate de documentos digitales con firma electrónica distinta a una firma electrónica avanzada o sello digital, para su valoración, **se estará a lo dispuesto por el artículo 210-A del Código Federal de Procedimientos Civiles.***

*Las demás pruebas quedarán a la prudente apreciación de la autoridad.*

*Si por el enlace de las pruebas rendidas y de las presunciones formadas, las autoridades adquieren convicción distinta acerca de los hechos materia del recurso, podrán valorar las pruebas sin sujetarse a lo dispuesto en este artículo, debiendo en ese caso fundar razonadamente esta parte de su resolucón.*

*Para el trámite, desahogo y valoración de las pruebas ofrecidas y admitidas, serán aplicables las disposiciones legales que rijan para el juicio contencioso administrativo federal, a través del cual se puedan impugnar las resoluciones que pongan fin al recurso de revocación, en tanto no se opongan a **lo dispuesto en este Capítulo.***

Consideramos que la representación más cercana de este tipo de relaciones en un sistema legal, estaría dada por un *grafo completo*, lo cual es poco provechoso desde la perspectiva de la recuperación de información, pues si quisiéramos obtener las normas que tienen relación con un caso concreto, sólo encontraríamos que todas las normas tienen relación, lo mejor que podríamos hacer en este caso sería ponderar las aristas midiendo de alguna forma las relaciones que existen entre dos normas jurídicas.



### 3.2.2. Relaciones semánticas entre normas jurídicas

La semántica es la rama de la lingüística que se encarga de estudiar los principios que dan lugar a la creación y comprensión del significado en las lenguas naturales del mundo [12]. Decimos que dos normas jurídicas tienen una relación semántica, si en cuanto a su contenido, estas comparten cierto significado. Un ejemplo de esta relación lo podemos encontrar en el artículo 336 del código penal del distrito federal y en el artículo 112 bis de la ley de instituciones de crédito, citados a continuación.

**Artículo 336.** *Se impondrán de tres a nueve años de prisión y de cien a cinco mil días multa al que, sin consentimiento de quien esté facultado para ello:*

*I. Produzca, imprima, enajene, distribuya, altere o falsifique tarjetas, títulos o documentos utilizados para el pago de bienes y servicios o para disposición de efectivo;*

*II. Adquiera, utilice, posea o detente tarjetas, títulos o documentos para el pago de bienes y servicios, a sabiendas de que son alterados o falsificados;*

*III. Adquiera, utilice, posea o detente, tarjetas, títulos o documentos auténticos para el pago de bienes y servicios, sin consentimiento de quien esté facultado para ello;*

*IV. Altere los medios de identificación electrónica de tarjetas, títulos o documentos para el pago de bienes y servicios;*

*V. Acceda a los equipos electromagnéticos de las instituciones emisoras de tarjetas, títulos o documentos para el pago de bienes y servicios o para disposición de efectivo;*

*VI. Adquiera, utilice o posea equipos electromagnéticos o electrónicos para sustraer la información contenida en la cinta o banda magnética de tarjetas, títulos o documentos, para el pago de bienes o servicios o para disposición de efectivo, así como a quien posea o utilice la información sustraída, de esta forma; o*

*VII. A quien utilice indebidamente información confidencial o reservada de la institución o persona que legalmente esté facultada para emitir tarjetas, títulos o documentos utilizados para el pago de bienes y servicios, o de los titulares de dichos instrumentos o documentos.*

*VIII. Produzca, imprima, enajene, distribuya, altere, o falsifique vales utilizados para canjear bienes y servicios.*

*Si el sujeto activo es empleado o dependiente del ofendido, las penas se aumentarán en una mitad.*

**Artículo 112 Bis.** *Se sancionará con prisión de tres a nueve años y de treinta mil a trescientos mil días multa, al que sin causa legítima o sin consentimiento de quien esté facultado para ello, respecto de tarjetas de crédito, de débito, cheques, formatos o esqueletos de cheques o en general cualquier otro instrumento de pago, de los utilizados o emitidos por instituciones de crédito del país o del extranjero:*

*I. Produzca, fabrique, reproduzca, introduzca al país, imprima, enajene, aun gratuitamente, comercie o altere, cualquiera de los objetos a que se refiere el párrafo primero de este artículo;*

*II. Adquiera, posea, detente, utilice o distribuya cualquiera de los objetos a que se refiere el párrafo primero de este artículo;*

*III. Obtenga, comercialice o use la información sobre clientes, cuentas u operaciones de las instituciones de crédito emisoras de cualquiera de los objetos a que se refiere el párrafo primero de este artículo;*

*IV. Altere, copie o reproduzca la banda magnética o el medio de identificación electrónica, óptica o de cualquier otra tecnología, de cualquiera de los objetos a que se refiere el párrafo primero de este artículo;*

*V. Sustraiga, copie o reproduzca información contenida en alguno de los objetos a que se refiere el párrafo primero de este artículo, o*

*VI. Posea, adquiera, utilice o comercialice equipos o medios electrónicos, ópticos o de cualquier otra tecnología para sustraer, copiar o reproducir información contenida en alguno de los objetos a que se refiere el párrafo primero de este artículo, con el propósito de obtener recursos económicos, información confidencial o reservada.*

Este tipo de relación no es única, y en general podemos decir que existen niveles de relación semántica. Mediante la adecuada definición de este tipo de relaciones un conjunto de normas jurídicas pueden agruparse en aquellas que:

- tienen el mismo contenido
- regulan los mismos objetos
- están dirigidas a cierto grupo de personas
- regulan una materia específica
- otorgan derechos
- imponen obligaciones

Este tipo de relaciones son de gran utilidad ya que nos ayudan a visualizar el sistema legal de un modo más estructurado, lo cual nos facilita su comprensión y análisis, tanto en el ámbito jurídico como computacional; y, de hecho ya han sido utilizadas en algunos los trabajos mencionados en el capítulo 2, como por ejemplo en el clasificador de normas jurídicas del sistema NIR, o el trabajo realizado en el CIC-IPN.

### 3.2.3. Relaciones pragmáticas entre normas jurídicas

Mientras la semántica lidia directamente con los significados del lenguaje, la pragmática debe buscar una intencionalidad, un sentido. La función pragmática del lenguaje se hace presente en enunciados como “quiero comprar un libro, ¿traes dinero?” en tal caso quién emite la oración no tiene la intención de saber si la persona trae al menos un peso en la bolsa, menos aún quiere escuchar que la persona no trae un sólo peso, su intención es saber si lleva dinero suficiente para comprar el libro. Este tipo de situaciones también las podemos observar en el ámbito jurídico, tal es el caso de la palabra “explote” que se encuentra presente en la fracción primera del artículo 189 del código penal para el distrito federal, en donde tal palabra se refiere a “obtener un beneficio” y no a “causar una explosión” [12].

**Artículo 189.** *Se sancionará con prisión de dos a diez años y de quinientos a cinco mil días multa, al que:*

*1. Habitual u ocasionalmente **explote** el cuerpo de una persona u obtenga de ella un beneficio por medio del comercio sexual;*

Una relación pragmática entre dos normas jurídicas se presenta cuando: para regular un caso en concreto dos (o más) normas deben ser interpretadas en forma conjunta. Un ejemplo de relación pragmática entre normas jurídicas lo encontramos en los artículos 750 del código civil federal y 395 del código penal federal. Para explicar esta relación, pensemos primero en las siguientes situaciones: en la primera, una mujer es sorprendida *robando* los elotes de una parcela para llevar comida a sus hijos; en la segunda, dos hombres son sorprendidos *robando* un sistema de aire acondicionado que se encuentra unido a la azotea de un edificio.

Resaltamos la palabra robando ya que los frutos y los aparatos eléctricos son *bienes muebles*, que bien pueden ser robados. Sin embargo, en las situaciones descritas, los elotes de la parcela y el sistema de aire acondicionado unido al edificio, el delito que se actualiza es el despojo, en el cual, la conducta recae sobre bienes inmuebles, esto en razón de que jurídicamente, en ambos supuestos se trata de bienes inmuebles por incorporación.

Sin soslayar que tal situación (tratar un mueble como un inmueble) parece un completo absurdo, es perfectamente legal y de hecho no debemos olvidar que ambos son conceptos jurídicos definidos por la ley. Visto de este modo, el trabajo de la jurisprudencia es en su parte esencial un análisis del lenguaje, y, de modo más preciso, de ese específico lenguaje en el que a través de proposiciones normativas se expresa el legislador [7]. Esto tiene múltiples consecuencias: en primer lugar en cuanto a la pena, pues el despojo en el código penal federal se castiga con prisión de tres meses a cinco años de prisión y multa de cincuenta a quinientos pesos. Respecto al robo, la pena va a depender de la cuantía de lo robado, en el primer caso, podríamos decir que el valor de lo robado es menor a cien veces el salario; por ello la pena que podría imponerse sería de hasta dos años de prisión y multa de hasta cien veces el salario; en el segundo caso, podríamos suponer que el valor del sistema de aire acondicionado es menor a quinientas veces el salario, así, la pena aplicable sería de dos a cuatro años de prisión, y multa de cien hasta ciento ochenta veces el salario.

Otro aspecto importante es que para la primera situación que describimos no podría ser aplicable la excusa absolutoria del artículo 379 del código mencionado (*robo de famélico*). Bien podría aplicarse el estado de necesidad genérico previsto en la fracción V del artículo 15 del mismo ordenamiento, pero tal situación no es la que nos interesa en este momento, lo importante aquí es observar, que por una relación que existe entre dos normas, el estatus jurídico de la situación cambia drásticamente al pasar de un robo (dirigido a la protección de bienes muebles) a un despojo (enfocado a la protección de bienes inmuebles).

Citamos aquí los artículos 367 y 395 del código penal federal, 750 del código civil federal, así como la tesis jurisprudencial que trata este tema (aunque en dicha tesis las leyes son diferentes a las aquí mencionadas, el análisis es el mismo).

**Artículo 367.-** *Comete el delito de robo: el que se apodera de una cosa ajena mueble, sin derecho y sin consentimiento de la persona que puede disponer de ella con arreglo a la ley.*

**Artículo 395.-** *Se aplicará la pena de tres meses a cinco años de prisión y multa de cincuenta a quinientos pesos:*

*I.- Al que de propia autoridad y haciendo violencia o furtivamente, o empleando amenaza o engaño, ocupe un inmueble ajeno o haga uso de él, o de un derecho real que no le pertenezca;*

*II.- Al que de propia autoridad y haciendo uso de los medios indicados en la fracción anterior, ocupe un inmueble de su propiedad, en los casos en que la ley no lo permite por hallarse en poder de otra persona o ejerza actos de dominio que lesionen derechos legítimos del ocupante, y*

*III.- Al que en los términos de las fracciones anteriores, cometa despojo de aguas.*

*La pena será aplicable, aun cuando el derecho a la posesión de la cosa usurpada sea dudosa o esté en disputa. Cuando el despojo se realice por grupo o grupos, que en conjunto sean mayores de cinco personas, además de la pena señalada en este artículo, se aplicará a los autores intelectuales y a quienes dirijan la invasión, de uno a seis años de prisión.*

*A quienes se dediquen en forma reiterada a promover el despojo de inmuebles urbanos en el Distrito Federal, se les aplicará una sanción de dos a nueve años de prisión. Se considera que se dedican a promover el despojo de inmuebles urbanos en forma reiterada, quienes hayan sido anteriormente condenados por esta forma de participación en el despojo, o bien, se les hubiere decretado en más de dos ocasiones auto de formal prisión por este mismo delito, salvo cuando en el proceso correspondiente se hubiese resuelto el desvanecimiento de datos, el sobreseimiento ó la absolución del inculpado.*

**Artículo 750.-** *Son bienes inmuebles:*

*I. El suelo y las construcciones adheridas a él;*

*II. Las plantas y árboles, mientras estuvieren unidos a la tierra, y los frutos pendientes de los mismos árboles y plantas mientras no sean separados de ellos por cosechas o cortes regulares;*

*III. Todo lo que esté unido a un inmueble de una manera fija, de modo que no pueda separarse sin deterioro del mismo inmueble o del objeto a él adherido;*

*IV. Las estatuas, relieves, pinturas u otros objetos de ornamentación, colocados en edificios o heredados por el dueño del inmueble, en tal forma que revele el propósito de unirlos de un modo permanente al fundo;*

*V. Los palomares, colmenas, estanques de peces o criaderos análogos, cuando el propietario los conserve con el propósito de mantenerlos unidos a la finca y formando parte de ella de un modo permanente;*

*VI. Las máquinas, vasos, instrumentos o utensilios destinados por el propietario de la finca directa y exclusivamente, a la industria o explotación de la misma;*

*VII. Los abonos destinados al cultivo de una heredad, que estén en las tierras donde hayan de utilizarse, y las semillas necesarias para el cultivo de la finca;*

*VIII. Los aparatos eléctricos y accesorios adheridos al suelo o a los edificios por el dueño de éstos, salvo convenio en contrario;*

*IX. Los manantiales, estanques, aljibes y corrientes de agua, así como los acueductos y las cañerías de cualquiera especie que sirvan para conducir los líquidos o gases a una finca o para extraerlos de ella;*

*X. Los animales que formen el pie de cría en los predios rústicos destinados total o parcialmente al ramo*

*de ganadería; así como las bestias de trabajo indispensables en el cultivo de la finca, mientras están destinadas a ese objeto;*

*XI. Los diques y construcciones que, aun cuando sean flotantes, estén destinados por su objeto y condiciones a permanecer en un punto fijo de un río, lago o costa;*

*XII. Los derechos reales sobre inmuebles;*

*XIII. Las líneas telefónicas y telegráficas y las estaciones radiotelegráficas fijas.*

*Época: Novena Época*

*Registro: 190653*

*Instancia: Primera Sala*

*Tipo de Tesis: Jurisprudencia*

*Fuente: Semanario Judicial de la Federación y su Gaceta*

*Tomo XII, Diciembre de 2000*

*Materia(s): Penal*

*Tesis: 1a./J. 32/2000*

*Página: 138*

***DESPOJO. EL APODERAMIENTO DE FRUTOS DEL INMUEBLE OCUPADO ES CONSECUENCIA DE AQUEL DELITO, POR LO QUE NO SE CONFIGURA EL DELITO DE ROBO (LEGISLACIÓN DEL ESTADO DE VERACRUZ Y JALISCO).***

*El delito de despojo previsto en los artículos 191, fracción I, del Código Penal para el Estado de Veracruz y 262, fracción II del mismo ordenamiento para el Estado de Jalisco, esencialmente queda circunscrito al hecho de ocupar un inmueble ajeno o propio, cuando la ley no lo permita, de propia autoridad y empleando a tal efecto la violencia, la furtividad, la amenaza o el engaño o bien usar de él o de un derecho real que no le pertenezca al usurpador. Por su parte, el ocupar, a que se refieren los tipos penales, es tomar posesión de una cosa, invadir, irrumpir, entrar o introducirse en el inmueble, lo que significa que el sujeto activo del delito, asiente o reafirme sus plantas sobre el inmueble con el fin de ejercer sobre él un poder de hecho. Asimismo, esa ocupación debe recaer en bien inmueble que atendiendo a su naturaleza, por regla general tiene una situación fija, como pueden ser los terrenos y edificios, conjunto de materiales consolidados para permanecer en la superficie o en el interior del suelo, así como los árboles, plantas y los frutos pendientes de los mismos. Ahora bien, si el activo del delito del despojo dispone de los frutos que penden de los árboles, no hace sino perfeccionar su dominio, pues éste da al que lo tiene, la facultad de disponer, usar y disfrutar todo lo que existe en el inmueble es decir, de enseñorarse de él; de ahí que, cuando el activo procede a la recolección de frutos de las plantaciones, no puede afirmarse que cometa el delito de robo, ya que ese corte de frutos implica el ejercicio de una potestad o dominio que por virtud del despojo se tiene del bien inmueble.*

*Contradicción de tesis 5/99. Entre las sustentadas por el Tribunal Colegiado en Materia Penal del Séptimo Circuito y el Primer Tribunal Colegiado en Materia Penal del Tercer Circuito. 27 de septiembre de 2000. Cinco votos. Ponente: Humberto Román Palacios. Secretario: Urbano Martínez Hernández.*

*Tesis de jurisprudencia 32/2000. Aprobada por la Primera Sala de esta Suprema Corte de Justicia de la Nación, en sesión de veintidós de noviembre de dos mil, por unanimidad de cuatro votos de los señores Ministros: presidente José de Jesús Gudiño Pelayo, Humberto Román Palacios, Juan N. Silva Meza y Olga Sánchez Cordero de García Villegas. Ausente: Ministro Juventino V. Castro y Castro.*

### **3.3 Resoluciones judiciales**

Por resolución judicial se entiende toda decisión o providencia que adopta un juez o tribunal en el curso de una causa contenciosa o de un expediente de jurisdicción voluntaria, sea instancia de parte o de oficio [22]. En general las determinaciones judiciales se clasifican en autos, decretos y sentencias.

Los decretos son aquellas resoluciones del juez por medio de las cuales dicta las medidas encaminadas a la simple continuidad del proceso. Los autos son resoluciones judiciales que afectan no sólo la cuestión procesal, sino también a cuestiones sustanciales que surgen durante el proceso y que es necesario atender antes de emitir a la sentencia. Por sentencia entendemos la resolución que pone fin a un juicio o instancia. Si además de poner fin al juicio, en la sentencia se resuelve el fondo del asunto, es decir, se resuelve la controversia planteada mediante la aplicación de la ley al caso concreto, se dice que se trata de una sentencia en sentido material. Por el contrario, si la resolución que pone fin al proceso no entra en el fondo del asunto, ni dirime la controversia, estamos frente a una sentencia de tipo formal [22].

Sobre un mismo asunto puede haber más de una sentencia. La *sentencia de primera instancia*, es aquella dictada por el juez que originalmente conoce de la demanda o le es consignado un asunto en materia penal, esta instancia se desarrolla de forma dialéctica, en ella, las partes fijan sus posturas y ofrecen las pruebas que consideran pertinentes. Si alguna de las partes no se encuentra conforme con esta sentencia, puede impugnarla, en tal caso se emitirá una *sentencia de segunda instancia*, en la cual sólo se revisará si las consideraciones del juez cumplen con los requisitos mínimos de legalidad, constitucionalidad y respeto a los derechos humanos. En nuestro sistema jurídico, puede existir además, una *sentencia de amparo* (considerada también como una tercera instancia, aún y cuando no se trata de un recurso). Anteriormente la constitucionalidad y el respeto a los derechos fundamentales en las sentencias eran tratados exclusivamente por el juez de amparo, a partir de las reformas de 6 y 10 de junio de 2011, todo juez puede entrar al estudio de dichas cuestiones, aun así, si una persona considera que sus derechos no han sido respetados por una sentencia de segunda instancia, puede solicitar la protección de la justicia federal mediante un amparo directo.

### 3.3.1. Requisitos formales de las sentencias judiciales

Las sentencias deben contener los requisitos comunes a toda resolución judicial es decir:

- Órgano jurisdiccional que las dicta.
- Lugar y fecha donde se emiten.
- Fundamentos jurídicos.
- Firma del juez o magistrado que la emite.
- Autorización del secretario.

Además de estos requisitos las sentencias deben contener:

- Relación sucinta de las cuestiones planteadas y las pruebas rendidas.
- Consideraciones jurídicas aplicables, tanto legales como doctrinales (motivación).
- Resolución de los puntos sujetos a consideración del tribunal.
- En su caso, el plazo dentro del cual debe darse cumplimiento a la resolución.

### 3.3.2. Estructura de las sentencias judiciales

Una sentencia está constituida de la siguiente forma:

**Preámbulo.** Es la parte de la resolución en la que contiene todos los datos que identifican el asunto, esto es, número de expediente, órgano jurisdiccional que la emite, la identificación del tipo de proceso, los nombres de las partes, así como el lugar y fecha en el que se pronuncia la sentencia.

**Resultandos.** Consideraciones de tipo histórico. En ellos se relata todo lo acontecido en el asunto, refiriendo la posición de cada una de las partes, sus afirmaciones, los argumentos que han esgrimido y las pruebas que se han ofrecido.

**Considerandos.** Esta es la parte medular de la sentencia. Es aquí donde se plasman los argumentos y conclusiones del tribunal resultado de la confrontación entre las pretensiones y resistencias de las partes, a la luz de las pruebas que se hayan presentado y que tengan relación con la controversia planteada [22].



**Puntos resolutivos.** En ellos, se precisa de forma muy concreta, el sentido de la resolución.

### 3.3.3. Requisitos de fondo de las sentencias

Por requisitos sustanciales de la sentencia deben entenderse aquellos aspectos esenciales que toda sentencia debe poseer, dichos requisitos son: congruencia, motivación y exhaustividad [22].

**Congruencia de la sentencia.** La congruencia se refiere a una correspondencia entre lo aducido por las partes y lo resuelto por el tribunal, es decir, una sentencia no puede referirse a cosas que no han sido materia del litigio.

**Motivación de la sentencia.** Es la obligación del tribunal de expresar los fundamentos, es decir, los preceptos jurídicos que sustentan la decisión; así como los razonamientos, esto es, los motivos que llevan a la autoridad a aplicar los fundamentos legales al caso concreto, en su resolución.

**Exhaustividad de la sentencia.** Se refiere a la obligación del tribunal de tratar todas y cada una de las cuestiones planteadas por las partes, sin dejar de considerar ninguna. Esto es, agotar todos los puntos aducidos por las partes y hacer alusión a cada una de las pruebas rendidas.

## 3.4 Procesamiento del Lenguaje Natural

El procesamiento del lenguaje natural (NLP) surge a partir de la relación que existe entre la lingüística, que es la ciencia que estudia el lenguaje humano, y la computación. Para nosotros, cualquier tarea que pretenda desarrollarse en el área de la computación y que se encuentre relacionada con el lenguaje (análisis de sentimientos, generación automática de resúmenes, respuesta automática a preguntas, traducción automática, recuperación de información,...) requiere de la elaboración de un modelo lingüístico-computacional capaz de describir, explicar y predecir la tarea a realizar.

Un modelo científico es una construcción mental que representa algunas características o propiedades del objeto correspondiente [49] y debe de estar cimentado tanto en intuiciones como en matemáticas, las

intuiciones nos ayudan a acercarnos a modelos que sean aceptables para la comunidad científica, la parte matemática es imprescindible para formalizar el modelo y asegurar su consistencia. Por ejemplo, el modelo de nuestro sistema solar, predice con exactitud la posición de los planetas en una fecha determinada [26]. Regresando a nuestra área, la lingüística computacional estudia cómo construir los modelos del lenguaje, de tal manera que sean entendibles para las computadoras [48]. La función de la lingüística en esta tarea consiste en proporcionar las directrices que nos ayuden a elegir las características y los valores que éstas pueden tomar, para posteriormente utilizarlas en diferentes algoritmos.

Antes de describir formalmente nuestra tarea (recuperación de información IR) explicaremos en que consiste la tarea de extracción de información (IE), esto en razón de que si bien, para llevar a cabo este trabajo es necesario generar un modelo específico, la extracción de información permite trasladar la información no estructurada de los textos, a *información estructurada* o *semiestructurada* que puede ser manejada computacionalmente.

### 3.4.1. Extracción de Información

La información la podemos encontrar de muchas maneras, una muy relevante en nuestra área es la información estructurada [6] un ejemplo<sup>9</sup> de información estructurada lo encontramos en la tabla 2.

Compañía	Ubicación
Omnicom	Nueva York
DDB Needham	Nueva York
Kaplan Thaler Group	Nueva York
BBDO South	Atlanta
Georgia-Pacific	Atlanta

**Tabla 2. Ejemplo de información estructurada.**

Esta tabla representa la relación que existe entre un conjunto de compañías y su ubicación. La misma información la podríamos encontrar en la siguiente lista de python:

---

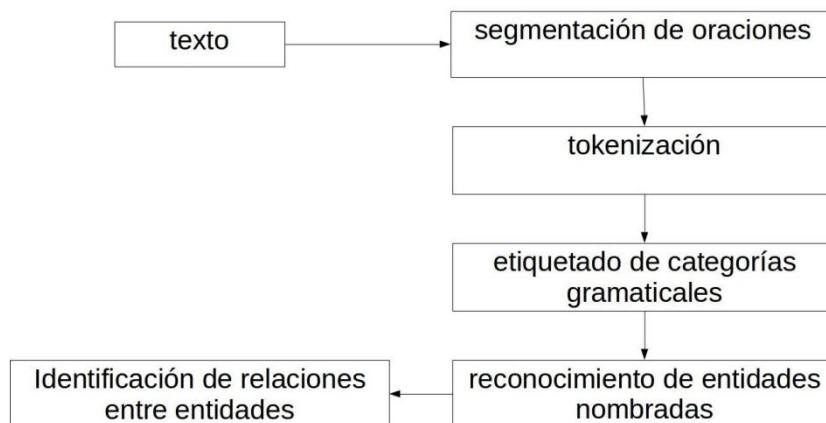
<sup>9</sup> El ejemplo es retomado de *Natural language processing with Python* [6].

*lista = [('Omnicom', 'in', 'Nueva York'), ('DDB Needham', 'in', 'Nueva York'), ('Kaplan Thaler Group', 'in', 'Nueva York'), ('BBDO South', 'in', 'Atlanta'), ('Georgia-Pacific', 'in', 'Alanta)]*

Cuando tenemos la información estructurada de esta forma, solicitar a la computadora las compañías ubicadas en Atlanta, resulta trivial. El problema es mucho más complejo cuando tratamos directamente con el texto, pues la misma información (*BBDO South* y *Georgia-Pacific*) la podríamos encontrar de la siguiente forma:

*“La división de Georgia-Pacific Corp. dedicada a productos de papel envasados es la cuarta cuenta en moverse de Wells hacia otra agencia, esta había llegado a Wells apenas el otoño pasado. Al igual que Hertz y the History Channel, han optado por mudarse hacia la agencia BBDO South filial de BBDO Worldwide propiedad de Omnicom. Ubicada en Atlanta BBDO South se encargaba hasta ahora del manejo de la publicidad de Georgia-Pacific, y asumirá ahora deberes adicionales, similares a los prestados actualmente a Angel Soft toilet tissue y Sparkle paper towels, dijo Ken Haldin vocero de Georgia-Pacific en Atlanta.”*

La *extracción de información* consiste en obtener información estructurada o semiestructurada de un conjunto de textos. Esta tarea ha cobrado mucha más relevancia en la actualidad debido a la cantidad de textos que se encuentran en internet, pues para la mayoría de las preguntas que se hacen, es probable que alguien haya escrito la respuesta en algún lugar; sin embargo, el problema consiste precisamente en ubicar la información que necesitamos y, para ello, es necesario en primer lugar una representación de los textos a partir de la cual podamos obtener cualquier tipo de información presente en un documento, es decir, es necesaria una representación de propósito general de la información textual. Desafortunadamente en este campo, el estado del arte aún se encuentra lejos de alcanzar una representación de este tipo; no obstante, podemos lograr avances significativos centrandos nuestros esfuerzos en detectar las relaciones que se presentan entre las entidades textuales [6]. El esquema general de un *sistema de extracción de información* lo encontramos en la figura 2.



**Figura 2. Esquema de un sistema de extracción de información.**

El etiquetado de categorías gramaticales (*POS tagging*) consiste justamente, en asignar a cada palabra la categoría gramatical que le corresponde, dependiendo el papel que desempeña en la oración [37]. Por ejemplo para la oración “el niño juega con la pelota” las categorías gramaticales que le corresponden a cada palabra serían el – determinante, niño – sustantivo, juega – verbo, con – preposición, la – determinante, pelota – sustantivo.

En el reconocimiento de entidades nombradas (NER), el objetivo es identificar dónde inicia y dónde termina una entidad lingüística en un texto, es decir, detectar en el texto las referencias a personas, instituciones, lugares, fechas, etc. Típicamente esto se logra reconociendo las frases nominales, los sustantivos propios y en algunos casos los sustantivos indefinidos [6]. Por ejemplo, en el siguiente texto:

*WASHINGTON, Estados Unidos. Oct. 27, 2015.- El presidente de EE.UU., Barack Obama, recibió en la Casa Blanca a la Selección Nacional de fútbol femenino, actual campeona mundial, y dio las gracias a las jugadoras por "inspirar" a todo el país.*

Las entidades nombradas son las siguientes:

- *Washington, Estados Unidos (lugar)*
- *Oct. 27, 2015. (fecha)*
- *El presidente de EE.UU., Barack Obama (persona)*
- *la casa blanca (lugar)*
- *la selección nacional de fútbol femenino (grupo de personas/organización)*
- *país (lugar)*

Una parte importante del NER consiste en determinar el tipo de entidad del que se trata, pues esto facilita la posterior identificación de relaciones que existen entre entidades. Otra parte importante, consiste en la detección de entidades dentro de entidades, por ejemplo, en el texto que se cita, es conveniente detectar que *Washington, Estados Unidos* es un lugar, pero resulta más conveniente aún el saber que *Washington* y *Estados Unidos* son dos entidades diferentes que también comparten una relación, al igual que en la frase *El presidente de EE.UU., Barack Obama* podríamos decir que *El presidente de EE.UU.* es una institución y *Barack Obama* es una persona, y que entre esta institución y esta persona existe una relación.

En cuanto a la identificación de relaciones entre entidades, podríamos decir que ésta consiste en el la búsqueda de patrones entre pares de entidades que ocurren en el texto [6]. Esto es de particular relevancia en la respuesta automática de preguntas pues por ejemplo, considerando el mismo texto, si deseáramos contestar a las preguntas ¿quién recibió a la selección nacional de fútbol femenino? o ¿en dónde recibieron a la

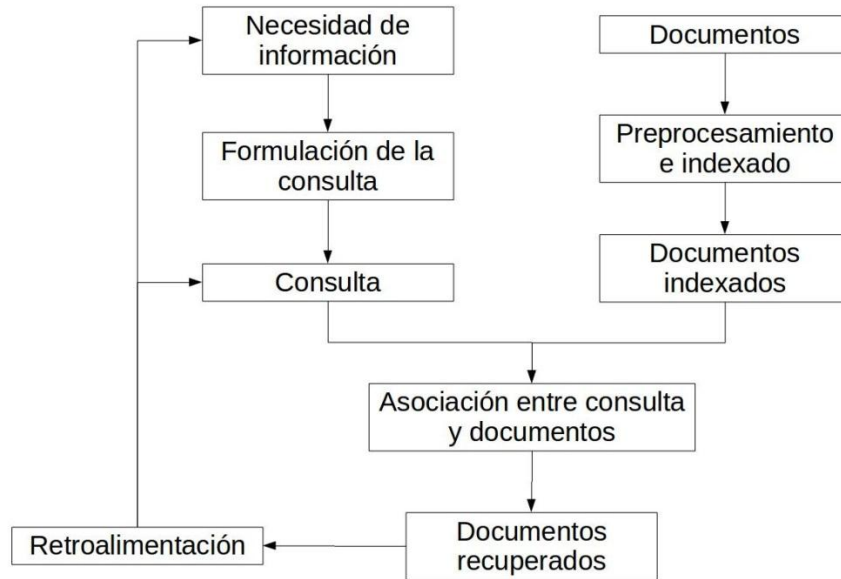
selección?, la búsqueda de las respuestas se hace posible si sabemos en primer lugar el tipo de entidades que participan en el texto, así como el tipo de relación que existe entre ellas; para la primera pregunta buscamos la relación *persona – recibe – persona/grupo de personas*; para la segunda *actividad – en – lugar*.

En el ámbito de nuestro trabajo únicamente consideramos necesario el reconocimiento de leyes, artículos, conceptos jurídicos y jurisprudencias. Propiamente sólo hicimos reconocimiento de leyes y artículos, pues la identificación de jurisprudencias, específicamente de rubros de jurisprudencias (232,851) así como de conceptos (4,874) se realizó a través de una lista. Esto tiene justificación ya que en las sentencias judiciales, los rubros de jurisprudencias se suelen citar de forma íntegra. En cuanto a los conceptos, lo más común es que sólo se altere el orden de las palabras que los componen; sin embargo, resulta ser un modelo muy limitado con el que no podemos proceder a buscar leyes y artículos, pues cualquier lista que se proponga no abarcará las nuevas leyes que se emitan, ni las distintas formas en las que se puede citar una ley o un artículo, los detalles sobre como realizamos el NER para leyes y artículos se encuentran descritos en el capítulo 4.

### **3.4.2. Recuperación de información**

Un sistema de recuperación de información, guarda y maneja documentos con la finalidad de asistir al usuario en la búsqueda de la información que necesita. El proceso mediante el cual se representa esta necesidad de información es la *consulta (query)*. La función esencial de este tipo de sistemas, consiste en indicar la existencia o inexistencia de la información, así como la localización de los documentos que podrían contener la información deseada. A estos documentos se les denomina *documentos relevantes*. El esquema general de un sistema de recuperación de información lo podemos visualizar en la figura 3 [26].

La etapa de pre-procesamiento e indexado de documentos consiste en obtener una representación de los documentos que nos permita posteriormente averiguar si estos contienen cierta información. En un sentido amplio, la etapa de consulta consiste en un diálogo interactivo entre sistema y usuario, que sea capaz de capturar la necesidad de información de éste, con el fin de que el sistema encuentre los documentos que satisfacen esta necesidad. Finalmente, la asociación entre la consulta y los documentos, usualmente resulta en una lista de documentos ordenados, conforma a la relevancia que estos tienen considerando la consulta que se realiza [26].



**Figura 3. Esquema de un sistema de recuperación de información.**

La manera específica en que estas etapas son llevadas a cabo depende del modelo de recuperación de información que se elija. En general podemos hablar de *modelos de coincidencia perfecta (exact match models)* entre la consulta y un documento o una parte de él, como son los modelos booleanos y los modelos de regiones; *modelos probabilísticos (probabilistic models)* y *modelos de espacio vectorial (vector space models, VSM)*. Existe una amplia literatura que describe con precisión las características de estos modelos, por lo cual sólo haremos referencia al VSM dado que fue el utilizado para realizar parte importante de nuestro trabajo.

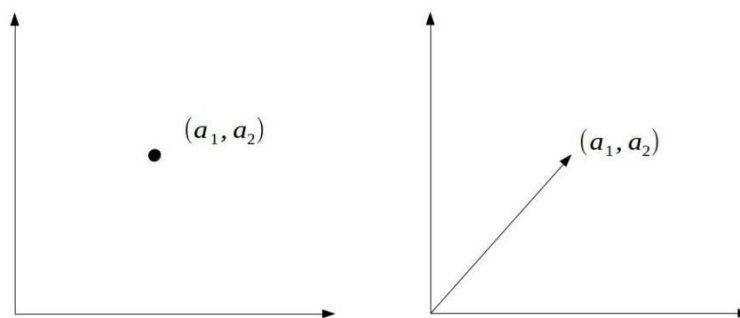
### 3.4.2.1. Modelo de espacio vectorial

El modelo de espacio vectorial (VSM) es un modelo ampliamente utilizado en el área de las ciencias de la computación para diversas tareas. Esto se debe en primer lugar a su simplicidad, la cual se conjuga con los resultados que a partir de su aplicación han sido obtenidos; es decir, es un modelo mediante el cual es posible resolver múltiples problemas dado que permite la comparación de objetos a través de métodos computacionales [48]. Comenzaremos por definir en qué consiste un espacio vectorial y posteriormente describiremos cómo es que éste se construye.

Matemáticamente se denomina *espacio n dimensional* y se denota  $R^n$  al conjunto de todas las *n-tuplas* ordenadas tales que, si  $n$  es un entero positivo, entonces un *n-teto* ordenado es una sucesión de  $n$  números reales  $(a_1, a_2, \dots, a_n)$ . Sea  $V$  un *espacio n dimensional*, si los objetos  $\mathbf{u}$ ,  $\mathbf{v}$ , y  $\mathbf{w}$  en  $V$  y los escalares (números)  $k$  y  $m$  satisfacen los siguientes axiomas, entonces  $V$  se denomina espacio vectorial y los objetos en  $V$  se denominan vectores [2].

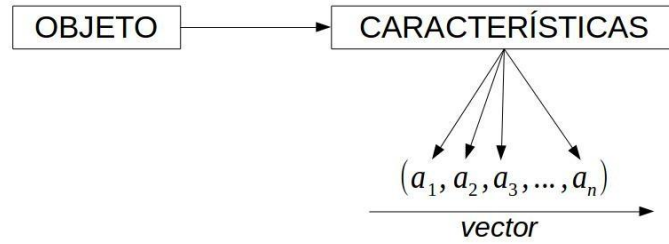
- 1.- Si  $\mathbf{u}$  y  $\mathbf{v}$  son objetos en  $V$ , entonces  $\mathbf{u} + \mathbf{v}$  está en  $V$ .
- 2.-  $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
- 3.-  $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$
- 4.- Existe un objeto  $\mathbf{0}$  en  $V$ , denominado vector cero tal que  $\mathbf{0} + \mathbf{u} = \mathbf{u} + \mathbf{0} = \mathbf{u}$
- 5.- Para todo  $\mathbf{u}$  en  $V$ , existe un objeto  $-\mathbf{u}$  en  $V$ , denominado negativo de  $\mathbf{u}$ , tal que  $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$
- 6.- Si  $k$  es cualquier escalar, y  $\mathbf{u}$  es cualquier objeto en  $V$ , entonces  $k\mathbf{u}$  está en  $V$
- 7.-  $k(\mathbf{u} + \mathbf{v}) = k\mathbf{u} + k\mathbf{v}$
- 8.-  $(k + m)\mathbf{u} = k\mathbf{u} + m\mathbf{u}$
- 9.-  $k(m\mathbf{u}) = (km)\mathbf{u}$
- 10.-  $1\mathbf{u} = \mathbf{u}$

Ahora bien, un objeto  $(a_1, a_2)$  tiene dos representaciones geométricas, como un “punto generalizado” o como un “vector generalizado”. En el primer caso hablamos de coordenadas; en el segundo de componentes o dimensiones. Para nuestro objeto es conveniente la representación de vectores.



**Figura 4. Formas de representar un vector.**

En la construcción de un VSM, lo que se pretende es describir los objetos a analizar a través de sus características, completando esta descripción asignando valores cada una de ellas. Este proceso se ilustra en la figura 5.



**Figura 5. Representación de características en un vector.**

La elección de las características, equivale a construir un modelo de nuestros objetos, en este sentido la selección de las características se realiza en forma subjetiva y depende del problema que se pretende abordar, sin embargo, la posterior comparación entre objetos es totalmente objetiva [48]. Para ilustrar este proceso, tomamos el siguiente ejemplo del libro *Construcción no lineal de n-gramas en la lingüística computacional* del Dr. Grigori Sidorov. Supongamos que nuestra intención es comparar dos libros. Las características a partir de las cuales podemos comparar los podríamos ser el número de páginas, el autor, el color de la portada, la editorial, etc. La selección de las características, equivale a la determinación de las dimensiones en un VSM, en donde cada característica equivale a una dimensión. Un aspecto importante en este modelo, es que todas las dimensiones deben tener valores numéricos, por lo que, el representar valores como el color de la portada, puede realizarse de diversas formas. Una de ellas es agregar a nuestro modelo los posibles colores de portada y colocar un 1 en el color que le corresponda y 0 en los demás posibles colores.

Como se puede apreciar, la construcción de un VSM es muy simple, y es uno de los métodos más utilizados en la comparación de textos. La idea más sencilla en este ámbito, es utilizar las palabras como características de los documentos. Los valores que estas características pueden tomar usualmente se determinan aplicando las medidas *tf* e *idf*, o una combinación de ambas *tf-idf*, esto, especialmente en el campo de la recuperación de información.

La medida *tf* (*term frequency*) se refiere al número de veces que una palabra se encuentra presente en un documento [48], aunque también puede referirse simplemente a si la palabra se encuentra presente o ausente del documento (0 ó 1 respectivamente). Es común que esta medida se encuentre normalizada de la siguiente forma:

$$tf(t, d) = \frac{f(t, d)}{\max\{f(w, d) : w \in d\}}$$



En donde  $f(t, d)$  es la frecuencia del término  $t$  en el documento  $d$ . De esta forma se pueden comparar de una mejor manera documentos grandes con documentos pequeños.

La medida *idf* (*inverse document frequency*) se basa en la siguiente intuición: si una palabra está presente en todos los documentos, entonces dicha palabra no es muy útil para distinguir entre dos textos [48]. El *idf* se calcula para cada palabra en una colección de la siguiente manera:

$$idf_i = \log\left(\frac{N}{DF_i}\right)$$

En donde  $N$  es el número total de documentos en nuestra colección,  $DF$  es el número de documentos donde la palabra  $i$  se encuentra al menos una vez.

### 3.5 Minería de texto

Definimos la minería de texto como el proceso de descubrimiento de patrones interesantes y posiblemente nuevos conocimientos en un conjunto de textos [16]. La minería de texto, toma su nombre a partir de la minería de datos, la cual se encarga del análisis de datos (estructurados) con el fin de descubrir patrones que, entre otras cosas, contribuyen en la toma de decisiones y mejora de estrategias en los negocios [24].

Estas disciplinas surgen como resultado de la evolución natural de las tecnologías de información, pues a partir del tremendo crecimiento de la información que se genera día con día (especialmente a través de internet) se ha producido una situación en la cual existen demasiados datos y poca información, excediendo incluso la capacidad humana para su comprensión sin las herramientas adecuadas. De esta forma, el propósito de estas disciplinas, es extraer (minar) el conocimiento presente en las grandes cantidades de datos que existen actualmente [24].

La extracción del conocimiento se realiza a partir de un proceso interactivo de las siguientes etapas [24]:

- Preparación, integración y selección de los datos.
- Transformación de los datos en formas apropiadas para extraer la información.
- Aplicación de los algoritmos para el descubrimiento de patrones no explícitos.
- Representación del conocimiento, que incluye la presentación de los datos al usuario.

La minería de texto es considerada una de las áreas más importantes de los sistemas de información e involucra, una integración entre múltiples disciplinas: bases de datos, estadística, aprendizaje automático, cómputo de alto rendimiento, redes neuronales, reconocimiento de patrones, recuperación de información, etc. Debido a esto, en esta sección nos limitamos a describir los métodos de clasificación y agrupamiento de textos que fueron utilizados para la realización del presente trabajo.

### 3.5.1. Clasificación de documentos

La clasificación es un método para la extracción de modelos que describen clases [27]. Dado un conjunto de datos, la tarea de clasificar los datos consiste en separarlos de modo tal que en la misma clase se encuentren objetos similares entre si y objetos de clases distintas sean lo menos similares posibles.

Amanera de ejemplo, si tomamos un conjunto de documentos, éstos pueden ser separados en clases considerando el tema, las palabras utilizadas, el autor del texto, el grupo de personas al cual se encuentra dirigido, la extensión del documento, etc. También es posible hacer una división tomando en cuenta conjuntos de características, verbigracia el tema y el autor; o, la extensión y el grupo de personas al que se encuentra dirigido etc.

Dado que la forma específica de realizar una clasificación va a depender completamente de nuestros objetivos, diversos métodos de clasificación han sido propuestos, principalmente provenientes del área de aprendizaje automático y reconocimiento de patrones [24].

Existen dos tipos de clasificación automática, a saber, clasificación supervisada (clasificación en sentido estricto) y clasificación no supervisada (agrupamiento). En el presente trabajo sólo se hace uso de clasificación no supervisada.

En la clasificación supervisada se cuenta con un conjunto de datos que previamente ya han sido clasificados (usualmente por expertos) es decir, cada dato cuenta con una etiqueta que indica la clase a la que pertenece. Este tipo de métodos consta de dos etapas: etapa de aprendizaje y etapa de clasificación. En la etapa de aprendizaje se analiza un subconjunto de los datos etiquetados (muestra de entrenamiento) con el objetivo de construir una función capaz de separar los datos en clases, esto es,  $f(objeto) = clase$ . En la etapa de clasificación, se procesan los datos que no pertenecen al conjunto de entrenamiento (muestra de supervisión o

muestra de control) y se *decide* a que clase pertenecen. Dado que previamente se sabe a qué clase corresponden los datos de la muestra de supervisión, es posible medir la precisión del algoritmo [24]. Algunos de los métodos más utilizados de clasificación supervisada son: clasificación por arboles de decisión, clasificadores bayesianos, clasificadores basados en reglas, redes neuronales y máquinas de soporte vectorial (SVM).

En la clasificación no supervisada (agrupamiento, *clustering*) se cuenta con un conjunto de datos en los que no se conoce la clase a la que pertenecen. Usualmente se utiliza para analizar grandes cantidades de datos pues realizar el etiquetado resultaría ser muy costoso en tiempo y en esfuerzo. El agrupamiento consiste en separar los datos en grupos utilizando las *divisiones propias de los datos* que existen entre ellos. Cada grupo constituye una colección de objetos que son similares entre sí, pero poco o nada similares a los objetos de otros grupos [24]. Existen diversos métodos para realizar agrupamiento, algunos de los más utilizados son: *agrupamiento particional, agrupamiento jerárquico y agrupamiento por densidad*.

La clasificación no supervisada tiene numerosas aplicaciones, por ejemplo, en el ámbito comercial puede ayudar a distinguir entre distintos grupos de consumidores y caracterizarlos utilizando sus patrones de compra. También puede ser empleada para descubrir desviaciones entre los patrones y detectar por ejemplo robos de tarjetas bancarias.

Dos de los principales usos del agrupamiento son el análisis exploratorio de datos (EDA) y la generalización. El EDA consiste en el entendimiento de las características básicas de un fenómeno para evitar planteamientos equivocados sobre los datos con los que trabajamos. Esto es especialmente importante en el análisis de textos pues a diferencia de otro tipo de datos, como pudieran ser los datos geográficos, no podemos tener una visualización de un conjunto de textos. Por lo que se refiere a la generalización, esta consiste en asumir que el tratamiento que se da a un miembro del grupo es correcto y, por tanto, podemos tratar a otros miembros del grupo de la misma manera, en este sentido agrupar es una forma de aprender a manejar objetos similares de forma similar [37]. En la generalización también podemos pasar de trabajar con datos individuales, a trabajar con datos agrupados como si constituyeran un sólo elemento, por lo cual, también podemos considerar el agrupamiento como una forma de reducir (comprimir) la información con la que trabajamos [24].

Como ya se mencionó, en la clasificación no supervisada, no se cuenta con un conjunto de datos etiquetados a partir de los cuales sea posible entrenar un algoritmo para que éste aprenda a clasificar los datos de acuerdo a lo que haya observado, por lo cual, la base del agrupamiento consiste en aprender a separar los datos, midiendo qué tan similares o diferentes son los datos entre sí.

### 3.5.2. Funciones de distancia y semejanza

La forma de medir que tan similares o diferentes son dos objetos, va a depender principalmente de tres factores: el aspecto del mundo con el que tratamos, el modelo que empleamos para entenderlo y el objetivo de la función, entendiéndose por éste, aquello que nos interesa medir entre dos objetos.

En un VSM de dos dimensiones podemos medir la distancia entre dos objetos de diversas formas; una de ellas se muestra en la figura 6.

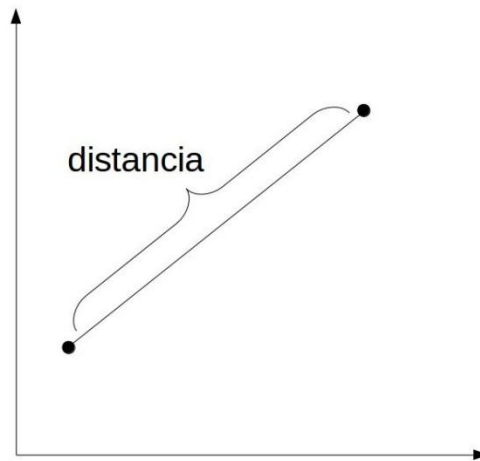
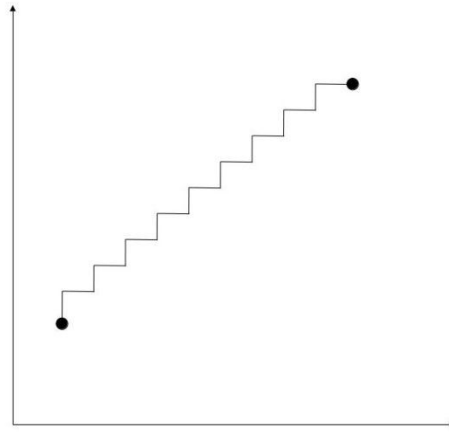


Figura 6. Distancia Euclidiana entre dos puntos.

La *distancia Euclidiana* representada en la figura 6 se encuentra dada por la siguiente fórmula:

$$d_E = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Donde  $x$  e  $y$  son vectores en un espacio  $n$  dimensional. Es conveniente aplicar esta distancia en espacios vectoriales continuos, sin embargo, en espacios vectoriales discretos, es preferible usar *distancia Manhattan*, o *distancia de Hamming* para vectores binarios.



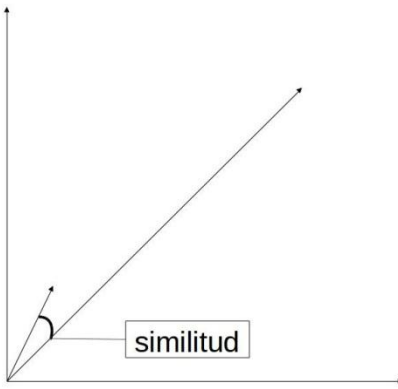
**Figura 7. Distancia Manhattan entre dos puntos.**

En la figura 7 mostramos la distancia entre dos puntos utilizando *distancia Manhattan*, la cual está definida de la siguiente forma:

$$d_M = \sum_{i=1}^n |x_i - y_i|$$

Otra forma de comparar dos objetos es verificar qué tan similares son, a diferencia de las funciones de distancia en las que el *codominio* se encuentra normalmente en los números reales (o naturales) positivos, en las funciones de semejanza el *codominio* usualmente está en el intervalo entre cero y uno  $[0, 1]$  siendo 1 la máxima semejanza que puede existir entre dos objetos, lo cual se presenta si los objetos son idénticos en nuestro modelo; y, 0 la mínima semejanza entre dos de ellos, idealmente cuando los objetos no comparten ninguna característica.

Una de las funciones más utilizadas para comparar textos en modelos de espacio vectorial, es la función de semejanza coseno, aunque el *codominio* de esta función se encuentra en el intervalo  $[-1, 1]$ ; al no ser factible medir qué tan ausente esta una palabra de un texto, lo más que podemos hacer es no considerarla, lo que equivale a colocar un cero en la dimensión que representa la palabra ausente, por lo cual, en la comparación entre textos el rango de la función de semejanza coseno suele estar en el intervalo  $[0, 1]$ ; podemos visualizar esta función en la figura 8.



**Figura 8. Representación de semejanza coseno entre dos vectores.**

La función de semejanza coseno se encuentra descrita de la siguiente forma, en donde  $v$  y  $u$  son vectores que se encuentran en el mismo espacio:

$$similitud_{\cos} = \frac{\mathbf{v} \cdot \mathbf{u}}{\|\mathbf{v}\| \times \|\mathbf{u}\|} = \frac{\sum_{i=1}^n (v_i \times u_i)}{\|\mathbf{v}\| \times \|\mathbf{u}\|}$$

## Estructura del trabajo propuesto

Como hemos mencionado, el objetivo de este trabajo es encontrar de manera automática las *relaciones pragmáticas* que existen entre las normas de un sistema legal, utilizando para ello un *corpus de decisiones judiciales*, y aprovechar estas relaciones para la recuperación de información en cuestiones jurídicas. Para ello hemos estructurado el trabajo en las siguientes etapas:

Etapa I. Recopilación de *sentencias judiciales*.

Etapa II. Procesamiento.

Etapa III. Desarrollo de métodos.

Métodos I y II.

Representación de los datos.

Recuperación de artículos relacionados a palabras.

Método III.

Representación de los datos.

Descubrimiento de relaciones entre artículos y palabras utilizando casos.

Recuperación de artículos relacionados a palabras.

Método IV.

Representación de los datos.

Descubrimiento de relaciones entre artículos utilizando similitud contextual.

Recuperación de artículos relacionados.

### 4.1 Recopilación de sentencias judiciales

Para este trabajo fueron utilizadas las sentencias públicas que se encuentran disponibles en el portal de la Dirección General de Estadística Judicial del Consejo de la Judicatura Federal del Poder Judicial de la Federación<sup>10</sup>, de conformidad con los artículos sexto y octavo de la Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental, el Reglamento de la Suprema Corte de Justicia de la Nación y del Consejo de la Judicatura Federal para la Aplicación de la Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental.

---

<sup>10</sup> <http://www.dgepj.cjf.gob.mx>

En total fueron descargadas 1,818,543 sentencias del periodo comprendido entre los años 2010 a 2014, cifra que representa el 39.23% de la totalidad de sentencias emitidas en dicho periodo (4,634,829) por Tribunales de Circuito y Juzgados de Distrito<sup>11</sup>.

Como fue explicado en el capítulo 3, la parte toral de las sentencias se encuentra en los *considerandos*. Por esta razón decidimos excluir de nuestro análisis las partes de la sentencia que corresponden al preámbulo, los resultandos y los puntos resolutivos. El *pseudocódigo* para obtener los considerandos de una sentencia es el siguiente:

*función obtener\_considerandos (sentencia):*

*lista\_considerandos = []*

*indice\_considerandos = 1*

*mientras no se encuentre inicio de considerando (indice\_considerandos):*

*avanzar en sentencia*

*mientras no se encuentre inicio de resolutivos:*

*variable\_considerando = []*

*indice\_considerandos += 1*

*mientras no se encuentre inicio de considerando (indice\_considerandos):*

*agregar palabras a variable\_considerando*

*avanzar en sentencia*

*agregar variable\_considerando a lista\_considerandos*

*regresar lista\_considerandos*

Cabe señalar que para una adecuada división de las sentencias es necesaria una variable que nos indique el considerando que buscamos, esto se debe a que en ocasiones un considerando puede contener una sentencia, aunque debemos reconocer que ésta es una práctica judicial terrible, pues sólo debería contener las partes relevantes de la sentencia. Esto ocurre cuando una sentencia dictada por *órgano de primera instancia* (Juzgados de Distrito) está siendo analizada por un *órgano de segunda o tercera instancia* (Tribunales de Circuito) a efecto de resolver si la sentencia cumple con los parámetros de validez legal, constitucional, e interpretación de los derechos fundamentales.

---

11 Datos tomados de los informes de labores que anualmente presenta la Suprema Corte de Justicia de la Nación, consultables en <https://www.scjn.gob.mx/Transparencia/Paginas/informe.aspx>



Al aplicar la función descrita se obtuvieron un total de 9,556,399 considerandos, esto es, un promedio de aproximadamente 5 considerandos por sentencia.

## 4.2. Procesamiento

Una vez que tuvimos los considerandos en archivos de texto separados, éstos fueron vinculados con los datos que identifican el tipo de asunto que se resolvió, a saber, materia del asunto y circuito en el que se atendió. Posteriormente los archivos fueron *tokenizados en* palabras y signos de puntuación; y, se eliminaron las *stop-words*, que son un conjunto de palabras, principalmente conjunciones, preposiciones y artículos, que tienen una alta frecuencia en la mayoría de los documentos y, que por tanto, se considera que no aportan información para distinguir entre diferentes documentos.

En esta etapa, también fueron estandarizados los números, esto es, los números ordinales y romanos, fueron convertidos a números decimales. De igual manera, los números ordinales y decimales que se encontraban descritos en palabras, fueron convertidos a números decimales; que por supuesto, siguieron conservando el tipo cadena (*string*).

Una vez hecho esto y considerando la cantidad de archivos así como la longitud de los mismos (un considerando puede abarcar varias páginas) se decidió aplicar un proceso de *lematización*<sup>12</sup>.

La *lematización* consiste en tomar la forma básica de una palabra en lugar de su forma flexionada. Para ello, mediante convención se han elegido ciertas reglas para elegir un representante de entre todas las formas flexionadas de una palabra. Por ejemplo, en el caso de los sustantivos, se elige como representante aquél que se encuentre en singular y masculino.

Después de haber *tokenizado* y *lematizado* los considerandos, se procedió al reconocimiento de jurisprudencias, leyes, conceptos jurídicos y artículos. Para una mejor explicación de la manera en que estas tareas fueron realizadas, decidimos abordarlas de forma separada.

---

12 Se utilizó el diccionario de lemas del sistema FreeLing. <http://nlp.lsi.upc.edu/freeling/>

## 4.2.1 Reconocimiento de jurisprudencias y leyes

El reconocimiento de las jurisprudencias se realizó localizando en el texto los rubros de las mismas. La lista de rubros se obtuvo de la página electrónica del Semanario Judicial de la Federación<sup>13</sup>, en total se obtuvieron 232,851 rubros, cada rubro con un promedio de 25 palabras.

Por otra parte, para la identificación de las leyes, utilizamos las leyes que se encuentran disponibles en el sitio electrónico de la Cámara de Diputados del Congreso de la Unión<sup>14</sup>, así como el catálogo de leyes estatales localizado en la página de internet del Instituto de Investigaciones Jurídicas de la UNAM<sup>15</sup> y la lista de tratados internacionales de los que México es parte, en los cuales se reconocen derechos humanos publicada en la dirección electrónica de la Suprema Corte de Justicia de la Nación<sup>16</sup>, obteniendo un total de 5,147 ordenamientos jurídicos. El título de cada ordenamiento jurídico tiene en promedio 8 palabras.

Tanto los rubros de las jurisprudencias, como los títulos de los ordenamientos jurídicos fueron considerados como *secuencias de palabras*. Para identificar estas secuencias de palabras (237,998) de manera eficiente, se construyó un *grafo dirigido* en el que cada nodo representa una secuencia de palabras y cada arista representa la relación “siguiente palabra” en una secuencia de palabras. El proceso de construcción del grafo (trie) se ilustra en la figura 9 para 5 secuencias de palabras.

---

13 <http://sjf.scjn.gob.mx/SJFSem/Paginas/SemanarioIndex.aspx>

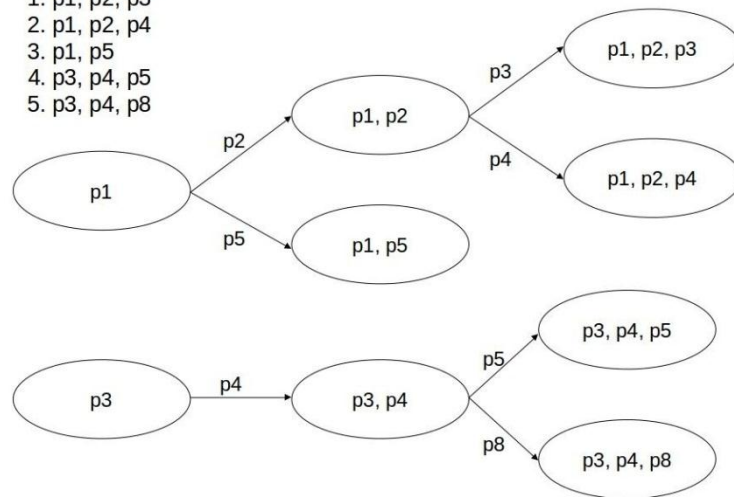
14 <http://www.diputados.gob.mx/LeyesBiblio/index.htm>

15 <http://www.juridicas.unam.mx/infjur/leg/legmexes.htm>

16 <http://www2.scjn.gob.mx/red/constitucion/TI.html>

Secuencias de palabras:

1. p1, p2, p3
2. p1, p2, p4
3. p1, p5
4. p3, p4, p5
5. p3, p4, p8



**Figura 9. Trie para la localización de secuencias de palabras.**

Mediante este método fue posible identificar todos los rubros de las jurisprudencias que fueron utilizadas en los considerandos, pues éstos se suelen citar de manera completa. Por otra parte, sólo fue posible reconocer las leyes que fueron invocadas utilizando su “denominación oficial” debido a que una misma ley puede ser nombrada de diversas maneras. Para ilustrar este hecho, ponemos algunos ejemplos de las formas en que es posible citar la Ley Federal del Trabajo.

Ordenamiento:

Ley federal del trabajo, reglamentaria del artículo 123 apartado A.

Formas de citar la ley:

Ley federal del trabajo.

Ley reglamentaria del artículo 123 apartado A.

Ley reglamentaria del artículo 123 constitucional, apartado A.

Ley reglamentaria del artículo 123, apartado A, de la constitución.

Ley reglamentaria del numeral 123, apartado A, de la constitución.

Ley reglamentaria del apartado A, del artículo 123, de la constitución.

Ley reglamentaria del apartado A, del artículo 123 constitucional.

Ley reglamentaria del apartado A, del artículo 123 del pacto federal.

Ley reglamentaria del apartado A, del artículo 123 constitucional de nuestro máximo ordenamiento.

...

Ahora bien, si consideramos que para referirnos a la constitución hay al menos 18 formas distintas, la palabra artículo puede ser reemplazada al menos con 3 términos diferentes; y, además, en algunos casos podemos variar la secuencia: artículo – apartado – constitución; en el ejemplo de la ley federal del trabajo, tendríamos más de 200 maneras de citar esta ley, sin tomar en cuenta que también podemos referirnos a ella de un modo más general, como por ejemplo “*ley de la materia*” o “*ley sustantiva*”.

Para abordar este asunto, utilizamos un *diccionario de sinónimos* que contiene únicamente las formas más comunes de referirse a las palabras: artículo, constitución y ley. Creamos también una lista con todos los *bigramas* presentes en los títulos de los 5, 147 ordenamientos legales que recabamos. Y, finalmente creamos una lista de lugares, con los estados y municipios del país; además, de la palabra “federal”. Para facilitar la explicación, decidimos denominar a estos recursos: *diccionario de sinónimos legales*, *bigramas legales* y *lista de lugares* respectivamente.

Los bigramas son todas las secuencias de dos palabras que componen una oración, por ejemplo, “tratado sobre el registro internacional de obras audiovisuales” está compuesto por los siguientes bigramas: “tratado sobre”, “sobre el”, “el registro”, “registro internacional”, “internacional de”, “de obras” y “obras audiovisuales”.

El *diccionario de sinónimos legales* nos permitió homologar las distintas formas en que se citan las palabras artículo, ley y constitución, que precisamente con esa denominación fueron etiquetados. Posteriormente identificamos todas las secuencias de palabras compuestas por *bigramas legales*, con una tolerancia de 2 bigramas, esto es, en la secuencia de palabras podían estar presentes hasta 2 bigramas que no estuvieran en nuestra lista. Por último, utilizamos la *lista de lugares*, para que estos fueran marcados en las secuencias de palabras.

Luego, para identificar a qué ley correspondían estas secuencias de palabras, creamos un modelo de espacio vectorial (VSM) en el que los vectores representan los títulos de las leyes, las dimensiones del espacio representan las palabras de estos títulos, y los valores que estas podían tener se obtuvieron aplicando la medida *tf-idf*.

Para cada secuencia de palabras, se calculó su similitud respecto a todas las leyes, utilizando la *función de semejanza coseno* descrita en el capítulo 3. Consideramos como leyes posiblemente citadas sólo aquellas que tenían una similitud mayor a 0.7.

Para decidir cuál de estas leyes había sido citada en una sucesión de palabras, o agregar una nueva ley que no estuviera incluida en nuestro catálogo, aplicamos las siguientes reglas, en el orden que se mencionan:

1. Si respecto a una ley, la secuencia de palabras sólo difiere en el lugar, entonces, consideramos la secuencia de palabras como una nueva ley.
2. Si la similitud de la ley con mayor semejanza a la secuencia de palabras, es mayor a 0.8, consideramos esta ley, como la ley citada.
3. Si el conjunto de palabras de una ley es un subconjunto de las palabras de la secuencia, consideramos esta ley como la ley citada.
4. Si el conjunto de palabras de la secuencia es un subconjunto de las palabras de una ley, consideramos esta ley como la ley citada.
5. Si la secuencia de palabras y todas las posibles leyes inician con la misma secuencia de términos, consideramos la secuencia de palabras como una nueva ley.

Considerando que en la aplicación de las reglas 2, 3 y 4 puede haber más de una ley que cumpla la condición, calculamos la probabilidad de que una ley  $l$ , aparezca en un considerando  $c$ , de la siguiente forma:

$$P(l) = \frac{|\{c \in C : l \in c\}|}{|C|}$$

En donde  $C$  es el conjunto de considerandos que analizamos. Así mismo se calculó la probabilidad de que una ley  $l$ , aparezca en un considerando  $c$ , dados la materia del asunto y el circuito en el que se resolvió:

$$P(l|I \cap M) = \frac{|\{c \in C : l \in c \wedge c \in I \cap M\}|}{|\{c \in C : c \in I \cap M\}|}$$

En donde  $C$  es el conjunto de considerandos analizados,  $I$  es el conjunto de considerados que fueron resueltos en el circuito  $i$  y  $M$  es el conjunto de considerados con la materia  $m$ .

De esta manera, en los casos en los que se aplicó alguna de las tres reglas mencionadas, y más de una ley cumplió con la condición, se seleccionó la ley con mayor probabilidad dada la materia del considerando y el circuito en el que fue resuelto. Para el caso en el que se volviera a presentar un empate, tomamos la ley con mayor probabilidad.

El programa mostró un desempeño eficiente para el reconocimiento de ordenamientos jurídicos que forman parte de nuestro catálogo, aunque esto no ocurrió para la detección de leyes que están fuera del mismo, pues el programa identificó 240 leyes, que de acuerdo con las reglas mencionadas, podrían ser agregadas a nuestra lista. De forma manual se advirtió la existencia de 128 errores, los cuales fueron corregidos, a guisa de ejemplo se muestran algunas de estas correcciones:

salida: ['ley', 'regular', 'juicio', 'amparo']

corrección: Ley de amparo, reglamentaría de los artículos 103 y 107 de la constitución.

salida: ['código', 'federal', 'represor']

corrección: Código penal federal

salida: ['ley', 'protección', 'ambiental', 'federal']

corrección: Ley general del equilibrio ecológico y protección ambiental.

salida: ['ley', 'general', 'trabajo']

corrección: Ley federal del trabajo

salida: ['código', 'procedimiento', 'materia', 'penal', 'yucatán']

corrección: Código de procedimientos penales del estado de Yucatán.

En esta línea, se agregaron sólo 112 leyes a nuestro catálogo, citamos aquí algunas de las leyes que fueron agregadas:

- Ley de hacienda municipal del estado de *Querétaro* (regla 1)
- Ley del servicio civil del estado de *Morelos* (regla 1)
- Ley de ejecución de penas y medidas de seguridad del estado de *Jalisco* (regla 1)
- Ley reglamentaria del servicio público de banca y crédito (regla 5, actualmente derogada)
- Tratado de libre comercio de América del Norte (regla 5)

También se consideró el hecho de que en nuestra colección de ordenamientos, estuviera presente una ley, no así su reglamento; pero que éste efectivamente existiera. Por lo cual, revisamos todas las leyes que fueron reconocidas (1,156) y en los casos en los que la palabra 'reglamento' se encontró exactamente una o dos posiciones antes de una ley detectada, procedimos de la siguiente manera: agregamos a nuestro conjunto de leyes el reglamento de la ley inicialmente identificada y, se cambió la etiqueta de la ley reconocida por la del reglamento añadido, integrando el término 'reglamento' a la secuencia de palabras, de esta forma fueron añadidos 23 reglamentos a nuestra lista de leyes.

Finalmente, cada una de las 1,179 leyes detectadas fue etiquetada como procesal o no procesal, y se procedió a resolver las siguientes referencias a leyes:

- **“Ley de la materia y ley sustantiva.** Para este tipo de citas, consideramos el ordenamiento con mayor probabilidad, dados el circuito en el que fue resuelto el considerando y la materia del mismo, siempre y cuando esta ley no fuera de carácter procesal.
- **“Ley adjetiva.** En cuanto a esta cita, consideramos el ordenamiento procesal con mayor probabilidad dados el circuito y la materia del considerando.

Otras referencias. Es común encontrar en las sentencias, referencias a leyes como las siguientes: *“la multicitada ley”, “la ley previamente referida”, “la ley anteriormente citada”*; este tipo de referencias no fueron consideradas en este trabajo, sin embargo, dada la cantidad de citas que se hacen de esta manera, resolvimos sólo las alusiones a leyes que se realizan inmediatamente después de haber mencionado la ley, como por ejemplo: *“ordenamiento legal en comento”, “la misma ley”, “la citada ley”*.

## **4.2.2. Reconocimiento de conceptos legales**

Para la identificación de *conceptos legales* utilizamos los rublos de las jurisprudencias, estos, tienen el siguiente formato:

***Tema.*** Argumento o resolución principal de la jurisprudencia.

Para efectos de este trabajo los temas de las jurisprudencias fueron considerados como *conceptos legales*. Por un lado, porque muchos de ellos tienen una definición legal o doctrinal, por otro, en razón de que los temas están compuestos a lo más de cinco palabras. A diferencia de los rubros, estos conceptos no pueden ser considerados como una secuencia de palabras, si no como un conjunto de palabras o bolsa de palabras, dado que el orden de las mismas no importa, ilustramos esta situación con algunos ejemplos:

- fianza y contrafianza
- querrela legítima
- falso testimonio
- legítima defensa
- testigo único

Aunque la búsqueda de bolsas de palabras, es más lenta que la búsqueda de secuencia de palabras, el reconocimiento de los conceptos resultó ser más rápido, por el hecho de que, la lista de conceptos recopilados consta únicamente de 4, 874 elementos, cantidad que es notablemente inferior en comparación con los rubros y leyes (casi 238 mil).

Cabe señalar que en primer lugar fueron identificados tanto los rubros como las leyes, y posteriormente los conceptos legales, debido a que algunos conceptos constan sólo de una palabra, y éstas palabras pueden estar presentes tanto en un rubro como en una ley; tal es el caso de las siguientes palabras: minería, municipio, ayuntamiento, sindicato, fideicomiso, quiebra, inmueble, etc.

### **4.2.3. Reconocimiento de artículos**

Para el reconocimiento de artículos se construyó un *etiquetador* de palabras basado en expresiones regulares (*Regular Expression Tagger*) con el cual se etiquetaron únicamente los siguientes tipos de palabras:

- Números
- Números latinos (bis, ter, quater...)
- Letras
- Símbolos
- Las palabras: artículo, apartado, fracción, inciso, párrafo y transitorio
- Los adjetivos ordinales: último penúltimo y antepenúltimo



Posteriormente, se realizó un análisis sintáctico superficial utilizando el *chunker* de la herramienta NLTK (Natural Language Toolkit versión 2.0b9). Las reglas con las que se realizó la primera parte del análisis son las siguientes:

- Estructura de párrafos citados:

La palabra “párrafo” seguida de al menos un número o un adjetivo ordinal

Al menos un número o un adjetivo ordinal seguidos de la palabra “párrafo”

- Estructura de incisos citados:

La palabra “inciso” seguida de al menos una letra.

- Estructura de fracciones:

La palabra “fracción” seguida de al menos un número o un adjetivo ordinal

- Estructura de apartados:

La palabra “apartado” seguida de al menos una letra

- Estructura de un identificador:

Un numeral latino

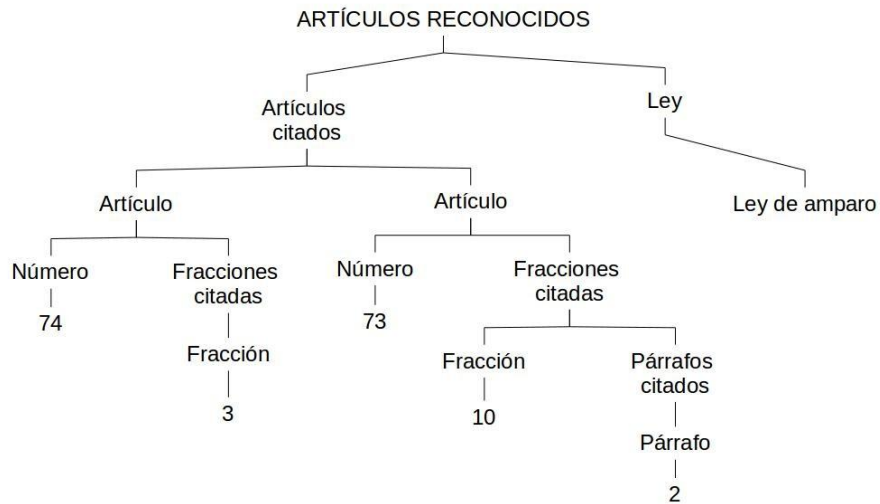
- Estructura un número:

Un número, que puede estar sucedido por un identificador

- Estructura de artículos:

La palabra “artículo”, seguida de al menos un número

La estructura completa de un precepto legal se compone al menos de una ley y un artículo, pudiendo existir además los siguientes elementos: identificador (numerales latinos), apartado, fracción, inciso y párrafo. A partir de este análisis obtuvimos árboles sintácticos como los siguientes:



**Figura 10. Artículos 74 fracción III y 73 fracción X párrafo segundo de la ley de amparo.**



**Figura 11. Artículo 42 bis de la ley federal de los trabajadores al servicio del estado.**

Estos árboles sintácticos fueron recorridos de manera que los artículos pudieran ser representados en el siguiente formato: ley – artículo – identificador – apartado – fracción – inciso – párrafo.

Los adjetivos ordinales: último penúltimo y antepenúltimo fueron representados con un -1, -2 y -3 respectivamente. En caso de que de alguno de estos elementos no estuviera presente se colocó un cero. La representación de los árboles sintácticos mostrados quedó de la siguiente manera:

- Ley federal de los trabajadores al servicio del estado, 42, bis, 0, 0, 0, 0.
- Ley de amparo, 74, 0, 0, 3, 0, 0.
- Ley de amparo, 74, 0, 0, 10, 0, 2.

Es importante señalar que elegimos esta representación, en razón de que un mismo artículo puede tener en cada una de sus partes un contenido diverso. Para ilustrar esta situación, citamos de manera íntegra el artículo 4 de nuestra constitución y colocamos entre paréntesis el tema de cada párrafo.

*Artículo 4o. El varón y la mujer son iguales ante la ley. Esta protegerá la organización y el desarrollo de la familia. (Igualdad entre el hombre y la mujer)*

*Toda persona tiene derecho a decidir de manera libre, responsable e informada sobre el número y el espaciamiento de sus hijos. (Libertad de procreación)*

*Toda persona tiene derecho a la alimentación nutritiva, suficiente y de calidad. El Estado lo garantizará. (Derecho a la alimentación)*

*Toda persona tiene derecho a la protección de la salud. La Ley definirá las bases y modalidades para el acceso a los servicios de salud y establecerá la concurrencia de la Federación y las entidades federativas en materia de salubridad general, conforme a lo que dispone la fracción XVI del artículo 73 de esta Constitución. (Derecho a la salud)*

*Toda persona tiene derecho a un medio ambiente sano para su desarrollo y bienestar. El Estado garantizará el respeto a este derecho. El daño y deterioro ambiental generará responsabilidad para quien lo provoque en términos de lo dispuesto por la ley. (Derecho al medio ambiente)*

*Toda persona tiene derecho al acceso, disposición y saneamiento de agua para consumo personal y doméstico en forma suficiente, salubre, aceptable y asequible. El Estado garantizará este derecho y la ley definirá las bases, apoyos y modalidades para el acceso y uso equitativo y sustentable de los recursos hídricos, estableciendo la participación de la Federación, las entidades federativas y los municipios, así como la participación de la ciudadanía para la consecución de dichos fines. (Derecho al agua)*

*Toda familia tiene derecho a disfrutar de vivienda digna y decorosa. La Ley establecerá los instrumentos y apoyos necesarios a fin de alcanzar tal objetivo. (Derecho a la vivienda)*

*Toda persona tiene derecho a la identidad y a ser registrado de manera inmediata a su nacimiento. El Estado garantizará el cumplimiento de estos derechos. La autoridad competente expedirá gratuitamente la primera copia certificada del acta de registro de nacimiento. (Derecho a la identidad)*

*En todas las decisiones y actuaciones del Estado se velará y cumplirá con el principio del interés superior de la niñez, garantizando de manera plena sus derechos. Los niños y las niñas tienen derecho a la satisfacción de sus necesidades de alimentación, salud, educación y sano esparcimiento para su desarrollo integral. Este principio deberá guiar el diseño, ejecución, seguimiento y evaluación de las políticas públicas dirigidas a la niñez. (Derechos de los niños)*

*Los ascendientes, tutores y custodios tienen la obligación de preservar y exigir el cumplimiento de estos derechos y principios.*

*El Estado otorgará facilidades a los particulares para que coadyuven al cumplimiento de los derechos de la niñez.*

*Toda persona tiene derecho al acceso a la cultura y al disfrute de los bienes y servicios que presta el Estado en la materia, así como el ejercicio de sus derechos culturales. El Estado promoverá los medios para la difusión y desarrollo de la cultura, atendiendo a la diversidad cultural en todas sus manifestaciones y expresiones con pleno respeto a la libertad creativa. La ley establecerá los mecanismos para el acceso y participación a cualquier manifestación cultural. (Derecho a la cultura)*

*Toda persona tiene derecho a la cultura física y a la práctica del deporte. Corresponde al Estado su promoción, fomento y estímulo conforme a las leyes en la materia. (Derecho a la cultura física y el deporte)*

Para finalizar esta sección sólo nos resta mencionar que fueron identificados un total de 70,101 artículos en 1,179 leyes, 104,451 jurisprudencias y 4, 352 conceptos jurídicos.

### **4.3. Métodos desarrollados**

Después de haber realizado el procesamiento de los datos, decidimos representar cada *considerando* como un conjunto de elementos que comprende la unión de: un conjunto de artículos, un conjunto de jurisprudencias, un conjunto de conceptos jurídicos, un conjunto de unigramas y un conjunto de bigramas. Posteriormente procedimos a la eliminación de los documentos con la misma representación quedando un total de 7,937,434 *considerandos* diferentes con los cuales realizamos experimentos a partir de los métodos que enseguida detallamos.

Para auxiliarnos en la explicación utilizaremos las siguientes convenciones:

- El conjunto de los considerandos utilizados, el conjunto de artículos recuperados, y el conjunto de jurisprudencias serán representados con las letras **C**, **A** y **J** respectivamente.
- El conjunto **T** (términos jurídicos) está conformado por la unión del conjunto de artículos y el conjunto de jurisprudencias identificados en la etapa del procesamiento de datos.
- El conjunto **E** (elementos relacionados) se integra por la unión del conjunto de unigramas, el conjunto de bigramas y el conjunto de conceptos jurídicos, que fueron observados en los considerandos.
- Una solicitud **S** es un conjunto de unigramas, bigramas y conceptos jurídicos, obtenidos a partir de una petición de información formulada como una secuencia de palabras (*tokenizada* y *lematizada* aplicado el procesamiento descrito en la sección 4.2).

### 4.3.1. Método 1

El objetivo de este experimento es recuperar los artículos relacionados a una petición de información, en donde esta petición pueda ser formulada en términos de una situación o hecho como por ejemplo: “qué pasa si una persona tiene dos credenciales de elector con diferente nombre”, o “cómo puedo solicitar mis ahorros del INFONAVIT”. Para ello decidimos utilizar un modelo de espacio vectorial (VSM) en el que los vectores representan *elementos relacionados*. Cada dimensión del espacio representa un *término jurídico*, y los valores que estas pueden tomar se calculan utilizando una adaptación de la medida *tf-idf* de la siguiente manera:

$$tf(t, e) = \frac{|\{c \in C : t \in c \wedge e \in c\}|}{\max_{x \in T} |\{c \in C : x \in c \wedge e \in c\}|} \quad idf(t, E) = \log \frac{|E|}{|\{e \in E : e \in c \wedge t \in c\}|}$$

En palabras simples, la frecuencia de términos se definió por el número de considerandos en los que se encontraron presentes tanto un elemento relacionado *e* como un término jurídico *t*; y la frecuencia inversa se determinó a partir del número de elementos que pertenecen a **E** con los que un término jurídico *t* coincidió en algún *considerando*.

Para la recuperación de información se procede de la siguiente forma: dada una solicitud **S**, obtenemos un conjunto **R** que corresponde a la intersección entre los conjuntos **S** y **E**; a partir de este conjunto, calculamos un *vector respuesta* sumando los vectores del conjunto **R**.

$$respuesta = \sum_{\vec{v} \in R} \vec{v}$$

Este *vector respuesta* puede ser visto como una lista de tupas. Cada una de ellas compuesta por un término legal, así como por el resultado de la sumatoria en la dimensión que le corresponde al término, esto es: (resultado, término legal).

Para facilitar la interpretación de los resultados, agrupamos los artículos de acuerdo a la ley a la que pertenecen; las jurisprudencias al no pertenecer a ninguna ley formaron un grupo independiente. Para presentar la respuesta, estos grupos son ordenados de mayor a menor, de acuerdo a su *peso*; este *peso* se obtiene sumando el *resultado* que de manera independiente obtuvo cada *término legal*. A su vez, los elementos de cada grupo (artículos o jurisprudencias) son ordenados de mayor a menor de acuerdo al resultado que obtuvieron.

Como respuesta a la solicitud obtenemos las primeras *n* leyes y los primeros *m* elementos de cada ley.

### 4.3.2. Método 2

El propósito de este experimento es el mismo que el mencionado al inicio del experimento 1; sin embargo, en este experimento procedimos a calcular la probabilidad de todos los *términos jurídicos* dada una solicitud **S** de la siguiente forma:

$$P(t|S) = \prod_{s \in S} \begin{cases} \text{si } s \notin E & \rightarrow \frac{1}{|C|+1} \\ \text{si } P(t|s) > 0 & \rightarrow \frac{|\{c \in C : t \in c \wedge s \in c\}|}{|\{c \in C : s \in c\}|} \\ \text{otro caso} & \rightarrow \frac{1}{|\{c \in C : s \in c\}|+1} \end{cases}$$

Una vez calculada la probabilidad de todos los términos, agrupamos los artículos de acuerdo a la ley a la que pertenecen; las jurisprudencias al no pertenecer a ninguna ley formaron un grupo independiente. Para presentar la respuesta, estos grupos son ordenados de mayor a menor de acuerdo a su *peso*. Éste *peso* se obtiene sumando la probabilidad que de manera independiente obtuvo cada *término legal*. A su vez, los elementos de cada grupo (artículos y jurisprudencias) son ordenados de mayor a menor de acuerdo a su probabilidad.

Como respuesta a la solicitud obtenemos las primeras  $n$  leyes y los primeros  $m$  elementos de cada ley.

### 4.3.3. Método 3

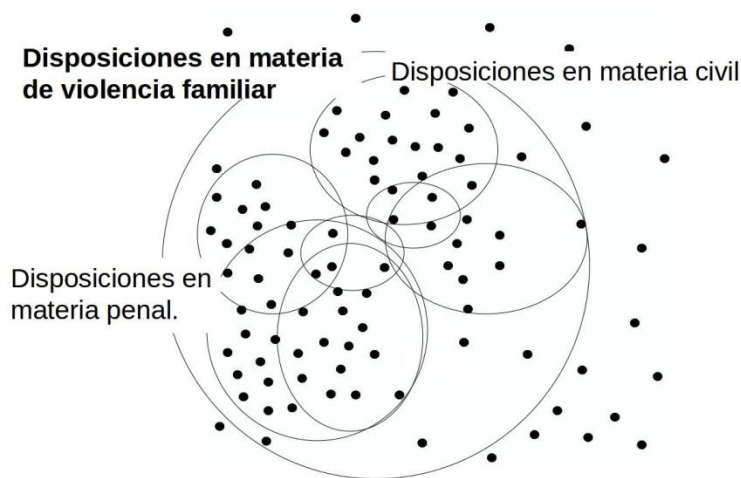
Los objetivos de este experimento consistieron por un lado, en agrupar los considerandos; por otro, utilizar este agrupamiento para realizar consultas de información en los términos descritos en el experimento 1; es decir, a partir del planteamiento de un hecho o una situación. Para alcanzar este fin, en primer lugar utilizamos una representación diferente de los *considerandos*, para efectos de este experimento, un considerando está definido por un conjunto de artículos y un conjunto de jurisprudencias.

En primer lugar, formamos un VSM en cual, los vectores representan a los *considerandos*; las dimensiones del espacio se encuentran definidas por los conjuntos **A** y **J** y los valores que éstas pueden tomar se calculan aplicando la medida *idf*.

Para calcular la similitud entre patrones utilizamos la función de semejanza coseno descrita en capítulo 3. El agrupamiento de los considerandos se realizó utilizando el algoritmo **DBSCAN** [24] las razones por las que se determinó utilizar este algoritmo son las siguientes:

Al resolver un asunto, los jueces deben fundamentar su sentencia en normas jurídicas. Estas normas tienen una mayor o menor relación con el asunto, y pueden tratar sobre aspectos jurídicos muy distintos; por ejemplo, en un caso de violencia familiar, el asunto puede ser abordado desde la materia penal (artículos 200 a 202 del código penal para el distrito federal) así como desde la vía civil (artículos 323 quater, 323 quintus y 323 sextus del código civil para el distrito federal). En el primer caso, esta conducta es sancionada con prisión así como con la pérdida de derechos respecto a la víctima; el segundo caso se enfoca al pago de daños y perjuicios.

Ahora bien, para dictar su sentencia un juez hará uso de todos los artículos que considere convenientes y estas disposiciones variarán de acuerdo al criterio del juez y las necesidades específicas del asunto, pero todos los casos resueltos con base en estos artículos son asuntos de violencia familiar, verbigracia, para resolver un asunto de este tipo, un juez penal podría aplicar el artículo 200 del código penal referido si la víctima fuera un familiar del agresor; de existir otro tipo de relación entre el agresor y la víctima (como lo es la custodia) el artículo aplicable sería el 201 bis del mismo código. Habrá además otras disposiciones que no siempre serán empleadas, por ejemplo, de carácter procesal, tal circunstancia ocurriría si se presenta una apelación en contra del auto de formal prisión o de vinculación a proceso; sin embargo, esta situación no se presenta en todos los asuntos. Ilustramos este hecho con el siguiente diagrama.



**Figura 12. Agrupamiento de términos legales.**

Considerando lo anterior, encontraremos artículos que estarán presentes en un mayor número de asuntos de este tipo y encontraremos otros que pueden casi no ser utilizados para el mismo tipo de casos. El emplear este método (agrupamiento por densidad) nos permite agrupar los considerandos, tomando en cuenta esta situación.

Otra razón que también influyó para aplicar este agrupamiento es la complejidad en tiempo del mismo  $O(n \log n)$  pues muchos de los algoritmos de agrupamiento tienen al menos una complejidad cuadrática.

Una vez que se tienen agrupados los *considerandos*, formamos un segundo VSM esta vez los vectores representan grupos de *considerandos*, las dimensiones del espacio están definidas por el conjunto  $E$  y los valores que estas pueden tomar se calculan utilizando la medida *tf-idf*, tomando cada grupo como si fuera un texto.



La recuperación de información se realiza a partir de una solicitud **S**; agregamos esta solicitud a nuestro VSM y calculamos su similitud con respecto a cada grupo utilizando la *función de semejanza coseno* descrita en el capítulo anterior. Una vez hecho esto ordenamos los grupos de acuerdo a la similitud que hayan tenido con **S** (de mayor a menor). La respuesta de nuestro programa está constituida por los primeros *n* grupos de esta lista, y los *m* artículos más frecuentes en cada grupo.

#### 4.3.4. Método 4

En este experimento nos enfocamos a descubrir las relaciones que existen entre los *términos jurídicos*, el propósito específico del experimento es el siguiente: sea **L** un subconjunto de **T**, nuestra tarea es encontrar los términos más relacionados a **L**, tales que, dichos términos se encuentren en el conjunto **T** y en el complemento de **L**.

Para cumplir con este objetivo, consideramos **L** como un grupo (*cluster*). De igual forma, los elementos del complemento de **L** fueron considerados como grupos conformados por un elemento. Para calcular la similitud entre términos nos basamos en el coeficiente de correlación de Pearson; para dos términos jurídicos *x* e *y*, éste se encuentra definido por la siguiente fórmula:

$$\text{correlación}(x, y) = \frac{\text{covarianza}(x, y)}{\text{desviación estándar}(x) * \text{desviación estándar}(y)}$$

Dado que cada *considerando* es visto como un conjunto, la covarianza y la desviación estándar se encuentran dadas por:

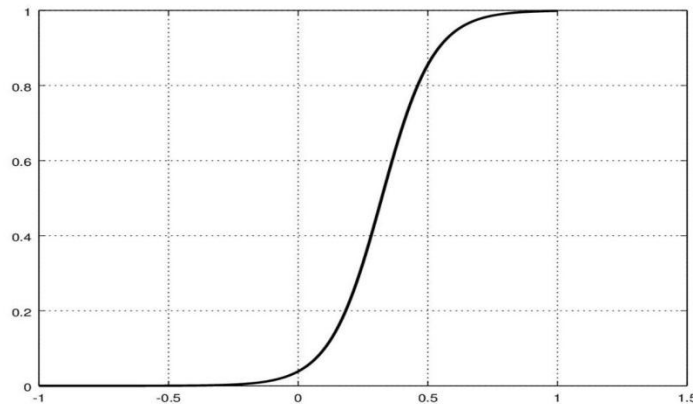
$$\text{covarianza}(x, y) = \frac{|\{c \in C : x \in c \wedge y \in c\}|}{|C|} - \left( \frac{|\{c \in C : x \in c\}|}{|C|} * \frac{|\{c \in C : y \in c\}|}{|C|} \right)$$

$$\text{desviación estándar}(x) = \sqrt{\frac{||c \in C : x \in c||}{|C|} - \left(\frac{||c \in C : x \in c||}{|C|}\right)^2}$$

El índice de correlación mide la relación lineal entre dos variables y se encuentra en el rango de -1 a 1; por tanto, para utilizarlo como una medida de semejanza fue necesario normalizarlo en el rango de 0 a 1. Para ello utilizamos la *función logística*:

$$f(x) = \frac{1}{1 + \beta(e^{\frac{x}{\alpha}})}$$

En este caso, se utilizó un  $\beta = 25$  y  $\alpha = 0.1$  para que la función iniciara su crecimiento cuando el valor de la correlación entre dos términos fuera mayor a cero.



**Figura 13. Crecimiento de la función sigmoide con valores de -1 a 1.**

Para medir la similitud entre grupos utilizamos la función de enlace promedio (*link average*):

$$\text{enlace\_promedio}(G1, G2) = \frac{\sum_{a \in G1} \sum_{b \in G2} \text{similitud}(a, b)}{|G1| \times |G2|}$$

Con esta función se calculó la similitud del grupo **L** con respecto a todos los grupos que se encuentran en el complemento de **L**, obteniendo una lista ordenada de mayor a menor de acuerdo a la similitud determinada por la función de enlace. De esta lista consideramos los primeros *n* elementos como los elementos más relacionados a **L**.

## Resultados y análisis

En este capítulo presentamos los resultados obtenidos en los experimentos realizados aplicando los métodos descritos en el capítulo anterior.

### 5.1. Descripción del conjunto de datos para evaluación

El conjunto para evaluación está compuesto por 16,492 sentencias judiciales (no incluidas en nuestro corpus de entrenamiento) todas ellas emitidas por Juzgados de Distrito en materia civil del Primer Circuito, en los años 2014 y 2015. Cabe mencionar que únicamente tomamos sentencias en materia civil, dado que en este periodo han entrado en vigor legislaciones de suma trascendencia como lo son la Nueva Ley de Amparo y Código Nacional de Procedimientos Penales, resultando la materia civil una de las menos afectadas por estos cambios, al menos en lo que se refiere a su aspecto sustantivo.

De estas sentencias se obtuvo un total de 71,681 considerandos, esto es, un promedio de 4.3 considerandos por sentencia. A estos considerandos les fue aplicado el procesamiento descrito en la sección 4.2 con lo cual, cada considerando es representado por un conjunto de elementos que comprende la unión de: un conjunto de artículos, un conjunto de jurisprudencias, un conjunto de conceptos jurídicos, un conjunto de unigramas y un conjunto de bigramas.

Los considerandos con exactamente los mismos elementos fueron eliminados; por otra parte, los considerandos que no incluían al menos un artículo o jurisprudencia no fueron tomados en cuenta, quedando así un total de 39,189 considerandos; de éstos, 3,808 contenían artículos que no están presentes en nuestro corpus debido a las reformas mencionadas (principalmente debido a la entrada en vigor de la nueva ley de amparo). Dado que detectar cambios en la legislación está más allá de los alcances de este trabajo, omitimos considerar para la evaluación los artículos no presentes en nuestro corpus de entrenamiento.

En la tabla 3 mostramos el promedio y la desviación estándar tanto de leyes como de artículos presentes en un *considerando*, en relación con el número de unigramas, bigramas y conceptos jurídicos (**S**).

Cardinalidad de S	Elementos evaluados	Promedio leyes	Desviación leyes	Promedio artículos	Desviación artículos
<=10	122	1.089	0.313	1.691	2.785
<=20	621	1.343	0.673	2.685	3.312
<=30	1154	1.576	0.874	2.756	2.873
<=40	1864	1.744	0.931	3.187	3.776
<=50	2701	1.89	0.994	3.397	3.369
<=60	2376	1.859	1.018	3.255	3.087
<=70	2538	2.292	1.232	3.858	3.066
<=80	1701	2.246	1.133	3.797	2.572
<=90	1277	2.056	0.949	3.722	2.869
<=100	1094	2.057	0.926	3.825	3.446
<=200	7773	2.407	1.019	4.897	3.582
<=300	3758	2.823	1.177	7.062	4.539
<=400	2492	3.458	1.218	9.931	5.687
<=500	1808	3.959	1.307	13.334	7.321
<=600	1654	4.505	1.303	16.268	7.532
<=700	1476	4.902	1.328	19.289	8.451
<=800	1227	5.3	1.49	22.492	9.228
<=900	958	5.538	1.465	24.982	10.337
<=1000	705	5.888	1.446	28.116	10.585
> 1000	1890	6.969	1.834	38.8	16.03

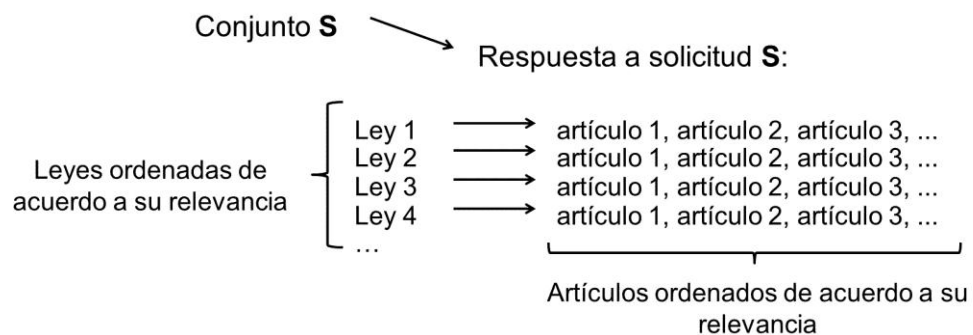
Tabla 3. Promedio de leyes y artículos en relación con el número de elementos que integran un considerando.

## 5.2. Resultados experimentales aplicando el método 1

Para analizar los resultados obtenidos a partir del método 1, cada considerando del conjunto de datos para evaluación fue dividido de la siguiente forma:

- Un conjunto **S** compuesto por los unigramas, bigramas y conceptos jurídicos del considerando.
- Un conjunto **T** compuesto por los artículos y jurisprudencias del considerando.

Nuestro objetivo fue utilizar los conjuntos **S** (solicitud de información) para encontrar los conjuntos **T** (términos jurídicos) para lo cual, aplicando el método descrito en la sección 4.3.1 a partir de cada conjunto **S** se obtuvo una lista de leyes ordenadas y, por cada ley se obtuvo una lista de artículos ordenados; la figura 14 muestra la forma en que se ordenan los resultados.



**Figura 14. Formato en el que son ordenados los resultados.**

Dado que conocemos la respuesta a cada conjunto **S**, la evaluación de los resultados se realiza utilizando las medidas de evaluación *precision* y *recall*. Mostramos en primer lugar *precision* y *recall* en la búsqueda de leyes relevantes considerando únicamente las primeras  $n$  leyes de la lista de leyes recuperadas.

Cardinalidad de S	Elementos evaluados	<i>precision</i>	<i>recall</i>
<=10	122	0.32	0.605
<=20	621	0.18	0.315
<=30	1154	0.167	0.254
<=40	1864	0.161	0.206
<=50	2701	0.176	0.228
<=60	2376	0.205	0.253
<=70	2538	0.235	0.25
<=80	1701	0.308	0.303
<=90	1277	0.361	0.353
<=100	1094	0.42	0.405
<=200	7773	0.603	0.498
<=300	3758	0.724	0.534
<=400	2492	0.812	0.498
<=500	1808	0.863	0.461
<=600	1654	0.908	0.434
<=700	1476	0.93	0.409
<=800	1227	0.956	0.389
<=900	958	0.962	0.371
<=1000	705	0.984	0.359
> 1000	1890	0.99	0.308

Tabla 4. *precision* y *recall* obtenidos en la recuperación de leyes con n=2 (método 1).

Cardinalidad de S	Elementos evaluados	<i>precision</i>	<i>Recall</i>
<=10	122	0.208	0.955
<=20	621	0.212	0.839
<=30	1154	0.244	0.827
<=40	1864	0.247	0.723
<=50	2701	0.266	0.72
<=60	2376	0.261	0.712
<=70	2538	0.342	0.745
<=80	1701	0.335	0.749
<=90	1277	0.296	0.717
<=100	1094	0.305	0.727
<=200	7773	0.381	0.778
<=300	3758	0.442	0.787
<=400	2492	0.5	0.74
<=500	1808	0.552	0.712
<=600	1654	0.606	0.701
<=700	1476	0.63	0.674
<=800	1227	0.65	0.644
<=900	958	0.658	0.621
<=1000	705	0.68	0.606
> 1000	1890	0.7	0.535

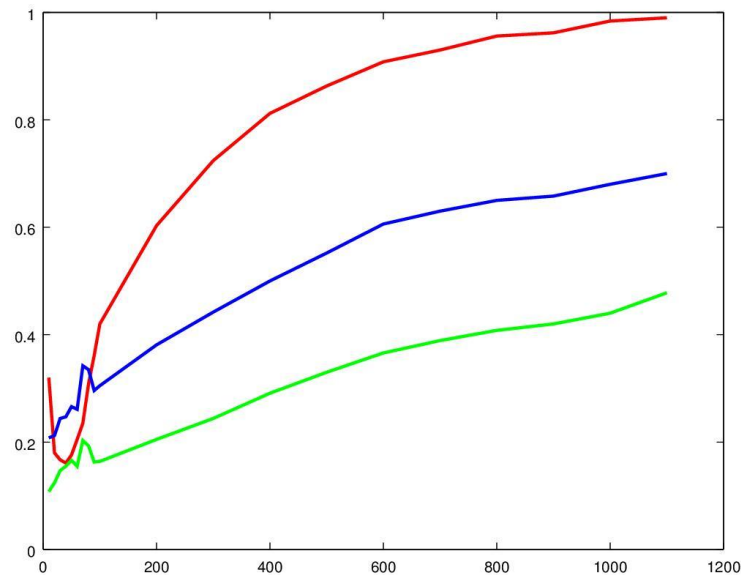
Tabla 5. *precision* y *recall* obtenidos en la recuperación de leyes con n=5 (método 1).

Cardinalidad de S	Elementos evaluados	<i>precision</i>	<i>recall</i>
<=10	122	0.108	0.992
<=20	621	0.124	0.925
<=30	1154	0.147	0.935
<=40	1864	0.155	0.854
<=50	2701	0.166	0.852
<=60	2376	0.155	0.806
<=70	2538	0.203	0.844
<=80	1701	0.193	0.838
<=90	1277	0.163	0.784
<=100	1094	0.164	0.802
<=200	7773	0.205	0.848
<=300	3758	0.244	0.877
<=400	2492	0.291	0.86
<=500	1808	0.33	0.848
<=600	1654	0.366	0.839
<=700	1476	0.389	0.823
<=800	1227	0.408	0.8
<=900	958	0.42	0.784
<=1000	705	0.44	0.775
> 1000	1890	0.478	0.719

**Tabla 6.** *precision* y *recall* obtenidos en la recuperación de leyes con n=10 (método 1).



En la figura 15 ilustramos la medida *precision* considerando únicamente las primeras  $n$  leyes recuperadas, en rojo se muestra para  $n=2$ ; en azul para  $n=5$  y en verde para  $n=10$ .



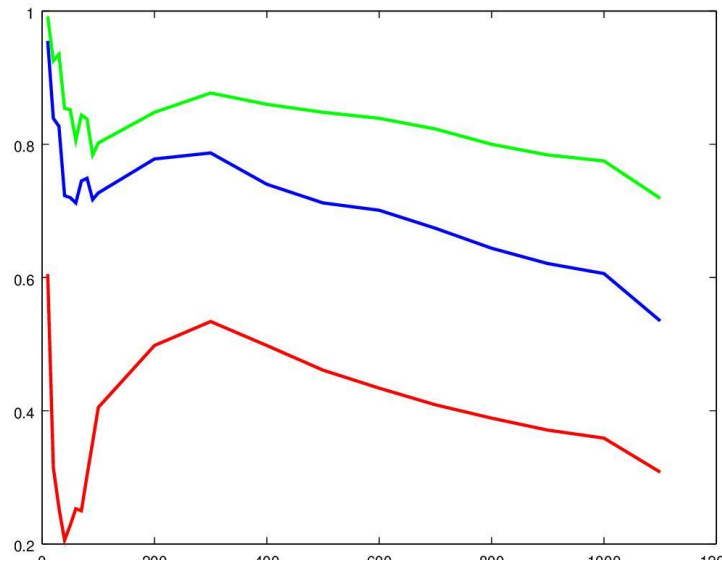
**Figura 15. Comparación de la medida *precision* en la recuperación de leyes (método 1).**

En general se observa que aumentando el número de leyes recuperadas disminuye la precisión y, aumentando el número de palabras esta aumenta. Pudiera parecer excesivo el número de palabras necesarias para obtener en *precision* un resultado considerable, sin embargo, observando las tablas anteriores vemos que el 39% de las consultas se realizó con conjuntos de cardinalidad mayor o igual a 11 y menor o igual a 100; por otra parte, el 29% de las consultas se hizo con conjuntos de cardinalidad mayor o igual a 101 y menor o igual a 300. Lo anterior, nos da una idea del número de palabras que son necesarias para describir un asunto jurídico.

La baja precisión con pocas palabras en este caso se debe a que un *considerando* corto incluirá pocos artículos, en la mayoría de los casos sólo uno y en consecuencia una ley, por lo que, el realizar esta evaluación tomando un número fijo de leyes recuperadas se refleja en una reducción para esta medida.

El incremento en la precisión aumentando el número de palabras se deriva de lo siguiente, un *considerando* extenso incluirá más artículos y por lo tanto un mayor número de leyes, aunado a esto, el tomar un número fijo de leyes recuperadas produce un aumento en la precisión sobre todo cuando evaluamos sobre un menor número de elementos.

En la figura 16 mostramos el *recall* tomando en cuenta sólo las primeras  $n$  leyes recuperadas. Al igual que en la gráfica anterior, en rojo se muestra para  $n=2$ , en azul para  $n=5$  y en verde para  $n=10$ .



**Figura 16. Comparación de la medida *recall* en la recuperación de leyes (método 1).**

En este caso observamos que el aumentar el número de palabras no se refleja en un incremento para esta medida de evaluación, sin embargo, si se aprecia una mejora al incrementar el número de leyes; además, los resultados para  $n=5$  y  $n=10$  son más altos cuando el número de palabras es menor o igual que 30 por lo cual consideramos que este método es de utilidad en la recuperación de leyes cuando existen pocas palabras en la petición.

A diferencia de lo que ocurre con la precisión, a partir de aproximadamente 300 palabras se ve un decremento en el *recall* en los tres casos, esto sugiere que una gran cantidad de palabras en la petición aumenta el valor para leyes no relevantes para un caso específico, lo que genera una reducción en el *recall*.

Mostramos ahora *precisión* y *recall* en la búsqueda de artículos relevantes considerando únicamente las primeras  $n$  leyes de nuestra lista y los primeros  $m$  artículos de cada ley.

Cardinalidad de S	<i>precision</i> <i>m=5</i>	<i>recall</i> <i>m=5</i>	<i>precision</i> <i>m=10</i>	<i>recall</i> <i>m=10</i>	<i>precision</i> <i>m=15</i>	<i>recall</i> <i>m=15</i>
<=10	0.02	0.148	0.014	0.186	0.011	0.244
<=20	0.026	0.137	0.017	0.167	0.014	0.201
<=30	0.016	0.119	0.01	0.146	0.008	0.162
<=40	0.01	0.055	0.008	0.084	0.006	0.1
<=50	0.013	0.058	0.012	0.095	0.009	0.11
<=60	0.021	0.09	0.013	0.113	0.011	0.131
<=70	0.026	0.1	0.015	0.117	0.011	0.132
<=80	0.034	0.109	0.02	0.131	0.015	0.149
<=90	0.042	0.131	0.024	0.151	0.018	0.171
<=100	0.053	0.146	0.031	0.173	0.022	0.197
<=200	0.093	0.198	0.054	0.232	0.04	0.258
<=300	0.121	0.172	0.075	0.213	0.058	0.249
<=400	0.12	0.134	0.08	0.176	0.063	0.206
<=500	0.109	0.084	0.083	0.128	0.068	0.157
<=600	0.123	0.079	0.094	0.121	0.08	0.156
<=700	0.139	0.072	0.103	0.107	0.089	0.139
<=800	0.141	0.063	0.104	0.093	0.09	0.122
<=900	0.166	0.066	0.12	0.095	0.101	0.122
<=1000	0.169	0.059	0.127	0.089	0.111	0.117
> 1000	0.188	0.05	0.139	0.074	0.12	0.096

**Tabla 7.** *precision* y *recall* obtenidos en la recuperación de artículos con  $n=2$ , se resaltan las columnas con los mejores resultados, en amarillo para *precision*, en naranja para *recall* (método 1).

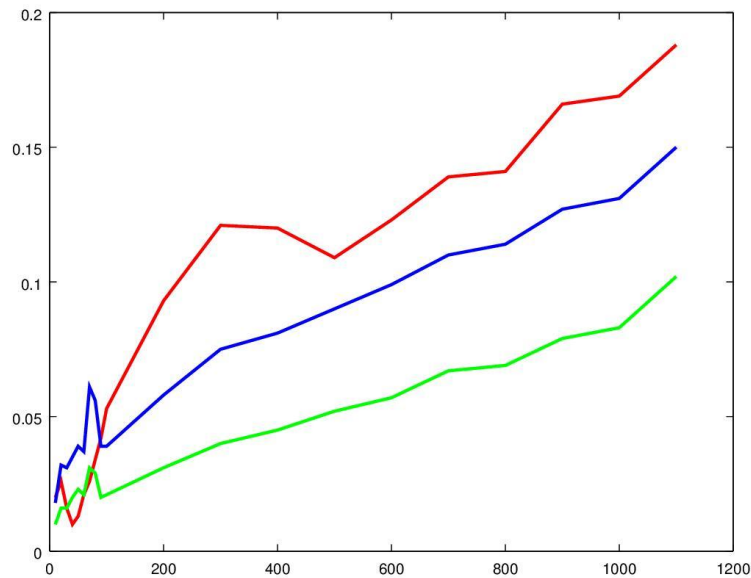
Cardinalidad de S	<i>precision</i> m=5	<i>recall</i> m=5	<i>precision</i> m=10	<i>recall</i> m=10	<i>precision</i> m=15	<i>recall</i> m=15
<=10	0.018	0.299	0.013	0.379	0.011	0.466
<=20	0.032	0.417	0.021	0.492	0.018	0.57
<=30	0.031	0.397	0.021	0.488	0.017	0.539
<=40	0.035	0.318	0.024	0.424	0.019	0.484
<=50	0.039	0.309	0.028	0.428	0.022	0.493
<=60	0.037	0.312	0.027	0.422	0.02	0.462
<=70	0.061	0.386	0.04	0.493	0.028	0.53
<=80	0.056	0.354	0.035	0.445	0.025	0.485
<=90	0.039	0.292	0.027	0.387	0.02	0.43
<=100	0.039	0.273	0.029	0.381	0.022	0.427
<=200	0.058	0.305	0.038	0.398	0.029	0.45
<=300	0.075	0.26	0.052	0.356	0.041	0.416
<=400	0.081	0.206	0.061	0.302	0.048	0.36
<=500	0.09	0.163	0.071	0.259	0.059	0.324
<=600	0.099	0.15	0.082	0.25	0.069	0.32
<=700	0.11	0.136	0.091	0.23	0.077	0.296
<=800	0.114	0.123	0.096	0.212	0.083	0.274
<=900	0.127	0.123	0.106	0.208	0.09	0.267
<=1000	0.131	0.11	0.113	0.195	0.098	0.255
> 1000	0.15	0.097	0.133	0.174	0.116	0.23

Tabla 8. *precision* y *recall* obtenidos en la recuperación de artículos con n=5, se resaltan las columnas con los mejores resultados, en amarillo para *precision*, en naranja para *recall* (método 1).

Cardinalidad de S	<i>precision</i> m=5	<i>recall</i> m=5	<i>precision</i> m=10	<i>recall</i> m=10	<i>precision</i> m=15	<i>recall</i> m=15
<=10	0.01	0.311	0.007	0.403	0.006	0.49
<=20	0.016	0.424	0.012	0.532	0.01	0.621
<=30	0.016	0.407	0.013	0.537	0.01	0.597
<=40	0.02	0.348	0.015	0.491	0.012	0.558
<=50	0.023	0.351	0.018	0.497	0.014	0.567
<=60	0.021	0.333	0.017	0.473	0.012	0.522
<=70	0.031	0.397	0.023	0.547	0.016	0.589
<=80	0.029	0.364	0.02	0.485	0.014	0.527
<=90	0.02	0.301	0.015	0.416	0.011	0.462
<=100	0.021	0.289	0.015	0.407	0.012	0.456
<=200	0.031	0.325	0.02	0.425	0.016	0.481
<=300	0.04	0.281	0.028	0.39	0.022	0.456
<=400	0.045	0.232	0.034	0.346	0.027	0.413
<=500	0.052	0.195	0.041	0.309	0.034	0.383
<=600	0.057	0.181	0.047	0.297	0.04	0.375
<=700	0.067	0.172	0.054	0.284	0.045	0.358
<=800	0.069	0.154	0.057	0.26	0.049	0.333
<=900	0.079	0.159	0.065	0.266	0.055	0.337
<=1000	0.083	0.148	0.071	0.252	0.06	0.322
> 1000	0.102	0.138	0.087	0.237	0.075	0.303

Tabla 9. *precision* y *recall* obtenidos en la recuperación de artículos con n=10, se resaltan las columnas con los mejores resultados, en amarillo para *precision*, en naranja para *recall* (método 1).

Tomando en cuenta únicamente las primeras  $n$  leyes y los primeros  $m$  artículos de cada ley se puede observar que se alcanzan mejores resultados en *precision* con  $m=5$ . En la figura 17 comparamos los resultados obtenidos variando únicamente el valor de  $n$ , en rojo se muestra para  $n=2$ , en azul para  $n=5$  y en verde para  $n=10$ .

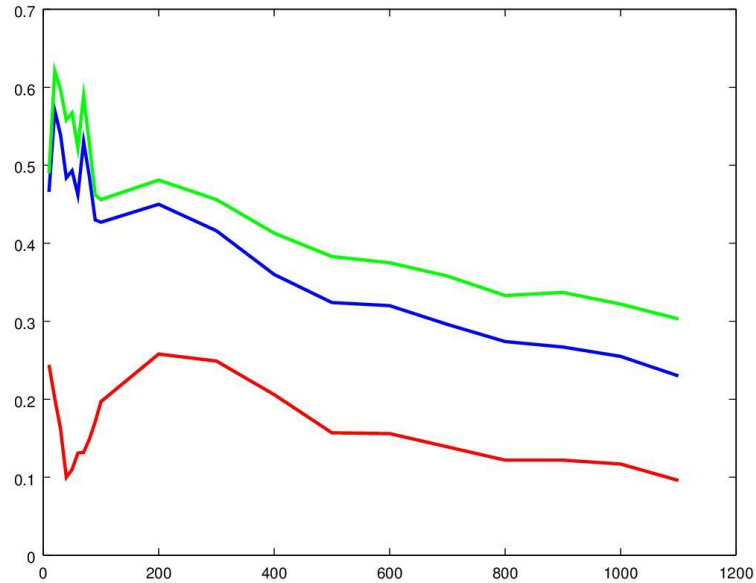


**Figura 17. Comparación de la medida *precision* en la recuperación de artículos (método 1).**

Aunque en general la precisión es baja, para consultas con menos de 100 palabras podemos ver una mejor evaluación tomando las primeras 5 leyes de nuestra lista, sin embargo, en general la precisión resulta ser mayor si tomamos menos leyes, además esta aumenta conforme se incrementa el número de palabras en la solicitud.

De manera análoga a lo que sucede en la recuperación de leyes, la mejora en la precisión al aumentar el número de palabras se debe a que *considerandos* extensos contendrán un mayor número de artículos relevantes, por lo que el tomar en cuenta sólo un número fijo de artículos recuperados, se refleja en un aumento en la precisión, sin embargo, también por este motivo obtenemos una precisión baja ya que la mayoría de los *considerandos* contienen en promedio menos de 10 artículos.

En cuanto a *recall*, los mejores resultados se muestran en la figura 18; en rojo se muestra para  $n=2$  y  $m=15$ ; en azul para  $n=5$  y  $m=15$  y en verde para  $n=10$  y  $m=15$ .



**Figura 18. Comparación de la medida *recall* en la recuperación de artículos (método 1).**

Es claro que la mejor evaluación se presenta tomando un mayor número de leyes, sin embargo, no se aprecia un incremento significativo al pasar de 5 a 10 leyes, lo que en este caso equivale a pasar de 75 a 150 artículos; además podemos ver que incrementar el número de palabras no mejora la búsqueda de artículos, por el contrario a mayor número de palabras, menor es la cantidad de artículos relevantes encontrados.

### 5.3. Resultados experimentales aplicando el método 2

Para evaluar los resultados obtenidos a partir de la aplicación del método 2, cada uno de los considerandos fue dividido de la misma manera que en la sección anterior, es decir:

- Un conjunto **S** compuesto por los unigramas, bigramas y conceptos jurídicos del considerando.
- Un conjunto **T** compuesto por los artículos y jurisprudencias del considerando.

Al igual que en el método 1, por cada conjunto **S** (solicitud de información) se obtuvo una lista de leyes ordenadas y, por cada ley se obtuvo una lista de artículos ordenados de acuerdo a los resultados obtenidos al aplicar el método descrito en la sección 4.3.2; la figura 14 muestra la forma en la que se ordenan los resultados.

En primer lugar, mostramos *precision* y *recall* en la búsqueda de leyes relevantes considerando únicamente las primeras  $n$  leyes de nuestra lista.

Cardinalidad de S	Elementos evaluados	<i>Precisión</i>	<i>recall</i>
<=10	122	0.344	0.667
<=20	621	0.182	0.278
<=30	1154	0.262	0.335
<=40	1864	0.334	0.391
<=50	2701	0.358	0.412
<=60	2376	0.341	0.402
<=70	2538	0.406	0.382
<=80	1701	0.418	0.377
<=90	1277	0.421	0.404
<=100	1094	0.404	0.398
<=200	7773	0.427	0.377
<=300	3758	0.5	0.388
<=400	2492	0.604	0.368
<=500	1808	0.674	0.353
<=600	1654	0.885	0.416
<=700	1476	0.942	0.408
<=800	1227	0.957	0.385
<=900	958	0.962	0.366
<=1000	705	0.983	0.353
> 1000	1890	0.992	0.303

**Tabla 10.** *precision* y *recall* obtenidos en la recuperación de leyes con  $n=2$  (método 2).



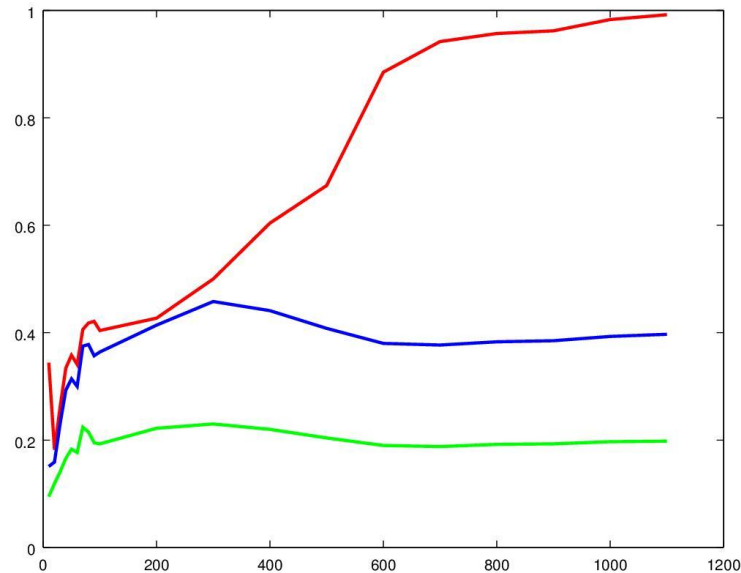
Cardinalidad de S	Elementos evaluados	<i>precision</i>	<i>recall</i>
<=10	122	0.151	0.732
<=20	621	0.159	0.537
<=30	1154	0.231	0.712
<=40	1864	0.293	0.805
<=50	2701	0.314	0.817
<=60	2376	0.3	0.818
<=70	2538	0.375	0.83
<=80	1701	0.378	0.839
<=90	1277	0.357	0.859
<=100	1094	0.364	0.85
<=200	7773	0.414	0.825
<=300	3758	0.458	0.799
<=400	2492	0.441	0.646
<=500	1808	0.408	0.527
<=600	1654	0.38	0.446
<=700	1476	0.377	0.408
<=800	1227	0.383	0.385
<=900	958	0.385	0.366
<=1000	705	0.393	0.353
> 1000	1890	0.397	0.303

**Tabla 11.** *precision* y *recall* obtenidos en la recuperación de leyes con n=5 (método 2).

Cardinalidad de S	Elementos evaluados	<i>precision</i>	<i>recall</i>
<=10	122	0.095	0.874
<=20	621	0.119	0.872
<=30	1154	0.141	0.896
<=40	1864	0.166	0.936
<=50	2701	0.183	0.951
<=60	2376	0.177	0.944
<=70	2538	0.224	0.955
<=80	1701	0.215	0.937
<=90	1277	0.195	0.927
<=100	1094	0.193	0.914
<=200	7773	0.222	0.899
<=300	3758	0.23	0.805
<=400	2492	0.22	0.646
<=500	1808	0.204	0.527
<=600	1654	0.19	0.446
<=700	1476	0.188	0.408
<=800	1227	0.192	0.385
<=900	958	0.193	0.367
<=1000	705	0.197	0.353
> 1000	1890	0.198	0.303

**Tabla 12.** *precision* y *recall* obtenidos en la recuperación de leyes con n=10 (método 2).

En la figura 19 se muestra la gráfica de la medida *precisión* considerando únicamente las primeras  $n$  leyes recuperadas, en rojo se muestra para  $n=2$ ; en azul para  $n=5$  y en verde para  $n=10$ .

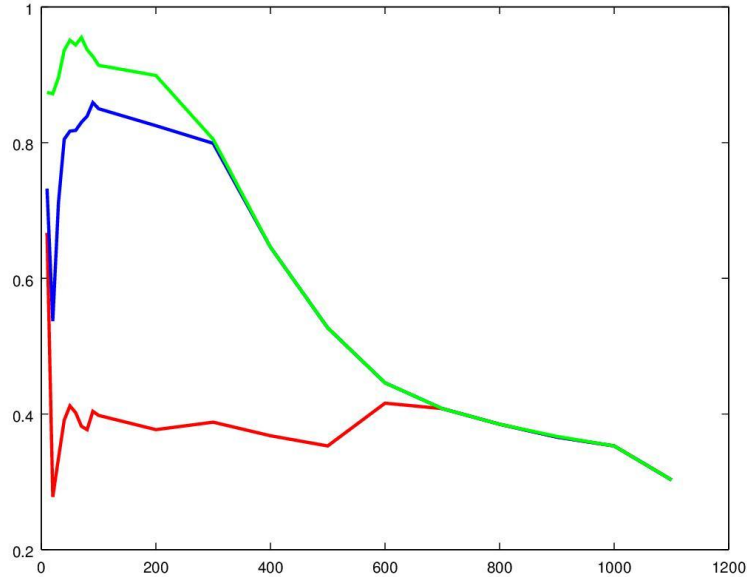


**Figura 19. Comparación de la medida *precisión* en la recuperación de leyes (método 2).**

Al aumentar el número de leyes observamos una disminución para esta medida de evaluación, por otro lado, inicialmente incrementar el número de palabras aumenta la precisión en los tres casos, aunque con más de 200 palabras no se aprecian cambios relevantes con  $n=5$  y  $n=10$ , esto no ocurre en el caso de  $n=2$  en donde incrementar el número de palabras si se refleja en una mejora en la evaluación.

La principal razón de una baja precisión es que en el 80% de los *considerandos* no hay más de 3 leyes relevantes, lo que genera que con  $n=5$  y  $n=10$  tengamos una precisión limitada. Por otra parte, cuando  $n=2$  si es visible una mejora en la precisión cuando tenemos un mayor número de palabras, pues también se incrementa el número de leyes.

En la figura 20 mostramos el *recall* tomando en cuenta sólo las primeras  $n$  leyes recuperadas, al igual que en la gráfica anterior, en rojo se muestra para  $n=2$ ; en azul para  $n=5$  y en verde para  $n=10$ .



**Figura 20. Comparación de la medida *recall* en la recuperación de leyes (método 2).**

En primer término observamos que el aumentar el número de palabras no representa un incremento para esta medida de evaluación; de hecho, con más de 200 palabras para  $n=5$  y  $n=10$  disminuye la cantidad de leyes relevantes recuperadas. Esto sugiere, que el tener una gran cantidad de palabras aumenta la probabilidad de leyes que no son relevantes para un caso específico y aumenta la probabilidad de las leyes más comunes a todas las palabras, aún y cuando estas no versen sobre el caso.

En segundo lugar, con menos de 300 palabras se observa una mejora significativa al aumentar el número de leyes. El principal motivo es que para esta cantidad de términos el número de leyes en promedio es menor a 3 por lo cual cuando  $n=5$  y  $n=10$  se recuperan la mayoría de leyes relevantes.

Presentamos ahora *precision* y *recall* en la búsqueda de artículos relevantes considerando únicamente las primeras  $n$  leyes y los primeros  $m$  artículos de los resultados experimentales.

Cardinalidad de S	precision m=5	recall m=5	precision m=10	recall m=10	precision m=15	recall m=15
<=10	0.026	0.158	0.014	0.17	0.01	0.172
<=20	0.023	0.131	0.012	0.138	0.008	0.144
<=30	0.036	0.182	0.019	0.193	0.013	0.202
<=40	0.087	0.291	0.044	0.302	0.03	0.306
<=50	0.102	0.319	0.054	0.333	0.036	0.34
<=60	0.091	0.292	0.048	0.308	0.033	0.333
<=70	0.111	0.281	0.058	0.295	0.04	0.321
<=80	0.102	0.246	0.053	0.259	0.038	0.284
<=90	0.1	0.257	0.054	0.272	0.039	0.29
<=100	0.083	0.231	0.044	0.242	0.032	0.258
<=200	0.098	0.211	0.055	0.232	0.039	0.247
<=300	0.127	0.192	0.072	0.215	0.05	0.229
<=400	0.143	0.161	0.074	0.17	0.05	0.17
<=500	0.143	0.119	0.072	0.12	0.048	0.12
<=600	0.131	0.09	0.066	0.09	0.044	0.09
<=700	0.126	0.076	0.063	0.076	0.042	0.076
<=800	0.132	0.067	0.066	0.067	0.044	0.067
<=900	0.136	0.061	0.068	0.061	0.046	0.061
<=1000	0.147	0.06	0.074	0.06	0.049	0.06
> 1000	0.104	0.033	0.052	0.033	0.035	0.033

Tabla 13. *precision* y *recall* obtenidos en la recuperación de artículos con  $n=2$ , se resaltan las columnas con los mejores resultados, en amarillo para *precision*, en naranja para *recall* (método 2).

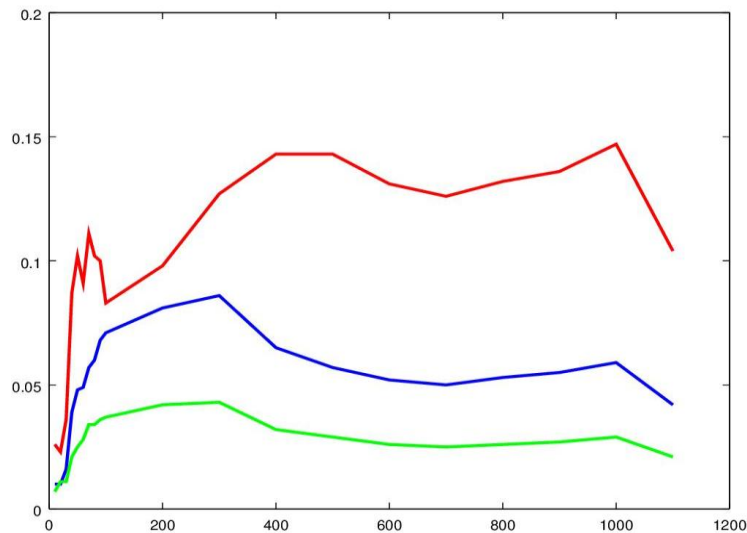
Cardinalidad de S	precision m=5	recall m=5	precision m=10	recall m=10	precision m=15	recall m=15
<=10	0.01	0.158	0.006	0.178	0.004	0.18
<=20	0.01	0.139	0.007	0.196	0.006	0.229
<=30	0.016	0.198	0.013	0.354	0.01	0.383
<=40	0.039	0.324	0.027	0.501	0.019	0.513
<=50	0.048	0.363	0.032	0.517	0.022	0.542
<=60	0.049	0.376	0.031	0.537	0.024	0.59
<=70	0.057	0.365	0.036	0.525	0.028	0.587
<=80	0.06	0.371	0.039	0.522	0.029	0.567
<=90	0.068	0.435	0.043	0.583	0.031	0.617
<=100	0.071	0.45	0.043	0.565	0.031	0.594
<=200	0.081	0.416	0.051	0.511	0.035	0.536
<=300	0.086	0.312	0.053	0.37	0.036	0.385
<=400	0.065	0.183	0.034	0.193	0.023	0.194
<=500	0.057	0.12	0.029	0.12	0.019	0.12
<=600	0.052	0.09	0.026	0.09	0.017	0.09
<=700	0.05	0.076	0.025	0.076	0.017	0.076
<=800	0.053	0.067	0.026	0.067	0.018	0.067
<=900	0.055	0.061	0.027	0.061	0.018	0.061
<=1000	0.059	0.06	0.029	0.06	0.02	0.06
> 1000	0.042	0.033	0.021	0.033	0.014	0.033

Tabla 14. *precision* y *recall* obtenidos en la recuperación de artículos con  $n=5$ , se resaltan las columnas con los mejores resultados, en amarillo para *precision*, en naranja para *recall* (método 2).

Cardinalidad de S	precision m=5	recall m=5	precision m=10	recall m=10	precision m=15	recall m=15
<=10	0.007	0.179	0.006	0.258	0.004	0.276
<=20	0.011	0.291	0.009	0.402	0.007	0.444
<=30	0.011	0.252	0.009	0.443	0.007	0.482
<=40	0.021	0.351	0.015	0.546	0.011	0.568
<=50	0.025	0.377	0.017	0.563	0.012	0.599
<=60	0.028	0.404	0.018	0.581	0.013	0.642
<=70	0.034	0.415	0.021	0.584	0.016	0.654
<=80	0.034	0.404	0.022	0.567	0.016	0.617
<=90	0.036	0.452	0.023	0.609	0.016	0.649
<=100	0.037	0.467	0.022	0.592	0.016	0.629
<=200	0.042	0.43	0.027	0.536	0.019	0.569
<=300	0.043	0.312	0.026	0.371	0.018	0.386
<=400	0.032	0.183	0.017	0.193	0.011	0.194
<=500	0.029	0.12	0.014	0.12	0.01	0.12
<=600	0.026	0.09	0.013	0.09	0.009	0.09
<=700	0.025	0.076	0.013	0.076	0.008	0.076
<=800	0.026	0.067	0.013	0.067	0.009	0.067
<=900	0.027	0.061	0.014	0.061	0.009	0.061
<=1000	0.029	0.06	0.015	0.06	0.01	0.06
> 1000	0.021	0.033	0.01	0.033	0.007	0.033

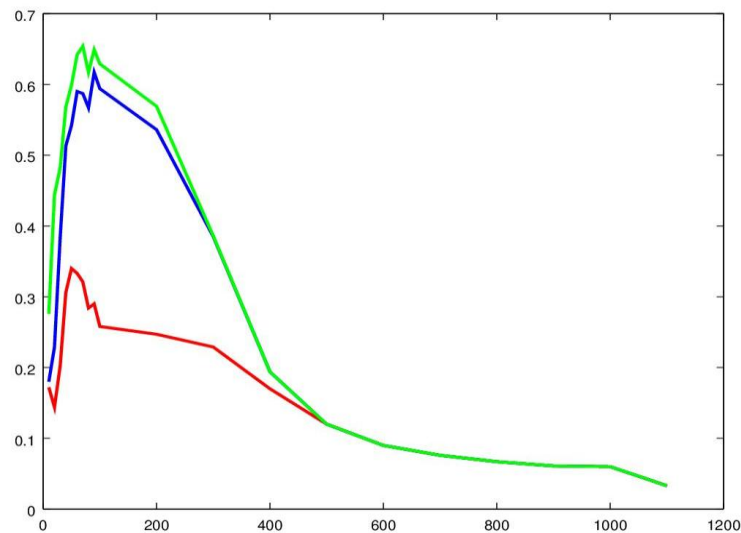
Tabla 15. *precision* y *recall* obtenidos en la recuperación de artículos con  $n=10$ , se resaltan las columnas con los mejores resultados, en amarillo para *precision*, en naranja para *recall* (método 2).

Tomando en cuenta las primeras  $n$  leyes y los primeros  $m$  artículos de cada ley se puede observar que los mejores resultados en *precisión* se alcanzan con  $m=5$ . En la figura 21 comparamos los resultados obtenidos variando únicamente el valor de  $n$ , en rojo se muestra para  $n=2$ , en azul para  $n=5$  y en verde para  $n=10$ .



**Figura 21. Comparación de la medida *precisión* en la recuperación de artículos (método 2).**

La precisión en la recuperación de artículos siempre es mayor si tomamos un menor número de leyes, sin embargo, el máximo aproximado que se llega a alcanzar es de 0.15, esto ocurre debido a que siempre regresamos un número fijo de artículos, aun y cuando en muchas de las peticiones sólo existe un artículo relevante, lo que genera una precisión baja sobre todo al regresar un mayor número de artículos.



**Figura 22. Comparación de la medida *recall* en la recuperación de artículos (método 2).**



En cuanto a *recall*, los mejores resultados se muestran en la figura 22. En verde para  $n=10$  y  $m=15$ ; en azul para  $n=5$  y  $m=15$ ; en rojo para  $n=2$  y  $m=15$ .

En este caso se aprecia que la medida *recall* aumenta si incrementamos el número de leyes, esta mejora no es significativa al pasar de 5 a 10 leyes, por otra observamos un decremento en la evaluación al aumentar el número de palabras, y finalmente una coincidencia a partir de 500 palabras, esto sugiere que un gran número de palabras en la petición aumenta la probabilidad de leyes y artículos no relevantes, desplazando los artículos y leyes relevantes a posiciones más lejanas, esto limita el desempeño del método a peticiones con pocas palabras.

## 5.4. Resultados experimentales aplicando el método 3

Para el experimento 3, cada considerando fue dividido de la siguiente manera:

- Un conjunto **S** compuesto por los unigramas, bigramas y conceptos jurídicos del considerando.
- Un conjunto **T** compuesto por únicamente por los artículos del considerando.

Nuestro objetivo fue utilizar los conjuntos **S** (solicitud de información) para encontrar los conjuntos **T** (términos jurídicos) para lo cual fue aplicando el método descrito en la sección 4.3.3; cabe señalar que sólo ocupamos artículos para realizar este experimento, por lo cual únicamente trabajamos con los considerandos que tuvieron al menos un artículo, resultando éstos en un total de 34,931.

Para realización del experimento, los considerandos fueron agrupados utilizando el algoritmo DBSCAN con los siguientes parámetros:

$e = 0.75$ . Este parámetro indica la similitud mínima que debe existir entre dos patrones para que estos sean considerados vecinos.

$m = 3$ . Indica el número mínimo de vecinos (con similitud  $\geq e$ ) para que un patrón sea considerado de alta densidad.

Los parámetros fueron seleccionados considerando el promedio de vecinos que se alcanza a cierta similitud, tomamos un número de vecinos pequeño y una similitud alta ya que nos interesaba que los grupos fueran compactos, así como generar una gran cantidad de grupos, esto a efecto de que al realizar una petición en los grupos más similares no existiera una gran cantidad de artículos.

Como se describió en el apartado 4.3.3, una vez agrupados los considerandos formamos un VSM en el cual los vectores representan grupos de considerandos y las dimensiones de estos vectores representan lo que denominamos *elementos relacionados*, es decir, unigramas, bigramas y conceptos jurídicos; los valores que puede tomar cada dimensión se asignan utilizando la medida *tf-idf*.

Posteriormente cada uno de los conjuntos **S** fue agregado al VSM y por cada conjunto se obtuvo una lista con todos los grupos ordenada de mayor a menor con base en la similitud que estos tuvieron con **S**. Finalmente se tomaron los **n** grupos más similares a **S** y, de cada grupo se obtuvieron los **k** artículos más frecuentes en el grupo, siendo ésta la respuesta del programa.

Dado que conocemos la respuesta a cada conjunto **S**, la evaluación de los resultados se realizó utilizando los conjuntos **T** y las medidas de evaluación *precision* y *recall*. Los resultados obtenidos se muestran en la tabla 16 y se ilustran en la figura 23. En rojo se muestran *precision* y *recall* tomando 10 grupos y 10 artículos por grupo; en azul se muestran las mismas medidas tomando 10 grupos y 5 artículos por grupo.

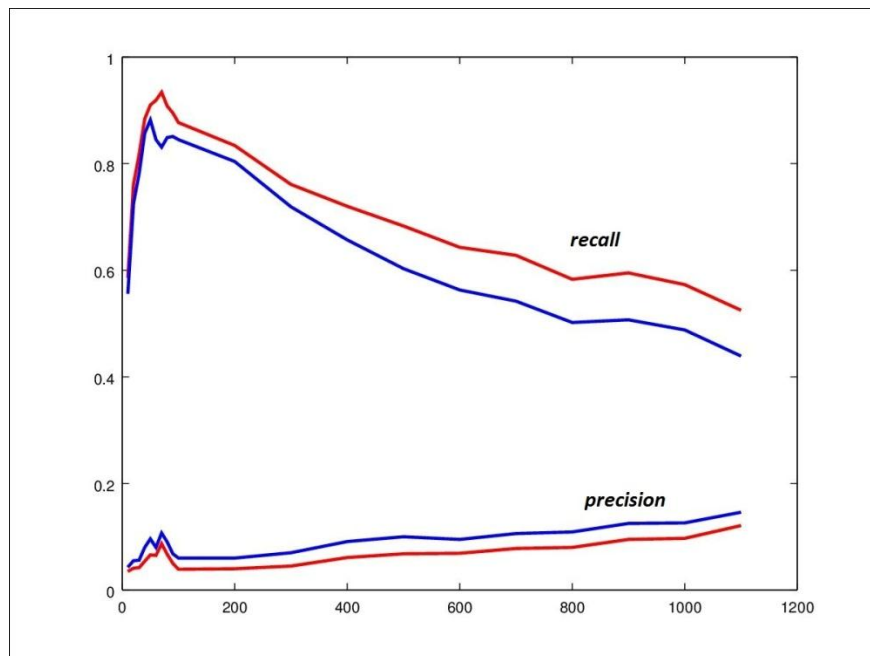


Figura 23. Comparación de las medidas *precision* y *recall* (método 3).

Cardinalidad de S	Elementos evaluados	<i>precision</i> k=10	<i>recall</i> k=10	<i>precision</i> k=5	<i>recall</i> k=5
<=10	49	0.035	0.586	0.043	0.556
<=20	490	0.041	0.761	0.055	0.725
<=30	926	0.042	0.817	0.056	0.781
<=40	1604	0.054	0.884	0.08	0.857
<=50	2278	0.066	0.91	0.096	0.882
<=60	2004	0.065	0.919	0.08	0.845
<=70	2239	0.087	0.934	0.107	0.831
<=80	1515	0.067	0.908	0.09	0.849
<=90	1115	0.05	0.895	0.068	0.851
<=100	928	0.039	0.877	0.06	0.845
<=200	6572	0.04	0.834	0.06	0.804
<=300	3254	0.045	0.761	0.07	0.719
<=400	2336	0.061	0.72	0.091	0.657
<=500	1747	0.068	0.683	0.1	0.603
<=600	1633	0.069	0.643	0.095	0.563
<=700	1472	0.078	0.628	0.106	0.542
<=800	1221	0.08	0.583	0.109	0.502
<=900	957	0.095	0.595	0.125	0.507
<=1000	704	0.097	0.573	0.126	0.488
> 1000	1887	0.121	0.525	0.146	0.439

Tabla 16. *precision* y *recall* obtenidos con 10 grupos (método 3).

Podemos observar que el mayor número de artículos relevantes recuperados se alcanza con peticiones de menos de 200 palabras, y que conforme aumenta el número de palabras, el número de artículos relevantes y recuperados disminuye aunque la medida *precision* aumenta ligeramente. A pesar de que en la gráfica no se aprecia con claridad, con menos de 20 palabras se alcanza un *recall* considerable ( $> 0.7$ ).

## 5.5. Comparación de resultados (métodos 1, 2 y 3)

La comparación que se presenta fue realizada tomando 50 artículos recuperados en cada uno de los métodos, para los métodos 1 y 2 se tomaron 10 leyes y 5 artículos de cada ley; para el método 3 se tomaron 10 grupos y 5 artículos de cada grupo.

En la figura 24 comparamos los resultados obtenidos utilizando la medida *precision*; en azul se muestra el resultado obtenido en el método 1, en rojo el resultado obtenido en el método 2 y en negro el resultado obtenido en el método 3.

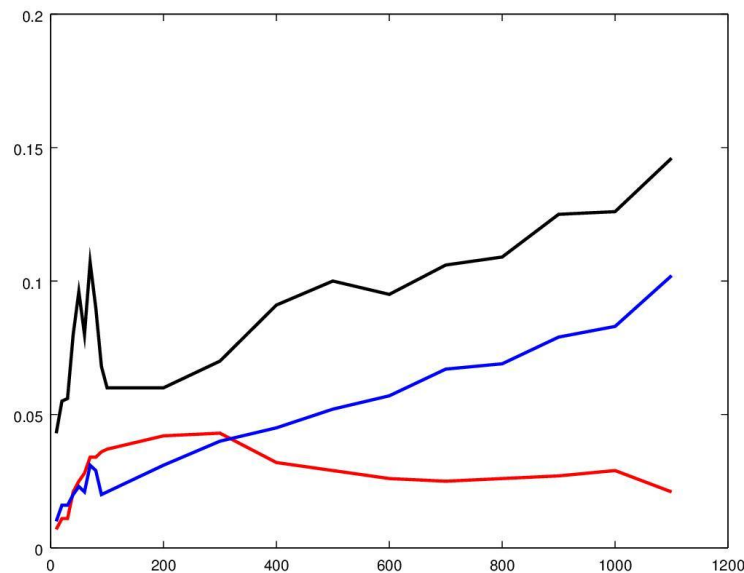
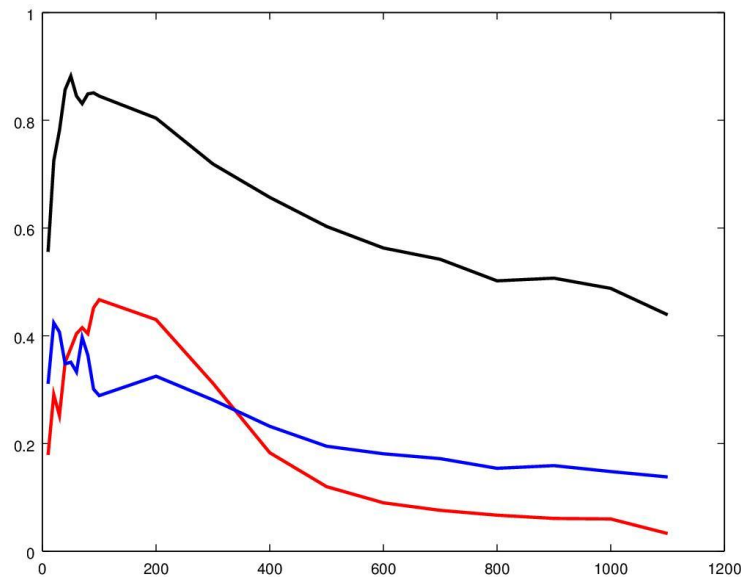


Figura 24. Comparación de la medida *precision* (métodos 1, 2, 3).

Aunque en los tres métodos la precisión es muy baja en la recuperación de artículos los mejores resultados fueron obtenidos al aplicar el método 3; además, tanto en el método 1 como en el método 3 la precisión aumenta si incrementamos el número de palabras en la petición. La razón de una precisión tan baja en los tres métodos, es que para esta comparación siempre se tomaron 50 artículos recuperados, lo cual genera una precisión baja sobre todo en peticiones en las que sólo existía un artículo relevante.

Respecto a *recall*, en la figura 25 ilustramos la comparación, en azul se muestra el resultado obtenido a partir del método 1, en rojo aplicando el método 2 y en negro el método 3.



**Figura 25. Comparación de la medida *recall* (métodos 1, 2, 3).**

En este caso también se observa que el método 3 obtiene un mejor desempeño que los métodos 1 y 2 en la recuperación de artículos relevantes, aunque en los tres casos con menos de 200 palabras se aprecia una mejora al aumentar el número de palabras en la solicitud. Por otra parte, con más de 200 palabras los tres métodos muestran un decrecimiento para esta evaluación, esto sugiere que demasiadas palabras en la petición, colocan en una mejor posición artículos no relevantes.

## 5.6. Resultados experimentales aplicando el método 4

El objetivo de este experimento es encontrar un conjunto **R** integrado por los términos legales más relacionados a un conjunto **L** compuesto también por términos legales, es decir,  $f(L) = R$ .

Para realizar esta evaluación sólo consideramos artículos; y, para obtener nuestro conjunto de datos para evaluar, de los 39,189 considerandos tomamos aquellos que tuvieran entre 2 y 10 artículos para formar los conjuntos **L**.

Por cada considerando seleccionado, formamos los conjuntos **L** a partir de las combinaciones de los artículos presentes en un considerando, es decir, dado un considerando con  $n$  artículos formamos todas las combinaciones desde 1 hasta  $n-1$  artículos; por cada una de estas combinaciones se formó un conjunto **L**; y, finalmente por cada conjunto **L** se integró un conjunto **R** con aquellos artículos del considerando no presentes en **L**. En total se formaron 1, 105, 447 conjuntos **L**.

Dado que sabemos que los artículos que contiene un considerando se encuentran relacionados, para evaluar los resultados experimentales utilizamos los conjuntos **R** para medir *precision* y *recall*.

Aplicando el método descrito en la sección 4.3.4, por cada conjunto **L** obtuvimos una lista de elementos ordenados de acuerdo a su relevancia, acotando la lista a  $n$  elementos logramos para *precision* los siguientes resultados:

**AV-1:** *precision* promedio variando el valor de  $n$  para conjuntos con la misma cardinalidad.

**AV-2:** *precision* promedio variando la cardinalidad de los conjuntos **L**.

Cardinalidad de L	Elementos evaluados	n=10	n=20	n=30	n=40	n=50	AV-1
1	3444	0.276	0.195	0.157	0.133	0.117	0.175
2	51918	0.207	0.14	0.11	0.093	0.081	0.126
3	162402	0.199	0.13	0.1	0.083	0.072	0.116
4	267250	0.188	0.12	0.092	0.076	0.065	0.108
5	283209	0.17	0.107	0.082	0.067	0.057	0.096
6	203856	0.146	0.092	0.07	0.057	0.048	0.082
7	98878	0.119	0.075	0.056	0.046	0.039	0.067
8	30015	0.091	0.057	0.043	0.034	0.029	0.05
9	4475	0.062	0.039	0.029	0.024	0.02	0.034
	<b>AV-2</b>	0.162	0.106	0.082	0.068	0.058	

**Tabla 17.** Comparación de la medida *precision* (método 4).

En cuanto a *recall* se obtuvieron los siguientes resultados:

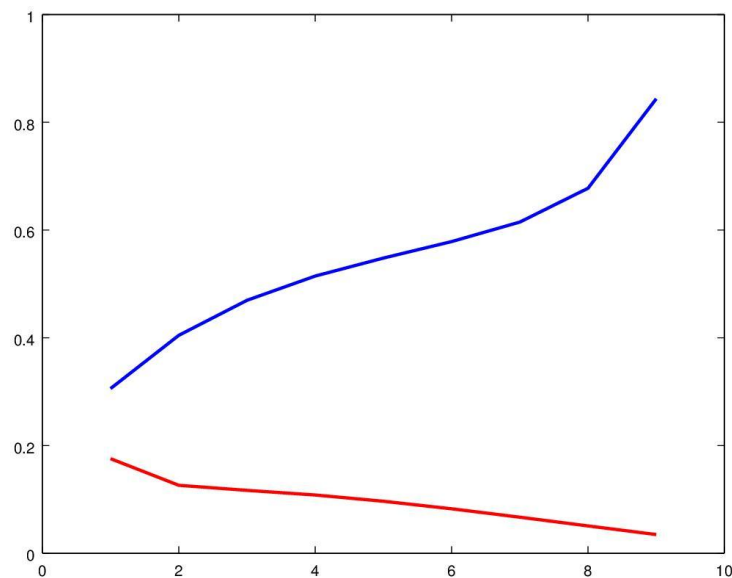
**AV-3:** *recall* promedio variando el valor de  $n$  para conjuntos con la misma cardinalidad.

**AV-4:** *recall* promedio variando la cardinalidad de los conjuntos L.

Cardinalidad de L	Elementos evaluados	n=10	n=20	n=30	n=40	n=50	AV-3
1	1223	0.214	0.277	0.319	0.348	0.371	0.305
2	32134	0.274	0.36	0.42	0.466	0.503	0.404
3	122352	0.329	0.423	0.487	0.535	0.574	0.469
4	214568	0.368	0.467	0.533	0.583	0.622	0.514
5	245583	0.397	0.5	0.568	0.618	0.657	0.548
6	172338	0.423	0.53	0.599	0.651	0.69	0.578
7	81619	0.452	0.565	0.636	0.69	0.731	0.614
8	29949	0.5	0.624	0.702	0.758	0.804	0.677
9	4475	0.623	0.779	0.874	0.943	0.999	0.843
	<b>AV-4</b>	0.397	0.502	0.57	0.621	0.661	

**Tabla 18.** Comparación de la medida *recall* (método 4).

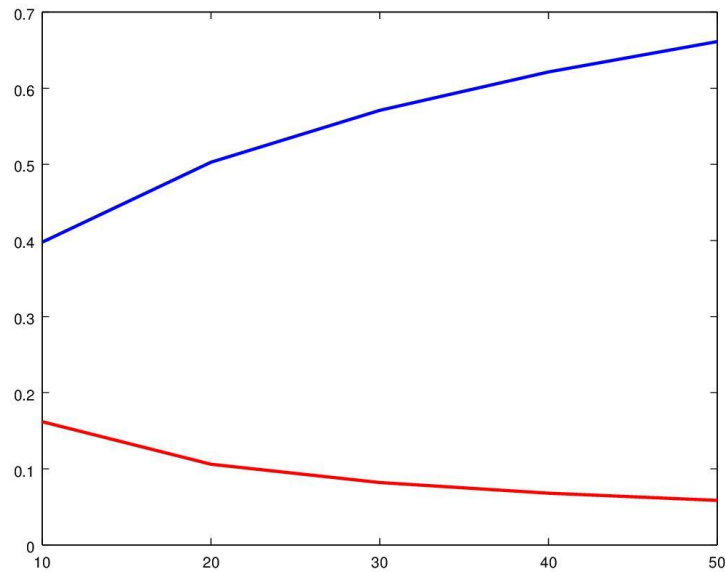
En la figura 26 mostramos en rojo los promedios de *precision* (AV-2) y en azul los promedios de *recall* (AV-4) variando la cardinalidad de los conjuntos L.



**Figura 26.** *precision* y *recall* obtenidos variando cardinalidad de L (método 4).

Como podemos apreciar, utilizar conjuntos con mayor cardinalidad para la búsqueda mejora el número de artículos relevantes recuperados. La medida *precision* en general es baja pero no se observa una afectación radical al cambiar la cardinalidad de los conjuntos.

En la figura 27 ilustramos en rojo los promedios de *precision* (AV-1) y en azul los promedios de *recall* (AV-3) variando el número de artículos recuperados.



**Figura 27. *precision* y *recall* obtenidos variando el número de artículos recuperados (método 4).**

En este caso, la mejora más significativa en la recuperación del número de artículos relevantes (*recall*) se observa al pasar de 10 a 20 artículos recuperados, sin embargo también podemos ver una pérdida significativa en la precisión.



## Conclusiones

El trabajo desarrollado se encuentra dedicado a localizar los fundamentos jurídicos aplicables a situaciones concretas de manera correcta e inmediata, su implementación en un sistema disponible para toda persona contribuiría a resolver los problemas jurídicos de manera mucho más eficiente y con un costo mínimo.

Se presentaron diferentes métodos para localizar fundamentos jurídicos sin considerar el contenido de las normas jurídicas, utilizando para ello un conjunto de *considerandos* extraído de un *corpus de decisiones judiciales*.

Para desarrollar estos métodos tomamos en cuenta las *visiones realista y constructivista del derecho*. Considerando los argumentos expuestos desde la *perspectiva realista*, en el sentido de que el derecho es una disciplina que trabaja con normas que tienen un déficit de claridad y coherencia, así como conceptos subjetivos ambiguos (y por ello poco entendidos) llegamos a la conclusión de que descubrir las relaciones que existen entre las normas jurídicas es mucho más viable utilizando casos ya resueltos, que directamente el contenido normativo.

Desde de la *perspectiva constructivista*, analizamos la forma en que son resueltos los casos, así como también la manera en que se encuentran estructuradas las decisiones judiciales, con lo cual determinamos que la forma más apropiada de alcanzar nuestros objetivos no era emplear las sentencias de manera íntegra, sino a través de sus considerandos.

A partir de los experimentos realizados se demuestra la viabilidad de encontrar relaciones entre fundamentos jurídicos sin considerar el contenido de las normas y únicamente observando los casos en que han sido aplicadas; también, se mostró que al utilizar sólo el contenido de las resoluciones judiciales, es posible localizar los fundamentos jurídicos aplicables a situaciones concretas a partir de búsquedas en términos flexibles. Esto es, utilizando palabras comunes para describir casos, esto supone una ventaja frente a los sistemas actualmente existentes en los cuales la búsqueda de fundamentos legales se realiza utilizando exclusivamente el contenido normativo.

Los resultados también demuestran que al agrupar los considerandos, mejora en mucho la cantidad de artículos relevantes recuperados (*recall*) sin embargo, es preciso realizar más pruebas con diferentes algoritmos de agrupamiento, modificando la representación de los considerandos y variando la función de comparación entre estos, con el fin de verificar la posibilidad de alcanzar una mejor evaluación utilizando una menor cantidad de palabras.

## Referencias

1. Alexy, R. (1993). *Derecho y razón práctica* (Vol. 1). Fontamara.
2. Howard, A. (1991). *Introducción al álgebra lineal*. Limusa.
3. Barragán, J. (2000). *Informática y decisión jurídica*. Fontamara.
4. Bartolini, R., Lenci, A., Montemagni, S., Pirrelli, V., y Soria, C. (2004, January). Automatic classification and analysis of provisions in italian legal texts: a case study. In *On the Move to Meaningful Internet Systems 2004: OTM 2004 Workshops* (pp. 593-604). Springer Berlin Heidelberg.
5. Biagioli, C., Francesconi, E., Passerini, A., Montemagni, S., y Soria, C. (2005, June). Automatic semantics extraction in law documents. In *Proceedings of the 10th international conference on Artificial intelligence and law* (pp. 133-140). ACM.
6. Bird, S., Klein, E., y Loper, E. (2009). *Natural language processing with Python*. " O'Reilly Media, Inc."
7. Bobbio, N. (2005). Ciencia del derecho y análisis del lenguaje. En *Filosofía analítica y filosofía del derecho en Italia*. Cajica.
8. Bonesana, C., de Beccaria, M., de Torres, G. C., y Dragonetti, J. (1993). *Tratado de los delitos y de las penas*. Heliasta.
9. Cáceres, E. (2008, July). EXPERTIUS: A Mexican Judicial Decision-Support System in the Field of Family Law. In *JURIX* (Vol. 8, pp. 78-87).
10. Calvo, H. (2013). *Procesamiento práctico del lenguaje natural*. SMIA.
11. Cardeñosa, J., y Lasala, P. (1996). ARPO-2: An expert system for legal advice on the breach of building contracts. *Artificial Intelligence and Law*, 4(2), 133-156.

12. Constantino, F., y León, M. (2010). *Elementos de lingüística jurídica. Deconstrucción conceptual: Herramientas hacia la interpretación normativa*. México, Fontamara-Universidad de Sonora.
13. Duncan Kennedy. (1999). *Libertad y restricción en la decisión judicial*. Siglo del Hombre Editores.
14. Dworkin, R. (1988). *El Imperio de la justicia*. Gedisa.
15. Dworkin, R. (1978). *Taking rights seriously* (Vol. 136). Harvard University Press.
16. Feldman, R., y Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge University Press.
17. Ferrajoli, L. (2001). Pasado y futuro del Estado de derecho. *Revista internacional de filosofía política*, (17), 31-46.
18. Francesconi, E., y Passerini, A. (2007). Automatic classification of provisions in legislative texts. *Artificial Intelligence and Law*, 15(1), 1-17.
19. Franssen, M. (2007). Automated Detection of Norm Sentences in Laws. In *Twente Student Conference on IT*.
20. Galgani, F., y Hoffmann, A. (2011). Lexa: Towards automatic legal citation classification. In *AI 2010: Advances in Artificial Intelligence* (pp. 445-454). Springer Berlin Heidelberg.
21. García, E. (2007). *Inteligencia artificial aplicada al derecho* (No. 104). Universidad Nacional Autónoma de México.
22. Gómez, L. C. (2004). *Teoría general del proceso*. Oxford University Press.
23. Gutiérrez, C. S., y Margadant S. G. F. (2004). *Sistemas jurídicos contemporáneos*. Porrúa.
24. Han, J., Kamber, M., y Pei, J. (2006). *Data mining: concepts and techniques: concepts and techniques*. Elsevier.
25. Hart, H. L. A. (2012). *The concept of law*. Oxford University Press.
26. Hiemstra, D. (2009). Information retrieval models. *Information Retrieval: searching in the 21st Century*, 2-19.

27. Houy, C., Niesen, T., Fettke, P., & Loos, P. (2013, July). Towards automated identification and analysis of argumentation structures in the decision corpus of the German Federal Constitutional Court. In *Digital Ecosystems and Technologies (DEST), 2013 7th IEEE International Conference on* (pp. 72-77). IEEE.
28. Jerome, F. (1991). *Derecho e Incertidumbre*. Fontamara, México.
29. Kennedy, D. (2009). *A Critique of Adjudication*. Harvard University Press.
30. Kim, M. Y., Xu, Y., y Goebel, R. (2013). Summarization of legal texts with high cohesion and automatic compression rate. In *New Frontiers in Artificial Intelligence* (pp. 190-204). Springer Berlin Heidelberg.
31. Li, T., Balke, T., De Vos, M., Satoh, K., y Padget, J. (2013). Detecting conflicts in legal systems. In *New Frontiers in Artificial Intelligence* (pp. 174-189). Springer Berlin Heidelberg.
32. López de Oñate, F. (1953). *La certeza del derecho*. Buenos Aires: EJEA.
33. Luhmann, N. (2005). *El derecho de la sociedad*. Universidad Iberoamericana.
34. Luhmann, N. (2009). *La política como sistema. Lecciones publicadas por Javier Torres Nafarrate*. Universidad Iberoamericana.
35. Maat, E., y Winkels R. (2008). Automatic classification of sentences in dutch laws. In *Legal Knowledge and Information Systems: JURIX 2008: the Twentieth First Annual Conference* (Vol. 21, p. 207). IOS Press.
36. Makri, S., Blandford, A., y Cox, A. L. (2008). Investigating the information-seeking behaviour of academic lawyers: From Ellis's model to design. *Information Processing & Management*, 44(2), 613-634.
37. Manning, C. D., y Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT press.
38. Mochales-Palau, R., y Moens, M. (2007). Study on sentence relations in the automatic detection of argumentation in legal cases. *Frontiers in Artificial Intelligence and Applications*, 165, 89.

39. Monroy, A. L., Calvo, H., Gelbukh, A., y Pacheco, G. G. (2013). Link analysis for representing and retrieving legal information. In *Computational Linguistics and Intelligent Text Processing* (pp. 380-393). Springer Berlin Heidelberg.
40. Nieto, A., y Gordillo, A. (2003). *Las limitaciones del conocimiento jurídico*. Trotta.
41. Raz, J., y Salmorán, R. T. (1986). *El concepto de sistema jurídico: una introducción a la teoría del sistema jurídico*. Universidad Nacional Autónoma de México.
42. Ríos Estavillo, J. J. (1997). Derecho e informática en México. *México: Instituto de Investigaciones Jurídicas (Universidad Nacional Autónoma de México)*.
43. Robles, G. (2008). *Comunicación, Lenguaje y Derecho*. Fontamara.
44. Norvig, P., y Rusell, S. (2004). Inteligencia Artificial: un enfoque moderno. *Phh Pretice Hall, Mexico*.
45. Sabahi, F., y Akbarzadeh-T, M. R. (2014). Introducing validity in fuzzy probability for judicial decision-making. *International Journal of Approximate Reasoning*, 55(6), 1383-1403.
46. Sergeant, A. (2013). Automatic argumentation extraction. In *The semantic web: Semantics and big data* (pp. 656-660). Springer Berlin Heidelberg.
47. Siches, L. R. (1955) ¿Oficio noble o diabólico?: las antinomias de la profesión jurídica. *En Revista de la Facultad de Derecho de México, Tomo V, p. 783-818*. Universidad Nacional Autónoma de México.
48. Sidorov, G. (2013). *Construcción no lineal de ngramas en la lingüística computacional: n-gramas sintácticos, filtrados y generalizados*. Sociedad Mexicana de Inteligencia Artificial.
49. Sidorov, G. (2013). *Modelos formales en la lingüística computacional*. Departament de Filologia Francesa i Romànica, Universitat Autònoma de Barcelona.
50. Soria, C., Bartolini, R., Lenci, A., Montemagni, S., y Pirrelli, V. (2007). Automatic extraction of semantics in law documents. In *Proceedings of the V Legislative XML Workshop* (pp. 253-266).
51. Sunstein, C. R. (2001). Of Artificial Intelligence and Legal Reasoning. *University of Chicago Law School Roundtable, 2001, vol. 8*.

52. Timmer, S. T., Meyer, J. J. C., Prakken, H., Renooij, S., y Verheij, B. (2015). A structure-guided approach to capturing Bayesian reasoning about legal evidence in argumentation. In *Proceedings of the 15th International Conference on AI and Law*.
53. Tiscornia, D. (2006, June). The LOIS project: Lexical ontologies for legal information sharing. In *Proceedings of the V Legislative XML Workshop* (pp. 189-204).
54. Wilkinson, M. A. (2001). Information sources used by lawyers in problem solving: An empirical exploration. *Library & Information Science Research*, 23(3), 257-276.
55. WINKELS, R., BOER, A., y PLANTEVIN, I. (2013, November). Creating Context Networks in Dutch Legislation. In *Legal Knowledge and Information Systems: JURIX 2013: The Twenty-Sixth Annual Conference* (Vol. 259, p. 155). IOS Press.