



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

**Reconocimiento de patrones
para la detección de
entidades nombradas**

TESIS

Que para obtener el título de
Ingeniero en Computación

P R E S E N T A

Guillermo López Velarde González

DIRECTOR DE TESIS

Dr. Gerardo Eugenio Sierra Martínez



Ciudad Universitaria, Cd. Mx., 2018



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradecimientos

El trabajo realizado en esta tesis, así como el realizado durante toda la carrera, no habría sido posible sin el apoyo incondicional de mi familia, mi mamá, mi papá, y mi hermano, con quienes estoy eternamente agradecido.

También agradezco a la Universidad Nacional Autónoma de México, a la Facultad de Ingeniería, al Instituto de Ingeniería, y particularmente al Grupo de Ingeniería Lingüística, cuyos miembros me ayudaron en el aprendizaje del área a la que aspiro dedicarme. Estoy en deuda y agradecido con todas estas instituciones.

Agradezco a todos los amigos que me acompañaron durante esta etapa de mi vida, a los que les hablo aún, y a los que ya no. Agradezco especialmente a mi amiga Victoria, quien además de brindarme su apoyo, me brindó asesoría para llevar este trabajo a buen puerto.

Estoy también en deuda con todos mis profesores, cuya dedicación y conocimiento, admiro, aprecio y aspiro a alcanzar.

Finalmente, agradezco al Consejo Nacional de Ciencia y Tecnología (CONACYT), y al Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT), con cuyos recursos, en momentos distintos, pude desarrollar plenamente el trabajo descrito en esta tesis.

1. Introducción	1
a) Objetivo	2
b) Hipótesis	2
c) Estructura de la tesis	3
2. Antecedentes	5
a) Reconocimiento de patrones	5
b) Extracción de información	6
c) Detección de entidades	8
d) Extracción de entidades nombradas	9
e) Extracción de eventos	11
f) Documentos legales	12
3. Estado del Arte	13
4. Marco Teórico	16
a) Corpus lingüísticos	16
I. ¿Qué es un corpus lingüístico?	16
II. ¿Qué características debe tener un corpus lingüístico?	16
III. Corpus anotados	18
b) Procesamiento de Lenguaje Natural	19
I. ¿Qué es el lenguaje natural?	19
II. ¿Qué es el lenguaje formal?	20
III. ¿Qué es el lenguaje legal?	20
IV. Lenguajes formales y expresiones regulares	21
c) Reconocimiento de patrones lingüísticos	24
I. Nivel fonético y fonológico	25
II. Nivel morfológico y sintáctico	25
III. Nivel léxico y semántico	26
IV. Nivel pragmático	27
d) Reconocimiento de entidades nombradas	28
I. Sistemas de reglas para reconocer entidades nombradas en texto legal.	29
e) Extracción de eventos y actuantes	30
f) Bases de datos no estructuradas	31
g) Lenguaje de programación Python	32

h)	Medidas de desempeño	33
5.	Metodología	36
a)	Detección de patrones lingüísticos	36
I.	Recopilación y descripción del corpus utilizado	37
II.	Patrones lingüísticos para entidades nombradas	38
i.	Patrones para entidades con alias	38
a.	Regla 0	39
1)	Buscar mayúsculas en orden	40
2)	Buscar minúsculas en orden	41
3)	Buscar mayúsculas en desorden	42
4)	Buscar con expresión regular	42
b.	Regla 1	42
1)	artículo + entidad + candidato a alias	43
2)	entidad + candidato a alias	44
c.	Regla 2	44
d.	Regla 3	44
e.	Regla 4	45
ii.	Patrones para entidades sin alias	45
III.	Patrones lingüísticos para detección de eventos	47
i.	Patrón para ¿Quién?	48
ii.	Patrón para ¿Qué?	48
iii.	Patrón para ¿Dónde?	48
iv.	Patrón para ¿A quién?	49
b)	Implementación de los patrones	49
I.	Implementación de patrones para entidades con alias	52
i.	Función MainNERalias()	53
ii.	Función aplicarReglasAlias(candidato, parrafo, fname, indiceOcurriencia)	55
iii.	Función regla1(candidato, parrafo)	57
iv.	Función resolverSiglas(siglas,parrafo,siglasOriginales)	59
v.	Función buscarArticulo(articulo,candidato,parrafo)	61
vi.	Función regla2(candidato,parrafo)	63
vii.	Función regla3(candidato,parrafo)	64

viii.	Función regla4(candidato,parrafo)	64
II.	Implementación de patrones para entidades sin alias	65
i.	Función MainNER()	66
ii.	Función buscarEntidades(texto,fname)	68
iii.	Función ReglasNER(candidato,fname)	70
III.	Implementación de la base de datos	70
IV.	Implementación de patrones para reconocimiento de eventos	72
i.	Funciones	72
V.	Consideraciones del software	76
c)	Experimentos	77
I.	Entidades con alias	77
i.	Ejemplo 1	77
ii.	Ejemplo 2	80
iii.	Ejemplo 3	80
iv.	Ejemplo 4	81
II.	Entidades sin alias	82
i.	Ejemplo 1	82
ii.	Ejemplo 1	84
iii.	Ejemplo 3	85
III.	Eventos	87
i.	Ejemplo 1	87
ii.	Ejemplo 2	88
iii.	Ejemplo 3	88
iv.	Ejemplo 4	89
v.	Ejemplo 5	90
6.	Resultados	92
a)	Documento del 2015	94
b)	Documento del 2016	96
c)	Documento del 2017	98
d)	Documento del 2018	99
e)	Documento ajeno al dominio	100
7.	Conclusiones	102

Bibliografía	103
Apéndices	105
Tablas de resultados de entidades nombradas	105
Documento 2015	105
Documento 2016	114
Documento 2017	125
Documento 2018	144
Glosario de términos	158

1. Introducción

Hoy en día, los abogados, servidores públicos, empresarios, y demás gente que quiere relacionarse de alguna forma con el Estado, inevitablemente debe recurrir al uso de documentos legales, tales como contratos, amparos, textos. Estos documentos, como bien sabemos, están escritos en un lenguaje casi inentendible para los que no conocemos del área y, generalmente, son muy extensos y aburridos.

La tarea de interactuar con documentos legales gasta mucho tiempo y, por lo tanto, supone un gasto importante para la entidad que sea responsable de la persona realizando el trabajo. Generalmente, esta labor puede ser considerada burocrática, pues en muchas ocasiones se requiere de documentos legales para avanzar en un proceso de mayor importancia.

Los gastos burocráticos en el gobierno aportan al gasto corriente, evitando que se utilicen los recursos en otras áreas donde puede hacer más falta. Como ejemplo de esto, en la Comisión Nacional del Deporte, según el periódico *El Economista*¹, se gasta más en burocracia, que en gastos propiamente deportivos. Así mismo, según *Forbes*², en el año 2017 se consideró ahorrar en gasto corriente (burocrático) entre 250,000 y 300,000 millones de pesos, siendo una de las formas, la eliminación de plazas que realizan trabajos repetitivos.

Para este fin la tecnología es útil, pues es posible reemplazar trabajos humanos sencillos, como los burocráticos, por computadoras, ayudando así al aligeramiento del gasto corriente del gobierno federal. Esto no es necesariamente malo para el empleado, o para la institución, pues la persona puede ser empleada para realizar un trabajo más complejo que aún no sea automatizado, mejorando la productividad propia, y de la institución.

Particularmente, la Cámara de la Industria Electrónica, de Telecomunicaciones y Tecnologías de la Información (Canieti) ha propuesto este año³ la creación de una secretaría especializada en tecnologías de la información y la comunicación, lo cual, a pesar de sonar a más gasto burocrático, realmente significa formalizar el área que se encarga de automatizar y digitalizar tareas del gobierno federal, de múltiples secretarías, que al final se traducen en ahorro en el gasto corriente.

¹ <https://www.economista.com.mx/deportes/Gastos-burocraticos-superaran-los-de-alto-rendimiento-20131003-0105.html>

² <https://www.forbes.com.mx/reducir-la-burocracia-opcion-viable-ajuste-al-gasto/>

³ <http://www.elfinanciero.com.mx/empresas/sector-telecomm-pide-la-creacion-de-una-secretaria-de-tic-s>

La organización que ha realizado esta propuesta está íntimamente ligada con el Instituto Federal de Telecomunicaciones, señalando, indirectamente, que existen procesos en ese Instituto que pueden ser absorbidos por la nueva secretaría, y que actualmente ella maneja. Por ejemplo, el desarrollo de interfaces web para buscar documentos en línea⁴ es una labor realizada desde dicha institución, y su existencia supone un gasto burocrático, pues es necesaria la revisión de los documentos de forma manual, así como su clasificación y digitalización, para después permitir su búsqueda.

Es en esta organización en particular, y por el conocimiento de este ejemplo concreto, que en este trabajo hemos decidido trabajar con documentos del Instituto Federal de Telecomunicaciones. De forma general, buscaremos facilitar la interacción entre la persona y el documento legal, ofreciendo automáticamente información relevante sobre el documento, sin que haya que leerlo completo.

Esto, para el IFT, se puede traducir en un aligeramiento en el trabajo de las personas que realizan la labor de clasificación de documentos, agilizando así el proceso por el cual se hacen disponibles los documentos en línea, y aportando al ahorro en el gasto corriente del gobierno federal.

A continuación, se establece un objetivo específico tanto en la técnica, como en el alcance de la solución al problema. También, se establece una hipótesis sobre cómo se desempeñará la solución realizada en el mundo real. Finalmente, explicaremos la estructura de la tesis para que el lector tenga claro cómo se lleva a cabo la solución al problema planteado.

a) Objetivo

Diseñar e implementar un sistema basado en reglas que utilice reconocimiento de patrones para identificar entidades nombradas en documentos legales provenientes del Instituto Federal de Telecomunicaciones, y después implementar un sistema que haga uso de dichas entidades nombradas para obtener información relevante sobre eventos.

b) Hipótesis

Es posible utilizar reconocimiento de patrones para obtener entidades nombradas de documentos legales. Además, dicho método permite obtener mejores medidas de desempeño que una herramienta que detecte entidades nombradas de propósito general.

⁴ <http://www.ift.org.mx/recursos-de-informacion/buscador>

c) Estructura de la tesis

Aquí explicaremos un poco para qué sirve la información contenida en cada una de las secciones de la tesis, para ofrecer al lector un mapa de este trabajo, y pueda moverse a través de él como su conveniencia indique.

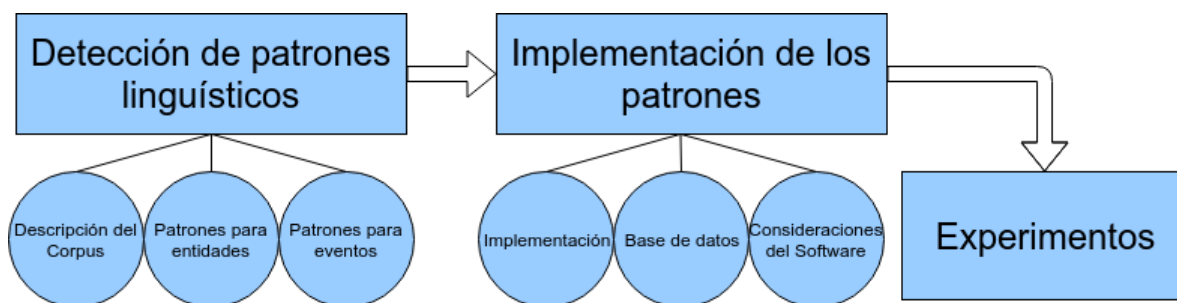
No perdamos de vista nunca el objetivo de la tesis, el cual es, de forma general, extraer información relevante de documentos legales, y proveerla al usuario, con el fin de que él mismo no tenga que revisar el documento entero para tener un esbozo inicial de lo que trata el mismo.

Primero, en la sección “Antecedentes” hablaremos de forma general sobre las áreas del conocimiento que se encargan de tratar con este tipo de problemas, como el reconocimiento de patrones, la extracción de información, extracción de entidades, etc.

En la sección “Estado del arte” analizamos los trabajos previos que atienden este mismo problema, para decidir así el mejor acercamiento al mismo. En esta sección se encontrará también la justificación de las técnicas utilizadas.

Después, en “Marco teórico” hablaremos específicamente sobre algunos de los temas vistos en los antecedentes, y que son de vital importancia para entender la solución propuesta y el programa realizado. Es decir, se tratará toda la información específica necesaria para la comprensión total de esta tesis. Cabe mencionar, que el contenido expuesto es tan profundo como se ha considerado necesario, y en la explicación del mismo no se busca ligarlo directamente con los documentos legales, a menos que se considere prudente.

En “Metodología” describiremos paso a paso, el proceso con el cual se ha realizado la solución al problema planteado, utilizando el marco teórico. A grandes rasgos, la metodología se muestra en el siguiente diagrama:



1.1 Diagrama de la metodología implementada

Cada rectángulo es un paso a seguir, primero se detectan los patrones para detectar entidades y eventos, después se implementan en Python, y finalmente se hacen experimentos.

Los círculos son especificaciones dentro de cada paso

Cada rectángulo representa un subtema dentro de esta sección, y los círculos describen un poco el contenido de dichas secciones. En cada rectángulo, se hará uso de los términos y los temas especificados en el marco teórico. Es decir, en esta sección se encuentra cómo hemos resuelto el problema planteado.

La sección de “Resultados” contiene las medidas de desempeño de la solución propuesta, así como análisis de los mismos resultados.

En “Conclusiones” analizaremos los problemas que aún se tienen, y propondremos trabajo a futuro para seguir mejorando la solución propuesta, además de validar la hipótesis.

Finalmente, dentro de los apéndices se agrega un breve glosario con definiciones de conceptos básicos propios del área de Procesamiento de Lenguaje Natural, que pueden servir como referencia rápida al lector.

A continuación empezaremos con los Antecedentes, hablando sobre qué entenderemos por reconocimiento de patrones, qué es la extracción de información, el reconocimiento de entidades nombradas, y algunas técnicas generales sobre cómo resolver el problema planteado.

2. Antecedentes

A continuación hablaremos de forma general sobre las áreas del conocimiento que aportan información importante para dar solución a la problemática de los textos legales. Empezaremos esbozando ¿qué es un patrón?, ya que podemos intuir que los documentos legales tienen muchos. Después hablaremos más específicamente sobre extracción de información, y luego sobre algunas técnicas que existen en el área para extraer dicha información, como los sistemas de reglas, sistemas estadísticos, etc. También hablaremos sobre ¿qué es una entidad nombrada?, ¿qué es el lenguaje natural?, y sobre algunas particularidades del lenguaje que aparecen en los textos legales, al cual denominaremos “lenguaje legal”. Empezamos entonces con los patrones.

a) Reconocimiento de patrones

En los documentos legales se puede intuir que existen patrones, pues muchas veces vemos que un documento se parece mucho a otro, aunque no podamos establecer propiamente las características específicas que lo hacen así. Entonces, ¿Qué es un patrón?

Se intuye que el proceso por el cual los humanos tomamos decisiones tiene que ver con el reconocimiento de patrones (Fukunaga 2013). Un patrón son los elementos que influyen en una decisión, y definirán el acto siguiente a realizar, o la conclusión a obtener. Por ejemplo, dado un contexto donde tenemos imágenes con personas, podemos reconocer los patrones que definen una cara, y así poder identificarla. Así mismo, existen patrones en un documento legal que nos permiten diferenciarlo de un texto común, un cuento, una novela, o un poema.

El reconocimiento de patrones, al estar inmiscuido en las decisiones humanas, tiene un dominio amplio donde puede actuar. Dicha amplitud de dominio nos presenta con el problema de tareas con complejidad muy alta, para los cuales no tenemos claro cómo decidimos, en general, en aquellas tareas que requieren de ‘intuición’ o ‘sensibilidad’. Por ejemplo, en el lenguaje legal, además de preparación en el área, se requiere cierto grado de intuición para reconocer información que para una persona no preparada es difícil encontrar.

Así, el objetivo del reconocimiento de patrones es clarificar estas tareas complejas que realizamos los humanos y automatizarlas utilizando computadoras (Fukunaga 2013). Cabe mencionar que el reconocimiento de patrones generalmente se usa para problemas estadísticos. Sin embargo, en esta tesis entenderemos el reconocimiento de patrones como el acto de detectar los elementos que nos permitirán tomar una decisión precisa respecto a un cúmulo de información.

Particularmente, el acto de reconocer información en un texto, se puede denominar como ‘extracción de información’. A continuación hablaremos un poco sobre esta área.

b) Extracción de información

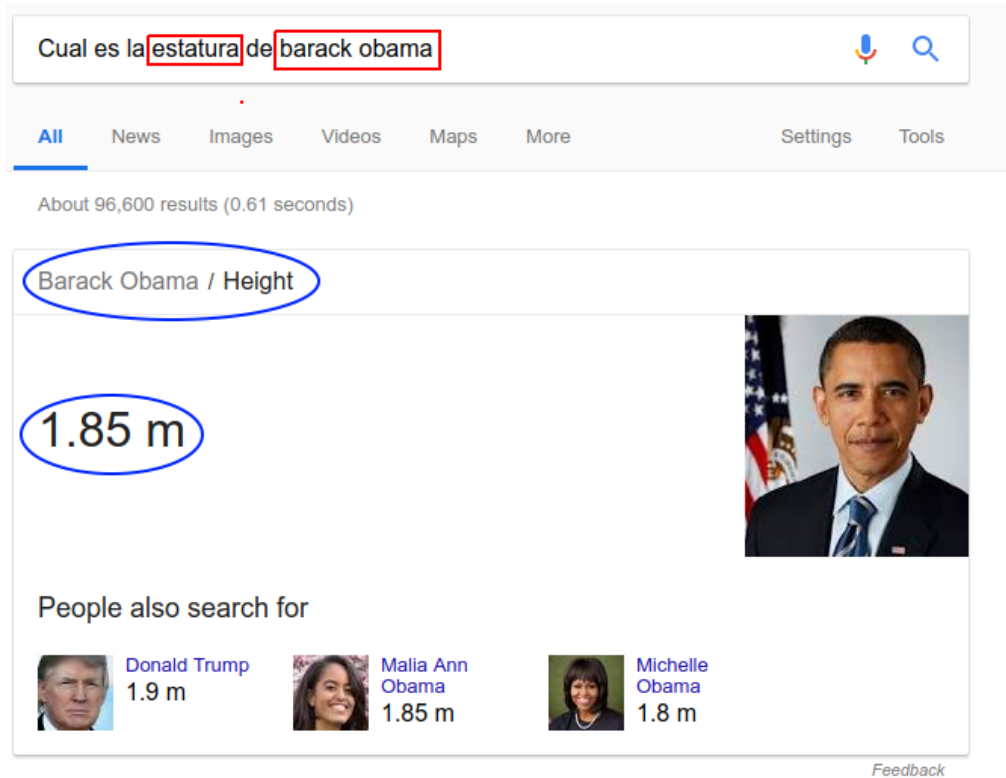
Una vez que sabemos que los documentos legales contienen patrones, podemos intuir que dichos patrones pueden decirnos qué cosa en el documento es información, y qué cosa es, por así decirlo, algo innecesario. Es parte del objetivo de esta tesis, extraer cierto tipo de información de estos textos de manera automática.

Tenemos que partir hablando sobre los documentos en sí. Estos documentos utilizan una variante del lenguaje que utilizamos día a día, una variante especializada, con características propias que la hacen diferente de la forma de hablar común. A esta variante le llamaremos 'lenguaje legal'.

El lenguaje legal es un medio para transmitir información que, a pesar de tener estructura, no es una estructura definida, estandarizada, y mucho menos clara. Esto hace que su procesamiento utilizando computadoras sea complicado, y es la principal razón por la cual el trabajo con estos documentos lo tiene que realizar un ser humano.

El área formal que se encarga de extraer la información que necesitamos, se denomina minería de textos, y se define como la encargada de obtener información de un texto no estructurado (Aguilar, C. A., Sierra Martínez, G. E. 2017), es decir, escrito en lenguaje natural . Nosotros utilizaremos la minería de textos, para poder extraer de los documentos la información necesaria, y después introducirla en una estructura computable.

No sólo en el lenguaje legal es necesario extraer información, también en el lenguaje común, o natural, es necesario realizar esta tarea. Por ejemplo, al realizar una búsqueda en Google, el motor de búsqueda debe establecer qué cosa estás buscando, a partir del texto que le has introducido.



2.1 Búsqueda en Google

Ejemplo de la necesidad de extraer información de texto escrito en lenguaje natural

En el ejemplo mostrado en la imagen, se observa la necesidad de detectar a Barack Obama como la persona que estás buscando, y la estatura como un atributo de esa persona para poder encontrar la información requerida en una base de datos. Así, Google es una herramienta familiar que utiliza el área de Extracción de Información para resolver un problema.

De la misma manera, nosotros queremos extraer información específica de un documento legal. Es como si escribiéramos todo un documento legal en Google, y Google nos contestara con la información más relevante de dicho texto.

Ahora, ¿qué información deseamos extraer exactamente? En el ejemplo del buscador de Google, se tuvo que identificar a una persona y a su estatura, porque son los elementos de interés que había en el texto “Cual es la estatura de barack obama”. Así, los elementos de interés que existirán en un documento legal, son diferentes, y les llamaremos entidades.

A continuación hablaremos más específicamente sobre qué es una entidad, y cuáles son algunos ejemplos de entidades dentro de los documentos legales.

c) Detección de entidades

En el documento legal puede haber muchos elementos de interés que pueden ser extraídos. Por ejemplo, nombres de personas, nombres de empresas, tipo de documento, leyes, artículos, referencias a empresas, eventos que se describen en el documento, nombres de instituciones, comisiones, grupos de personas, etc. Estos elementos de interés se denominan entidades, y el área que se encarga de encontrarlos y extraerlos, se llama Detección de Entidades Nombradas (por sus siglas en inglés, NER). Esta área es subárea del área de Extracción de Información.

El área de NER tiene como finalidad extraer a lo que hemos denominado como 'entidad' e introducirlo en una estructura computable. De forma general, se considera entidad a personas, lugares y organizaciones (Galicia-Haro, S. N., Gelbukh, A., & Bolshakov, I. A 2004), aunque la definición de entidad, como hemos dicho anteriormente, puede adaptarse dependiendo del documento y de la información específica que buscamos extraer.

Así, todos los textos tienen necesidad de nombrar entidades, y al hacerlo se establece información sobre ellas que puede ser útil extraer y estructurar. En este caso, nuestro objetivo será extraer entidades de los documentos legales y estructurarlas de forma que sea más fácil saber qué elementos existen en dicho documento, o quienes participan de él, sin tener que leerlo.

Particularmente, buscamos extraer nombres de personas, organizaciones (instituciones, comisiones, empresas, etc.), leyes, documentos (oficios, expedientes y acuerdos) y lugares. Pues considero que un buen resumen de qué trata el documento puede verse conociendo las entidades que en él participan, pues generalmente es lo que varía de documento a documento.

Es importante que no sólo extraigamos esta información, sino que también le demos una estructura, y la hagamos disponible a la computadora, con el fin de que pueda ser utilizada en otro sistema que pueda llegar a ser más útil y completo que este acercamiento.

En nuestro caso, para probar que efectivamente extraer las entidades mencionadas es útil, y se ha hecho de manera correcta, crearemos un sistema que nos pueda decir, con base en el texto y las entidades que en él se detectaron, qué eventos sucedieron, y cuáles son las entidades que participan en dicho evento.

Así, a continuación hablaremos sobre cómo extraer entidades nombradas y las técnicas generales que existen para este fin, después estableceremos qué es un evento, y qué es lo que nos interesará obtener de él.

d) Extracción de entidades nombradas

Para realizar reconocimiento de entidades nombradas se han seguido varios métodos, algunos de los cuales son descritos por Mansouri et al. (2008). En esta sección los retomaremos y explicaremos brevemente. En general, pueden clasificarse en tres tipos:

- 1) Métodos de aprendizaje de máquina.
- 2) Sistemas de reglas.
- 3) Métodos híbridos.

Los métodos de identificación de entidades nombradas que utilizan aprendizaje de máquina convierten el problema de detección a un problema de clasificación (Mansouri et al. 2008). Es decir, del texto se busca clasificar las palabras en diferentes categorías, donde la categoría más relevante es la de entidad, y después, pueden existir subcategorías de entidad que sirvan para hacer una clasificación más específica, por ejemplo 'ley', 'nombre de persona', 'nombre de institución', etc.

El aprendizaje de máquina se puede dividir en dos acercamientos:

- 1) Supervisado
- 2) No supervisado

El supervisado necesita de ejemplos de entidades suficientes para enseñarle a la computadora qué palabras son entidades, y qué palabras no lo son. Así, la computadora es capaz de realizar nuevas clasificaciones correctamente. Los ejemplos son documentos donde seres humanos han indicado manualmente las entidades nombradas, y a este proceso se le denomina 'etiquetado'.

El proceso de etiquetado es costoso en horas hombre, pues la cantidad de ejemplos que son necesarios podría ser muy alta si el fenómeno que se quiere caracterizar aparece pocas veces en los textos. Sin embargo, de lograrse pagar este costo, el método ofrece alta portabilidad a otros dominios (temas). Es decir, si tienes suficientes ejemplos de qué es una entidad en un documento legal, es probable que también funcione, en alguna medida, para documentos que no tengan lenguaje legal.

El enfoque no supervisado no necesita de ejemplos. Su eficacia recae en si las entidades nombradas tienen o no un patrón subyacente capaz de ser encontrado y agrupado por la computadora, usando algún algoritmo de aprendizaje. Estos patrones varían de dominio a dominio, por lo que realmente los sistemas no supervisados no son muy recurridos para el reconocimiento de entidades nombradas, aunque hay algunos intentos mezclando aspectos de los sistemas

supervisados con sistemas no supervisados para resolver el problema, como el utilizado por Kim, Kang, & Choi (2002).

El método de identificación usando sistemas de reglas se basa en definir un dominio concreto, y para él, identificar una serie de patrones gramaticales, sintácticos, léxicos y semánticos que permitan encontrar las entidades y su clasificación. (Mansouri et al. 2008).

Son sistemas costosos pues un humano tiene que realizar el análisis para encontrar los patrones. Además, sólo funcionan para dominios específicos, pues un patrón que puede utilizarse para encontrar entidades en un tema, podría no funcionar para otro tema. Además, hacer cambios al programa con fines de mantenimiento también es costoso.

A pesar de todo esto, los sistemas de reglas son capaces de encontrar entidades que los sistemas supervisados tienen problemas encontrando, pues la especificidad de las reglas lo permite, y muchas veces los fenómenos específicos no se repiten lo suficiente como para que un sistema estadístico pueda aprenderlos.

Un sistema híbrido utiliza una mezcla entre los sistemas de reglas y los sistemas de aprendizaje de máquina, que trata de maximizar las virtudes de cada sistema. La mayor desventaja que hereda es la portabilidad limitada por el hecho de utilizar reglas.

Las ventajas y desventajas de los tres métodos descritos por Mansouri et al. (2008) y Oudah, M., & Shaalan, K. (2017), a mi interpretación, pueden resumirse en la siguiente tabla:

		Costo en tiempo de implementación	Tiempo de mantenimiento	Precisión de los resultados	Portabilidad de dominio.
Aprendizaje	Supervisado	Alto	Bajo	Medio	Alto
	No supervisado	Bajo	Bajo	Bajo	Alto
Reglas		Alto	Alto	Alto	Bajo
Híbrido		Alto	Medio	Alto	Bajo

2.2 Tabla comparativa entre métodos

Se describen las características de cada método, medidos con “Bajo”, “Medio” y “Alto”

Así, por la alta precisión de los resultados, en esta tesis se utiliza un sistema de reglas, pero sólo para un dominio bien delimitado. La delimitación correcta del dominio contrarresta la desventaja de la baja portabilidad, y maximiza la precisión de los resultados por la especificidad de las reglas.

Ahora, una vez que sabemos el método con el cual trabajaremos para extraer las entidades mencionadas, hablaremos sobre los eventos que vamos a extraer, y de cómo se relacionan estos eventos con las entidades que hemos extraído.

Este sistema de detección de eventos, distinto al de detección de entidades nombradas, busca probar que efectivamente la información obtenida por el primer sistema es computable y útil para realizar otra tarea con un potencial más significativo.

e) Extracción de eventos

Generalmente en un documento legal se habla de hechos que han ocurrido en alguna fecha, y dichos eventos refieren a una entidad. Por ejemplo en el fragmento:

“Con fecha 14 de mayo de 2015 la Unidad de Competencia Económica a través de la Dirección General de Concentraciones y Concesiones emitió el oficio IFT mediante el cual remite la opinión correspondiente a la Solicitud de Prórroga.”

Podemos identificar la fecha y el evento marcado por el verbo “emitió”. Sería útil entonces, que antes de leer el documento en su totalidad, supiéramos que en el documento, el 14 de mayo del 2015 la Unidad de Competencia Económica emitió un oficio. Así, extraemos del documento un esbozo que nos permite tener un primer acercamiento al documento sin tener que leerlo en su totalidad.

Observemos que de ese extracto específico podemos responder algunas preguntas: ¿Quién? la Unidad de Competencia Económica, ¿Qué? un oficio, ¿Cuándo? el 14 de mayo del 2015. Así, el sistema de detección de eventos buscará extraer el verbo que define al evento, y la respuesta a preguntas que nos den más información sobre el evento.

A continuación hablaremos sobre el dominio específico de los documentos legales de los cuales vamos a extraer las entidades. No es lo mismo extraer entidades de un amparo, que un contrato, o un acta constitutiva, etc. Es importante delimitar el tipo de documentos legales con los que trabajaremos, para poder asegurarnos de que el sistema de reglas funcione de manera óptima.

f) Documentos legales

Existen en el lenguaje legal usos particulares de vocabulario, estructuras sintácticas y significados que lo diferencian del lenguaje natural ordinario, sin separarlo del mismo (Mattila, H. E. 2016). La existencia de estos patrones y características únicas hacen del lenguaje legal uno más estructurado, del que es más fácil extraer información utilizando sistemas de reglas.

También es importante acotar el lenguaje a español mexicano, pues los patrones que se presentan en el ámbito legal de México varían con lo que puede encontrarse en el ámbito legal de otro país con otra lengua, o incluso otro país con la misma lengua como España (Mattila, H. E. 2016).

Así, el dominio elegido es el ámbito legal utilizando español mexicano para documentos del Instituto Federal de Telecomunicaciones (IFT), el cual es un órgano autónomo operante en México, cuya finalidad es desarrollar eficientemente las telecomunicaciones en México con apego a la Constitución.⁵

Se ha elegido esta institución, porque actualmente el IFT ofrece un buscador de documentos⁶ en su sitio web, el cual podría ser mejorado y extendido, si se tuviera más información sobre las entidades que aparecen en cada documento. Además, dicho buscador ya sirve en sí mismo como fuente de documentos para poder realizar el análisis y el sistema.

Además, el buscador pone en evidencia que existen personas que necesitan la información que está dentro de esos documentos. Así, esta herramienta puede facilitarles el trabajo a los abogados, servidores públicos, y empleados que buscan esta información.

Concluimos así la sección de ‘Antecedentes’, conociendo qué es el reconocimiento de patrones, qué es la extracción de información, qué es la extracción de entidades nombradas y sus principales técnicas; También sabemos que buscaremos extraer eventos, y responder algunas preguntas sencillas sobre los mismos. Finalmente establecimos el tipo de documentos con los que trabajaremos.

A continuación, en la sección “Estado del Arte” hablaremos sobre los trabajos que se han realizado para solucionar problemas parecidos al planteado en esta tesis, y los analizaremos con el fin de establecer con mayor certeza las técnicas y particularidades que habremos de tener en cuenta en la resolución de la problemática planteada en esta tesis.

⁵ <http://www.ift.org.mx/que-es-el-ift/que-es-el-ift>

⁶ <http://apps.ift.org.mx/cumplimientoStp/secured/adminficum.faces>

3. Estado del Arte

Para resolver el problema de extraer entidades nombradas y eventos de documentos legales, no pudimos encontrar información en español. En cambio, hay algunos acercamientos a este problema para documentos legales en inglés que pueden arrojar algo de luz acerca de las técnicas y soluciones que puede tener este problema. Particularmente hablaremos de tres trabajos, el primero enfocado a extraer entidades nombradas, el segundo enfocado a encontrar particularidades del lenguaje legal, y finalmente un trabajo sobre extracción de eventos.

Dozier et al. (2010) describen en su trabajo un sistema que utiliza un sistema de reglas, un modelo estadístico, y diccionarios para reconocer entidades nombradas en documentos legales escritos en inglés. Es decir, han implementado un sistema híbrido de reconocimiento.

Dividieron su sistema en 5 partes denominadas *Taggers*, cada una dedicada a encontrar un tipo de entidad. Las entidades que buscaron en sus documentos fueron Jurisdicción, Corte, Título, Tipo de documento y Juez, obteniendo las siguientes medidas de desempeño:

Tagger	Precision	Recall
Jurisdiction	97	87
Court	90	80
Title	84	80
Doctype	85	80
Judge	98	72

3.1 Medidas de desempeño para NER de 5 entidades.

Tabla que describe, utilizando las medidas de precisión y exhaustividad, el rendimiento del programa implementado por Dozier et al. Sobresale el rendimiento del etiquetador de jurisdicciones.

Tomado de Dozier et al. (2010)

La división de su sistema resulta bastante conveniente, pues cada parte utilizó un método distinto. El etiquetador de jurisdicción usó un sistema de reglas, el de cortes utilizó un diccionario, para títulos se usó un método estadístico, el tipo de documento y el etiquetador de jueces utilizaron sistemas de reglas.

Vemos que el buscador de títulos fue el que tuvo la menor precisión, siendo este el que utilizó un sistema estadístico, mientras que el etiquetador de juez obtuvo la mejor precisión, siendo uno que utilizó un sistema de reglas.

En general, el sistema que funcionó mejor fue el etiquetador de jurisdicciones con la mayor precisión y exhaustividad.

El etiquetador de cortes que utilizó un diccionario obtuvo buena precisión y exhaustividad. Sin embargo, los autores advierten que diccionarios con entradas ambiguas como nombres propios pueden traer muchos falsos positivos, pues no porque se detecte un nombre en el texto, se asegura que dicho nombre tenga el rol específico que queremos encontrar. Para resolver este problema, eliminan la ambigüedad de su diccionario asignando a cada corte, atributos específicos como jurisdicción, división y circuito al que pertenecen.

Con base en la información anterior, puedo confirmar que el mejor acercamiento para detectar entidades nombradas en documentos legales es utilizar un sistema de reglas, como el etiquetador de jueces y jurisdicciones, junto con un diccionario que tenga entradas no ambiguas, como el etiquetador de cortes.

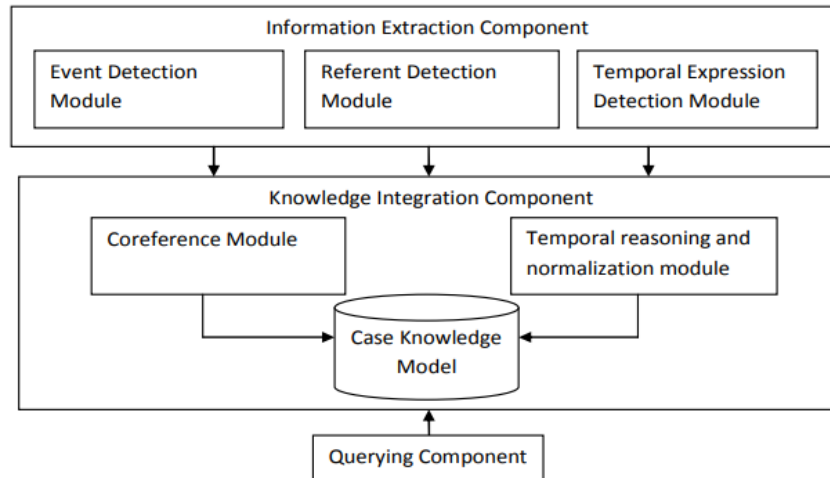
Existe también trabajo que da información sobre las particularidades del lenguaje utilizado en documentos legales. Dichas particularidades luego pueden ser descritas en términos de reglas que pueden ayudar a identificar entidades nombradas o eventos.

Venturi (2010) analiza estas particularidades, señalando que las mismas son un factor que está siendo subestimado en el área. Para encontrarlas, se toma un corpus de lenguaje legal en italiano e inglés, y se compara con un corpus de lenguaje ordinario también en italiano e inglés. Esta comparación permite obtener características propias del lenguaje legal que se comparten entre dos lenguas, para así concluir que son características propias del tema y no de la lengua.

Se encontró que es parte del lenguaje legal la ocurrencia sistemática de frases preposicionales y una baja presencia de verbos en infinitivo.

Lagos et al. (2010, October) describen un sistema para obtener eventos significativos de documentos legales, así como las personas que participan en dichos eventos. Este sistema tiene un módulo de detección de eventos, otro de detección de referencias y finalmente uno sobre detección de expresiones temporales.

Esos tres módulos arrojan su resultado a un módulo que resuelve correferencias, y a otro módulo que normaliza las expresiones temporales, para finalmente integrar toda la información del evento en una base de datos, la cual puede ser accedida utilizando búsquedas de un usuario.



3.2 Estructura de los módulos implementados por Lagos et al.

Se muestran dos módulos, uno que extrae información, y el otro que la integra y la hace disponible en una base de datos a la que un usuario puede acceder haciendo búsquedas.

Tomado de Lagos et al. (2010)

Análogamente, nuestro sistema tendrá un sistema de detección de entidades nombradas, y luego un sistema de detección de eventos, el cual incluye la detección de expresiones temporales como requisito para la detección de dichos eventos.

Una vez conociendo algunos trabajos que se han hecho en esta área, hablaremos sobre los temas que se necesitan conocer para el correcto desarrollo de la metodología. Es decir, en “Marco Teórico” ahondaremos en los temas teóricos que sustentan la solución planteada en esta tesis. Hablaremos de los temas de forma más profunda que en la sección “Antecedentes”, pero sin ligarlos concretamente al ámbito legal, a menos que se considere prudente para dar ejemplos.

4. Marco Teórico

En este capítulo trataremos la teoría que sustenta las aplicaciones realizadas en esta tesis. Hablaremos sobre corpus lingüísticos, procesamiento de lenguaje natural, reconocimiento de patrones lingüísticos, reconocimiento de entidades nombradas, extracción de eventos y actuantes, bases de datos no estructuradas, lenguaje de programación python, y medidas de desempeño, en ese orden. Se han elegido estos temas, pues se ha considerado que son los que se necesitan para la realización de este trabajo.

a) Corpus lingüísticos

En esta sección trataremos sobre corpus lingüísticos, definiendo primero ¿Qué es un corpus lingüístico? y después las características que deben de satisfacer para asegurarnos de poder obtener información útil de ellos. Además, hablaremos un poco sobre la anotación de corpus, y de cómo puede ayudarnos esto a la hora procesarlo computacionalmente.

I. ¿Qué es un corpus lingüístico?

“Un corpus lingüístico consiste en un conjunto de textos de materiales escritos y/o hablados, debidamente recopilados para realizar ciertos análisis lingüísticos” (Sierra 2010).

Con materiales escritos nos referimos desde artículos, revistas y libros, hasta reseñas y textos más pequeños. Como materiales hablados podemos tener grabaciones con audio y video, o sólo con audio. En nuestro caso, el corpus sería el conjunto de documentos legales obtenidos del IFT.

Es importante para recopilar un corpus tener claro el análisis lingüístico que se quiere realizar, pues dependiendo de éste serán las características de los textos a buscar. Por ejemplo, si se quiere hacer un análisis lingüístico sobre las entidades en documentos legales, sería ilógico recolectar textos ajenos al dominio legal.

Además de esta consideración, debemos tener en cuenta una serie de características para recolectar nuestro corpus, y de eso hablaremos a continuación.

II. ¿Qué características debe tener un corpus lingüístico?

El conjunto de textos no sólo está delimitado por el análisis lingüístico que queremos hacer, también hay otras consideraciones. Según Sierra (2017), las consideraciones son: contención de datos reales, representatividad, variedad, equilibrio, selectividad y tamaño, que a continuación

describimos brevemente. Más detalle en estos aspectos puede encontrarse en Sierra (2017), *Capítulo 1.3 Características que debe cumplir un corpus*.

Primero, todos los textos que tomemos deben tener datos reales y no fabricados, esto garantiza que los fenómenos que allí se encuentran se repiten en cierta medida dentro del uso normal del lenguaje.

La representatividad del corpus se refiere a que los textos deben provenir de todos los subdominios que definen al dominio que escogimos al elegir el análisis lingüístico que queremos realizar. Por ejemplo, si se quisiera tener un corpus para analizar algún aspecto del español mexicano en jóvenes de 15 a 20 años, tendríamos que conseguir hablantes de español mexicano con dichas características, que provengan de todo el territorio mexicano. Si tomáramos más hablantes de una zona que de otra, no podríamos generalizar que los fenómenos encontrados son característicos del español mexicano, sino del español mexicano de la zona con más hablantes.

Variedad refiere a escoger las características importantes de nuestro corpus, para que las mismas funcionen como filtro de información más adelante. Por ejemplo, localidad geográfica, información personal, tópico, tipo de texto, fuente, tiempo, etc.

Los corpus deben guardar equilibrio en cuanto a la posición que tienen respecto a la característica que comparten los documentos entre sí. Es decir, se debe evitar tener un corpus tendencioso. Esto asegura que el corpus tenga suficiente variedad en las opiniones y los fenómenos lingüísticos.

La selectividad está ligada a los criterios con los cuales decidimos qué texto incorporamos al corpus y qué texto dejamos fuera. Dichos criterios deben ser coherentes con el fenómeno lingüístico que buscamos investigar.

Finalmente, el tamaño del texto debe ser finito, pues una colección de textos infinita, o con una longitud desconocida por no estar bien involucrado en el proceso de recolección, haría difícil un análisis estadístico preciso de los fenómenos que interesan. Además, lo más importante en el corpus es que los textos en él sean una muestra representativa del dominio elegido, lo cual puede alcanzarse con un corpus grande o uno pequeño, pero siempre finito.

Antes de continuar, es necesario mencionar que de esta colección de textos tenemos que poder extraer información utilizando la computadora. Es decir, los textos deben estar disponibles para ser modificados en un editor de texto, no en físico, pues su análisis se complicaría demasiado. Tenerlos disponibles en la computadora permite hacer anotaciones en el texto que ayudan al análisis lingüístico que queremos realizar. Hablaremos de dichas anotaciones a continuación.

III. Corpus anotados

Una vez que tenemos nuestro corpus recopilado y cumpliendo todas las características necesarias para que sea útil, debemos informatizarlo. Es decir, hacerlo procesable para una computadora con un editor de texto.

Tener nuestro corpus en digital nos abre las puertas para realizar el análisis de información de manera automática y eficiente. Si tenemos los textos físicos, se escanean y se pasan por un programa OCR (Optical Character Recognition) que obtiene de la imagen escaneada, el texto procesable por un editor de textos.

Una vez digitalizado nuestro corpus, podemos hacer anotaciones en él, con el fin de identificar más fácilmente ciertas características que nos interese conocer. Dichas anotaciones en el texto deben cumplir ciertos requerimientos para ser útiles, como inteligibilidad, facilidad de extracción, intercambio, documentación y estandarización. (Sierra, 2017)

```
<html>
  <title>Corpus                                lingüísticos</title>
  <body>
    <p> Los corpus lingüísticos son recopilaciones de textos que comparten ciertas
características lingüísticas las cuales desean ser estudiadas </p>
    <p> Deben de cumplir ciertos requerimientos para ser útiles, por ejemplo: ... </p>
  </body>
</html>
```

4.1 Ejemplo de texto anotado con HTML.

En el ejemplo anterior, tenemos un fragmento de un texto anotado con el lenguaje de anotación HTML que cumple con los requerimientos antes mencionados. Observamos que las notas son inteligibles, pues podemos identificar rápidamente dónde empieza el título y dónde termina.

El texto es fácilmente extraíble con expresiones regulares, pues tenemos la certeza de que al extraer lo que está encerrado entre las etiquetas “title” nos devolverá el título de nuestro texto. De la misma forma, detectar las etiquetas en sí permite intercambiarlas a ellas, y al texto que indican, de forma sencilla.

Finalmente, el lenguaje HTML que se utilizó para anotar el ejemplo es un lenguaje conocido, estandarizado y documentado.

Con este tipo de etiquetas en nuestro texto, podemos anotar cosas que queremos saber de él, y al mismo tiempo, indicarle a la computadora los roles de cada elemento. Esto permite poder diferenciar ciertos fragmentos del texto, del resto del mismo, y así estudiarlos cuantitativa o cualitativamente con ayuda de la computadora.

En nuestro caso específico, nos interesa poder etiquetar las entidades nombradas, eventos, y algunos metadatos de los documentos legales. Para este fin, podría usarse una notación como la de HTML, encerrando las entidades nombradas con una etiqueta "<EN>". Sin embargo, como nuestro fin por el momento es introducir las entidades a una base de datos, podemos dejar esta tarea como parte del trabajo a futuro.

Ya que tenemos nuestro corpus recolectado y listo para ser procesado por la computadora, es momento de aplicarle procesos propios del Procesamiento de Lenguaje Natural, el cual explicaremos en la siguiente sección.

b) Procesamiento de Lenguaje Natural

El procesamiento de lenguaje natural es un conjunto de técnicas y procesos computacionales que trabajan sobre textos escritos en lenguaje natural para, con base en análisis lingüísticos, resolver tareas útiles que requieren de dicho conocimiento lingüístico (Liddy 2001).

Como parte de este conocimiento lingüístico, hablaremos sobre qué es el lenguaje natural, qué son los lenguajes formales, qué es el lenguaje legal, y cómo puede funcionar la teoría de lenguajes formales y autómatas para resolver problemas del lenguaje natural y del lenguaje legal.

I. ¿Qué es el lenguaje natural?

Polanco (2000) define el lenguaje como la capacidad que tenemos los humanos para comunicarnos por medio de lenguas. Es decir, es el lenguaje que practicamos día a día al intercambiar ideas con otros humanos.

Consta de dos características que lo definen: La primera es que no es artificial, es decir, no fue creado por el ser humano estableciendo una teoría y luego un lenguaje, sino es consecuencia de un proceso evolutivo de la misma lengua. La otra característica es que posee semántica, es decir, tiene un significado que se descubre en el mismo acto de hablar, y que no se define previamente.

Así, este lenguaje es diferente del lenguaje formal, el cual explicaremos a continuación.

II. ¿Qué es el lenguaje formal?

Un lenguaje formal es un modelo que describe un conjunto de cadenas finito o infinito, y sirve como acercamiento para entender o generar algunos aspectos del lenguaje natural.

Tiene dos características principales, la primera es que surge desde reglas y modelos, es decir, es artificial. La segunda es que su semántica está dada por la sintaxis o diversos modelos de representación (Polanco 2000).

La diferencia con el lenguaje natural es que el lenguaje formal surge a partir de reglas y el natural a partir de la experiencia. También, el significado en el lenguaje formal está dado únicamente por la sintaxis, o el orden de los componentes, mientras que en el lenguaje natural, la semántica está dada por todo el contexto que rodea al acto de habla. Dicho contexto le permite al lenguaje natural un carácter polisémico, y no hay atributos en la oración o las palabras en sí que nos permitan discernir la semántica correcta de dicha manifestación lingüística. (Jurado Málaga, E. 2008)

	Surgimiento	Semántica
Lenguaje Formal	Artificial o creado.	Dada por la sintaxis, sin ambigüedad
Lenguaje Natural	A partir de experiencia.	Dada por el contexto, con ambigüedad

4.2 Tabla comparativa entre lenguaje natural y lenguaje formal. Describe sus diferencias en términos de su surgimiento, y el origen de su semántica.

A continuación, sabiendo las diferencias entre lenguaje formal y lenguaje natural, explicaremos a qué nos referimos cuando hablamos de lenguaje legal, pues es el lenguaje específico con el que trataremos a lo largo de este trabajo.

III. ¿Qué es el lenguaje legal?

Para efectos de esta tesis, definiremos al lenguaje legal como una manifestación o una variante del lenguaje natural que se utiliza específicamente en el contexto legal, en donde se manejan particulares normas lingüísticas como la sintaxis, vocabulario, términos y significados (Mattila, H. E. 2016).

El lenguaje legal se divide a su vez en otros tipos de lenguaje, dependiendo del área de especialización del profesionista que lo utiliza. Es decir, pueden haber particularidades en los términos utilizados dependiendo del tipo de documento que se esté tratando (Mattila, H. E. 2016).

También, como hemos visto ya en el estado del arte con Venturi (2010), existen particularidades como la ocurrencia sistemática de frases preposicionales y baja presencia de verbos en infinitivo que caracterizan al lenguaje legal como una variante del lenguaje natural que sigue patrones y reglas propias que son impuestas por el uso y la experiencia de los especialistas del área.

A continuación revisaremos la teoría de los lenguajes formales. Específicamente la herramienta que nos ayudará a trabajar con el lenguaje legal.

IV. Lenguajes formales y expresiones regulares

La teoría que sustenta a los lenguajes formales nos sirve para modelar características específicas del lenguaje natural. Particularmente, la teoría sobre expresiones regulares. A continuación, presentaremos una serie de definiciones propias del área de lenguajes formales y autómatas para entender qué son las expresiones regulares y cómo nos pueden ayudar a caracterizar y extraer información escrita en lenguaje legal.

Definiremos primero los conceptos de alfabeto, cadena, longitud de cadena, sigma kleene y lenguaje formal, según Jurado Málaga (2008).

Alfabeto: Un alfabeto es un conjunto no vacío y finito $\Sigma = \{a_1, a_2, \dots, a_n\}$, donde cada $a_i \in \Sigma$, $i = 1, \dots, n$ es un elemento al que llamamos símbolo. Por ejemplo: $\Sigma = \{a, b, c, d, \dots, z\}$

Cadena: Es una secuencia finita de símbolos de un alfabeto. Por ejemplo, dado el alfabeto $\Sigma = \{0, 1\}$, una cadena sería $x = (0, 1, 1, 1, 0) = 01110$

Sigma Kleene (Σ^*): Conjunto de todas las cadenas posibles sobre el alfabeto Σ . Este lenguaje, bajo cualquier alfabeto, siempre es infinito. Por ejemplo, bajo el alfabeto $\Sigma = \{0, 1\}$, $\Sigma^* = \{, 0, 1, 10, 11, 100, 101, 110, 111, 1000, 1001, 1010, 1011, \dots\}$, donde Σ^* representa a los números naturales escritos en sistema binario, y donde el primer elemento, marcado con un espacio, es el elemento nulo.

Lenguaje: Es un conjunto cualquiera de cadenas definidas sobre dicho alfabeto. Por ejemplo, un lenguaje bajo el alfabeto $\Sigma = \{0, 1\}$, puede ser $L = \{00, 01, 10, 11\}$ el cual define los números 0, 1, 2, y 3 en sistema binario.

Con los conjuntos que hemos definido anteriormente, podemos realizar operaciones para crear otros lenguajes o cadenas diferentes. Para operar con cadenas, existe la concatenación, potencia i -ésima de una palabra, y reflexión. Para operar con lenguajes existe la unión, concatenación, clausura o cierre de Kleene, y la clausura positiva. A continuación explicaremos brevemente cada operación según Jurado Málaga (2008).

Concatenación con cadenas: Sean $x^- = (a_1, a_2, \dots, a_n)$ y $y^- = (b_1, b_2, \dots, b_n)$ dos cadenas sobre el alfabeto Σ , la concatenación de x^- y y^- denotada por $x^- \cdot y^-$ es la cadena $(a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_n)$. Por ejemplo, sea $x^- = (c, a, s, a)$ y $y^- = (l, a)$, $x^- \cdot y^- = (c, a, s, a, l, a) = \text{casala}$ o $y^- \cdot x^- = (l, a, c, a, s, a) = \text{lacasa}$.

Potencia i-ésima de una palabra: Consiste en concatenar una cadena x consigo misma, n veces. Por ejemplo: $x = \text{si}$, $x^3 = \text{sisisi}$.

Reflexión de una palabra: Consiste en revertir el orden de una cadena. Por ejemplo: $x = \text{si}$, $x^{-1} = \text{is}$

Unión de lenguajes: La unión de dos lenguajes L_1 y L_2 definidos sobre el alfabeto Σ , es un lenguaje L_3 que contiene todas las cadenas de L_1 y de L_2

$$L = L_1 \cup L_2 = \{ x \mid x \in L_1 \vee x \in L_2 \}$$

Concatenación de lenguajes: La concatenación de dos lenguajes L_1 y L_2 definidos sobre el alfabeto Σ está formada por todas las cadenas obtenidas al concatenar una palabra cualquiera de L_1 con otra de L_2 .

$$L = L_1 L_2 = \{ xy \mid x \in L_1 \wedge y \in L_2 \}$$

Clausura o cierre de Kleene: Es el conjunto de todas las cadenas que se generan del operador potencia i-ésima, incluyendo el elemento vacío.

Clausura positiva: Es el conjunto de todas las cadenas generadas por el operador exponente, a excepción de la potencia cero.

Existen diferentes tipos de lenguajes dependiendo de las restricciones o reglas que acotan las cadenas que lo componen. Dichas restricciones son establecidas por una gramática, y después clasificadas en la jerarquía de Chomsky. Formalmente:

Gramática: Es una 4-tupla $G = (Q, \Sigma, R, q)$ donde Q es un alfabeto de símbolos no terminales, Σ es un alfabeto de símbolos terminales, R es un conjunto de reglas y q es un símbolo inicial, donde $q \in Q$, tal que $Q^* \cap \Sigma^* = \emptyset$ y $Q \neq \emptyset \neq \Sigma$.

Jerarquía de Chomsky: Dependiendo del conjunto de reglas que definen a la gramática, podemos dividirla en 4 tipos: recursivamente enumerable, dependiente del contexto, independiente del contexto y regular. Así, la jerarquía de Chomsky nos dice que las gramáticas recursivamente enumerables contienen a las dependientes del contexto, que a su vez contienen a las independientes del contexto, y finalmente contienen a las gramáticas regulares.



4.3 Jerarquía de Chomsky.

Se observa que un lenguaje va incluyendo al otro, conforme se sube en la jerarquía.

En esta tesis nos interesan los lenguajes regulares, es decir, los lenguajes que son acotados por una gramática regular. Formalmente, una gramática regular se define como:

Gramática regular o tipo 3: 4-tupla $G=(Q,\Sigma,R,q)$ donde Q es un alfabeto de símbolos no terminales, Σ es un alfabeto de símbolos terminales, R es un conjunto de reglas con la forma $A \rightarrow aB | \varepsilon$ donde $a \in \Sigma$ y $B \in Q$, y q es un símbolo inicial, donde $q \in Q$.

Por ejemplo, la gramática $G=(Q,\Sigma,R,q)$ compuesta por $Q = \{S,A\}$, $\Sigma = \{0,1\}$, $q = \{S\}$, $R = \{ S \rightarrow 0A, A \rightarrow 10A | \varepsilon \}$, genera el lenguaje que contiene a todas las cadenas con ceros y unos que tienen sólo un cero, o que empiezan con cero y terminan con cero, siempre y cuando los ceros y unos se vayan intercalando.

Por ejemplo, se puede derivar la cadena 01010 de la siguiente forma: $S \rightarrow 0A \rightarrow 010A \rightarrow 01010A \rightarrow 01010$. Así, el lenguaje regular que genera la gramática regular, puede ser expresado como

$L(G) = \{0,010,01010,0101010,010101010,\dots\}$.

Para definir al lenguaje, no sólo podemos expresarlo como lo hemos hecho en el ejemplo anterior, también podemos utilizar una expresión regular.

Expresión regular: notación algebraica para caracterizar lenguajes regulares.

Retomando el ejemplo anterior, y las operaciones sobre cadenas que hemos visto anteriormente, podemos construir una expresión regular que caracteriza al lenguaje $L(G)$. Dicha expresión regular sería $0(10)^*$, y se lee como "Un cero, seguido o no de un 1 y un 0 repetidos n veces". En este caso, utilizamos las operaciones de concatenación de cadenas y de potenciación de cadena para construir nuestra expresión regular.

Autómata finito: 5-tupla $AF=(Q,\Sigma,q_0,A,\delta)$ donde Q es un conjunto finito de estados, Σ es un alfabeto, q_0 es un estado inicial donde $q_0 \in Q$, A es un conjunto de estados finales donde $A \subseteq Q$, δ es una función de transición $\delta: Q \times \Sigma \rightarrow Q$.

Computacionalmente, un lenguaje regular, una expresión regular, y un autómata finito son equivalentes (Hopcroft, Motwani, & Ullman, 2008).

Con expresiones regulares podemos describir lenguajes (conjuntos de cadenas) específicos, y encontrarlos en un texto utilizando un autómata, o implementando dicha expresión regular con alguna librería disponible en algún lenguaje de programación.

En cierto modo, una expresión regular sirve para generalizar un patrón o una regla que define a un conjunto de cadenas. Es una herramienta útil para esta tesis pues el lenguaje legal tiene patrones específicos que pueden ser modelados, buscados y encontrados con expresiones regulares. Así, hablaremos a continuación sobre el reconocimiento de patrones lingüísticos, donde algunos podrán ser modelados utilizando expresiones regulares.

c) Reconocimiento de patrones lingüísticos

Reconocer patrones lingüísticos, con base en nuestra definición propuesta de reconocimiento de patrones, significa identificar los elementos con los cuales tomamos decisiones respecto a un fenómeno lingüístico. Por ejemplo, en un texto escrito en español, podemos identificar una pregunta cuando está encerrada entre los símbolos '¿' y '?'. Así, el proceso por el cual tomamos la decisión de si una oración es o no una pregunta, es encontrar los símbolos que la delimitan. Siendo así, los símbolos y su ubicación, el patrón que nos permite discernir.

Existen niveles del lenguaje de donde podemos extraer la información que nos ayudará a tomar una decisión respecto al texto y al fenómeno que queramos identificar o estudiar. Estos niveles del lenguaje son, según Domínguez, & Agelvis (2003), los siguientes:

- a) nivel fonético y fonológico;
- b) nivel morfológico y sintáctico;
- c) nivel léxico y semántico
- d) nivel pragmático.

Explicaremos cada uno de ellos y cómo podemos obtener patrones en cada nivel. Así como indicar los niveles donde no se encontraron patrones para esta tesis.

I. Nivel fonético y fonológico

La fonética es el estudio de los sonidos de una lengua, mientras que la fonología se encarga de formalizar los sonidos que existen en la lengua, con ciertos criterios. En nuestro caso, este nivel no aporta información, pues tratamos siempre con lengua escrita.

II. Nivel morfológico y sintáctico

Este nivel puede apreciarse con más facilidad de forma escrita, a diferencia del nivel fonético y fonológico. Por ello, a partir de este nivel empezaremos la búsqueda de patrones lingüísticos.

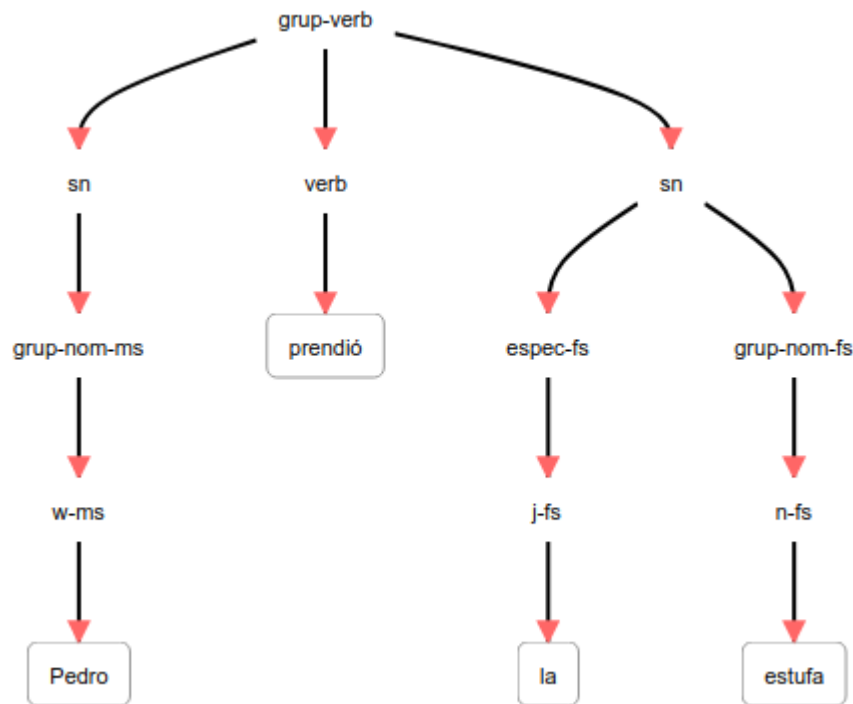
La morfología estudia la forma de las palabras. Es decir, las partes de las mismas que les dan cierto significado. En este nivel podemos encontrar diversos patrones, pues encontramos flexiones, sufijos y prefijos, los cuales podemos identificar utilizando expresiones regulares.

Por ejemplo, el sufijo “-able” sirve para formar adjetivos a partir de un verbo: rentable, irritable, engañable, corrible, etc. Una expresión regular que represente el lenguaje que contiene todas las palabras con este sufijo sería “[a-z]*able”, significando cualquier letra de la ‘a’ a la ‘z’, repetida 0 o n veces, seguido de ‘able’.

Esta expresión regular aplicada a todo el universo de cadenas posibles bajo el alfabeto del español podría traer palabras que no son válidas, como las cadenas “asñldkj”, “fiowuh”, “sdubnxz”. Sin embargo, al aplicarla en un texto escrito en lenguaje natural, podría asumirse que todas las cadenas que encuentre serán correctas, pues difícilmente encontraremos casos como los mencionados, en un escrito que pretende ser comunicativo.

La sintaxis estudia el orden, los roles y la jerarquía en la cual aparecen las palabras en una oración. Por ejemplo, en la oración ‘Pedro prendió la estufa’, ‘Pedro’ es un nombre, ‘prendió’ un verbo, ‘la’ un artículo y ‘estufa’ un sustantivo. Así, una forma alterna de ver la oración, sería ‘nombre verbo artículo sustantivo’. De esta secuencia de categorías gramaticales, pueden encontrarse patrones mediante expresiones regulares.

También, de una oración puede obtenerse un árbol sintáctico que muestra el proceso de derivación de la gramática para obtener dicha oración. Este árbol contiene elementos que pueden ser utilizados como constituyentes en un patrón.



4.4 Árbol sintáctico generado con Freeling 4.0

III. Nivel léxico y semántico

El léxico de una lengua es el conjunto de unidades básicas que componen a dicha lengua Domínguez & Agelvis (2003). Es decir, para el español, todas las unidades básicas pueden ser todas las palabras válidas.

Podemos, por ejemplo, resolver el problema de detectar qué lengua se está escribiendo si ponemos atención a las palabras que se utilizan y logramos identificar a qué léxico pertenecen. Así, del nivel léxico podemos obtener patrones para resolver problemas lingüísticos.

La semántica estudia el significado en la lengua. En este nivel se complica la detección de patrones, pues como hemos visto anteriormente, la lengua al ser de carácter polisémico, vuelve complicado establecer qué significado específico se utiliza. Sin embargo, existen técnicas computacionales para encontrar patrones semánticos en las palabras.

Uno de ellos, se basa en obtener una representación vectorial que defina el significado de la palabra en cuestión. Si tomamos otra representación vectorial de una palabra distinta, podemos medir la distancia entre ellas, y así determinar un grado de similitud entre ambas, dando así a cada una, un significado.

Un ejemplo sencillo de representación vectorial es el siguiente:

dadas las siguientes oraciones

- a) El gato movió ágilmente sus patas.
- b) El perro lamió sus patas.
- c) La tarántula tejió su telaraña.
- d) El gato lamió a sus crías.

Construimos la siguiente tabla, poniendo un 0 si las palabras no concurren en una oración, y un 1 si ambas palabras aparecen en la misma oración.

	movió	lamió	tejió	patas	telaraña	crías
gato	1	1	0	1	0	1
perro	0	1	0	1	0	0
tarántula	0	0	1	0	1	0

4.5 Representación vectorial de las cadenas “gato”, “perro” y “tarántula”, con base en sus contextos

Así, gato = (1,1,0,1,0,1), perro = (0,1,0,1,0,0,) y tarántula = (0,0,1,0,1,0). Obteniendo las distancias entre los 3 vectores, utilizando distancia euclidiana, obtenemos

- a) distancia gato-perro: 1.4142
- b) distancia gato-tarántula: 2.4494
- c) distancia perro-tarántula: 2

La distancia más corta la representa la mayor similitud. Así, con base en los contextos de “gato”, “perro” y “tarántula”, podemos concluir que un gato y un perro se parecen más que un gato y una tarántula, o un perro y una tarántula.

De esta forma, también del nivel semántico podemos obtener patrones e información lingüística para tomar decisiones respecto a un texto.

IV. Nivel pragmático

El nivel pragmático estudia los aspectos relacionados con el uso específico que se le da a la lengua, y es en este nivel donde se presenta una ambigüedad difícilmente computable.

Por ejemplo, en México la frase “colgar los tenis” significa “morir”. Para saber esto, es necesario conocimiento sobre la cultura, o simplemente haber escuchado muchas veces la frase “colgar los tenis” en el mismo contexto que “morir”, lo cual es difícil de implementar en una computadora.

Usando el método descrito en el nivel semántico, podríamos descubrir el significado literal de “colgar los tenis” pero difícilmente el significado coloquial del término, es por esto que no obtendremos patrones de este nivel de la lengua.

Los patrones lingüísticos se utilizan en el área de NER para lograr su objetivo. A continuación hablaremos sobre Reconocimiento de Entidades Nombradas, y del uso de reglas y patrones lingüísticos en dicha área.

d) Reconocimiento de entidades nombradas

El reconocimiento de entidades nombradas surge a partir del área de extracción de información, la cual se encarga de extraer información de texto en lenguaje natural o sin una estructura clara (Nadeau & Sekine 2007). En este sentido, el reconocimiento de entidades nombradas es la extracción de un conjunto de cadenas específico de dicho texto, es decir, entidades nombradas.

El término ‘entidad nombrada’ surgió en la conferencia MUC-6, y abarca las unidades básicas de información en un texto, como pueden ser personas, organizaciones, localidades, expresiones numéricas, expresiones temporales, etc. (Nadeau, D., & Sekine, S. 2007).

Dentro del reconocimiento de entidades nombradas también es importante la clasificación de dichas entidades. Por ejemplo, una organización puede ser clasificada como civil, gubernamental, internacional, o la etiqueta que se necesite. El área actualmente se conoce como Reconocimiento y Clasificación de Entidades Nombradas (NERC por sus siglas en inglés).

Así, el objetivo del Reconocimiento y Clasificación de Entidades Nombradas es extraer de un texto no estructurado, entidades nombradas, para después clasificarlas y estructurarlas. Donde las entidades a encontrar y su clasificación son definidas por la persona realizando la búsqueda.

Como hemos visto anteriormente en *Extracción de entidades nombradas*, dentro de *Antecedentes*, Mansouri et al. (2008), menciona que existen tres principales acercamientos para resolver el problema de NERC: sistemas de reglas, sistemas estadísticos y sistemas híbridos. En este trabajo se ha elegido usar un sistema de reglas, por las razones expuestas anteriormente.

En la sección *Estado del arte* hemos citado el trabajo de Dozier et al. (2010) y analizado sus resultados. Ahora procederemos a hablar sobre el sistema de reglas que han implementado en su

trabajo para solucionar el problema de Reconocimiento de Entidades Nombradas en documentos legales.

I. Sistemas de reglas para reconocer entidades nombradas en texto legal.

Dozier et al. (2010) buscan las entidades de juez, tipo de documento, título, corte y jurisdicción. Para encontrar jurisdicción, jueces y tipo de documento, han usado sistemas de reglas.

La primera regla que establecen está en dividir los documentos legales en dos tipos: con encabezado y sin encabezado. Así, han decidido sólo trabajar con documentos que tengan encabezado, pues en él se encuentra de forma un tanto estructurada a priori, la mayor parte de la información que están buscando.

Antes de pasar al reconocimiento de entidades, existe un preprocesamiento que se aplica a los textos para facilitar la tarea. Este preprocesamiento consiste en dividir el texto en palabras, identificar el encabezado y agrupar líneas de texto relevantes.

Una vez hecho este proceso, el detector de jurisdicciones toma los siguientes pasos para encontrar su jurisdicción.

1. Busca en el documento la palabra *corte*.
2. Obtiene las 5 líneas anteriores a la línea donde se encontró la palabra, así como las 5 líneas posteriores, y les llama *Contexto de corte*.
3. Utilizando cada elemento de una lista que contiene posibles jurisdicciones, y el *Contexto de corte*, se ejecuta un algoritmo que encuentra la subcadena más larga, y ese resultado se etiqueta como jurisdicción.

De forma general, el algoritmo ejecutado en el paso tres necesita dos cadenas. Por ejemplo $X = \text{"ABCDEFGF"}$ y $Y = \text{"FKSIDEFL"}$. Después, encuentra la subcadena más larga que se comparte entre X y Y, es decir $Z = \text{"DEF"}$.

De esta forma, en el paso tres, una cadena está compuesta por el *Contexto de corte*, y la otra cadena son las posibles jurisdicciones. Por ejemplo, tomando $X = \text{"For the district of wyoming"}$ como el *Contexto de corte* y $Y = \text{"district of wyoming, district of columbia"}$ como las posibles jurisdicciones, la subcadena más larga entre X y Y sería $Z = \text{"district of wyoming"}$, lo cual corresponde efectivamente con la jurisdicción.

El reconocedor de tipo de documento utiliza el mismo algoritmo que el reconocedor de jurisdicciones, sólo que en vez de *Contexto de corte*, se toma el título del documento, y en vez de la lista de posibles jurisdicciones, se toma una lista de posibles tipos de documento.

Finalmente, el detector de jueces se vale del nivel léxico del lenguaje, y se dedica a detectar títulos como “Honorable” o “Hon.” o “Judge”, para decir que el nombre que venga después pertenece a un Juez.

El detector de jueces es un sistema de reglas más parecido al implementado en esta tesis, pues los otros dos detectores que usan reglas se basan en la comparación de dos cadenas y en el conocimiento a priori de cortes y tipos de documentos válidos, mientras que los nombres de los jueces pueden variar a través del tiempo sin necesidad de actualizar listas.

Ahora que conocemos un sistema de reglas que funciona para extraer entidades nombradas, hablaremos sobre un sistema de extracción de eventos y actuantes, el cual crearemos utilizando las entidades encontradas.

e) Extracción de eventos y actuantes

Hogenboom et al. (2011) mencionan que existen tres acercamientos a la extracción de eventos. El que utiliza sistemas estadísticos, el que utiliza sistemas de reglas, y el híbrido. También apunta que para los sistemas de reglas es necesario conocimiento lingüístico especializado para la correcta extracción de los eventos.

De forma similar al área de NERC, la extracción de eventos también busca información de documentos escritos en lenguaje natural, sólo que ahora se buscan eventos.

En trabajos como el de Lagos et al. (2010), Borsje, Hogenboom & Frasincar (2010) y Nishihara, Sato & Sunayama (2009), observamos que para detectar eventos es necesario detectar verbos y algunos otros elementos, como fechas, lugares, objetos, sujetos, etc.

Lagos et al. (2010) describen que un evento en un texto se caracteriza por los factores que hacen al evento relevante, y que dichos factores pueden variar dependiendo del documento. Principalmente, menciona que los factores relevantes en eventos son:

1. El tema del evento.
2. Los roles de los participantes del evento.
3. El tiempo en el que ocurre el evento.
4. La localización donde ocurre el evento.

Así, podemos establecer que un evento está caracterizado, primero, por el verbo que representa la acción principal del evento y, segundo, por alguna otra característica que depende directamente del tipo de documento con el que tratamos.

En el caso de documentos legales, hemos establecido que un evento será tal mientras pueda ser detectado en él un verbo y una fecha. Como información adicional a este evento, podremos obtener respuestas a preguntas como ¿quién?, ¿dónde?, ¿a quién?, ¿qué?, ¿cuándo?, donde las respuestas serán entidades nombradas, y cuya detección dependerá de las reglas lingüísticas elaboradas para cada verbo en cuestión.

Así, las respuestas a las preguntas mencionadas serán conocidos como los actuantes del evento, el evento en sí será el tiempo en el que ha ocurrido, y el verbo que representa la acción principal del evento.

Para que este sistema pueda trabajar de manera correcta, es necesario que las entidades nombradas se encuentren disponibles en una estructura computacional accesible, como lo es una base de datos. Se ha decidido utilizar una base de datos no estructurada, cuya teoría explicaremos a continuación.

f) Bases de datos no estructuradas

Las bases de datos no relacionales son una alternativa a las bases de datos relacionales. Han et al. (2011) mencionan algunas razones por las cuales surgió esta alternativa, entre las que destacan:

- 1) La necesidad de mejorar la escalabilidad de las bases de datos, así como su disponibilidad a la hora de hacer actualizaciones.
- 2) Mejorar la velocidad de lectura y escritura.
- 3) Tener capacidad de almacenar grandes cantidades de información.
- 4) Disminuir los gastos de manejo y costos operacionales.

Las bases de datos no relacionales mejoran estos cuatro aspectos.

Así mismo, Han et al. (2011) describen tres tipos de bases de datos no relacionales basados en el modelado que hacen de sus datos, las cuales se explican a continuación:

- a) Llave-Valor.
- b) Bases de datos orientadas en columnas.
- c) Bases de datos con documentos.

Las bases de datos Llave-Valor asocian cada valor a una llave o ID. Esto simplifica la estructura y aumenta la velocidad para realizar búsquedas. Algunos ejemplos de manejadores disponibles actualmente son Redis⁷ y Tokyo Cabinet⁸.

Las orientadas en columnas utilizan tablas tradicionales como modelo de datos, pero no permiten asociarlas entre sí, como se haría normalmente en un modelo relacional. En vez de eso, la información se guarda en cada columna de manera independiente. Esto permite mayor capacidad de compresión de la base de datos, disminuyendo los costos de almacenamiento de grandes cantidades de información a la que no se necesita acceder de forma frecuente.

Finalmente, las bases de datos con documentos son las que utilizamos aquí. En ellas también se asocia una llave a un valor, pero ese valor tiene formato JSON o XML, lo cual le agrega complejidad a la estructura que puede guardarse dentro de la base de datos.

MongoDB es el manejador de base de datos que he decidido usar, pues a pesar de ser una base de datos no relacional, los conceptos para entenderla se acercan mucho a las bases relacionales.

Base de datos relacional	MongoDB
Tabla	Colección
Tupla	Documento
Columna	Campo

4.6 Tabla de equivalencia entre BD relacional y MongoDB⁹

Además, MongoDB ofrece facilidad para cambiar la estructura de la base de datos sin tener que repensar todo el esquema. Es decir, los documentos, al poder ser archivos JSON, son fácilmente manejables, accesibles y modificables.

Para poder acelerar el proceso de creación de una estructura accesible, rápida y costeable, he decidido utilizar MongoDB. También, para este mismo fin, he decidido usar el lenguaje de programación Python, del cual hablaremos a continuación.

g) Lenguaje de programación Python

Este lenguaje es un lenguaje orientado a objetos e interpretado. Tiene una curva de aprendizaje bastante fácil de superar, así como una gran cantidad de librerías enfocadas al procesamiento de lenguaje natural que pueden hacer más sencillo nuestro trabajo (Madnani, N. 2007).

⁷ <https://redis.io/>

⁸ <http://fallabs.com/tokyocabinet/>

⁹ <https://www.mongodb.com/compare/mongodb-mysql>

Python en sí mismo ofrece librerías que facilitan algunas de las tareas expuestas en la sección de *Antecedentes*. Otras librerías se ofrecen de manera externa. A continuación mencionaremos cada una de ellas y el uso específico que se le da en el proyecto.

Para manejar el corpus, usamos la librería de python docx2txt¹⁰ que puede extraer el texto de documentos de Word en formato docx, así como la librería os¹¹ para acceder a cada documento del corpus que se encuentra dentro de una carpeta específica.

Después, una vez con el texto plano del corpus, necesitaremos expresiones regulares para buscar en él los patrones que encontremos para identificar eventos y entidades nombradas. Esto se logra con la librería re¹², la cual cuenta con su propia notación para crear expresiones regulares. Esta librería ofrece diversas opciones para la búsqueda de expresiones regulares, y el manejo de sus respectivos resultados.

Para conectarnos a MongoDB utilizaremos la librería PyMongo¹³ la cual provee MongoDB. Esta librería nos permite acceder a todas las capacidades de MongoDB sin tener que utilizar la terminal misma de Mongo.

Además de estas librerías, cabe mencionar que existen otras que ya ofrecen el servicio de reconocimiento de entidades nombradas y eventos. Por ejemplo, existe una API para acceder a FreeLing¹⁴, la cual nos ofrece reconocimiento de entidades nombradas, entre otras funcionalidades. También es accesible la librería Stanford CoreNLP¹⁵ que realiza tareas similares a FreeLing.

Una vez terminada la programación, tenemos que evaluar de alguna forma el desempeño de nuestra solución, para lo cual necesitaremos establecer métricas. Dichas medidas de desempeño se explican a continuación.

h) Medidas de desempeño

Es necesario establecer una métrica para evaluar el desempeño del programa. En este caso, utilizaremos los conceptos de precisión, exhaustividad y medida f1, las cuales son generalmente usadas para programas que buscan extraer información contenida en textos (Makhoul et al. 1999).

¹⁰ <https://pypi.org/project/docx2txt/>

¹¹ <https://docs.python.org/2/library/os.html>

¹² <https://docs.python.org/3/library/re.html>

¹³ <https://api.mongodb.com/python/current/>

¹⁴ <https://github.com/TALP-UPC/FreeLing/tree/master/APIs/python3>

¹⁵ <https://pypi.org/project/stanford-corenlp/>

Para entenderlos, definiremos primero cuestiones básicas y después hablaremos de cada medida una por una.

Verdadero Positivo (TP): Es un dato que se ha detectado como positivo, y que, en efecto, es positivo.

Falso Positivo (FP): Es un dato que se ha detectado como positivo, pero no lo es.

Verdadero Negativo (TN): Es un dato que se ha detectado como negativo, y que, en efecto, es negativo.

Falso Negativo (FN): Es un dato que se ha detectado como negativo, pero en la realidad es positivo.

Por ejemplo, una entidad clasificada correctamente como entidad es un ejemplo de dato Verdadero Positivo. Una entidad (nosotros sabiendo a priori que sí lo es) la cual no se clasifica como tal, es un Falso Negativo. Es decir, la palabra “Positivo” y “Negativo” refieren al resultado que arroja el programa en cuestión, y las palabras “Verdadero” y “Falso” refieren a la clasificación real del dato estudiado.

A partir de estos cuatro tipos de dato, podemos calcular las medidas de desempeño que hemos mencionado al inicio de la siguiente manera:

$$Precisión = \frac{TP}{TP+FP}$$

$$Exhaustividad = \frac{TP}{TP+FN}$$

$$Medida F1 = \frac{2 * Precisión * Exhaustividad}{Precisión + Exhaustividad}$$

La precisión busca decirnos la proporción de entidades identificadas correctamente. Es decir, la proporción del total de entidades que encontramos, contra el total de entidades.

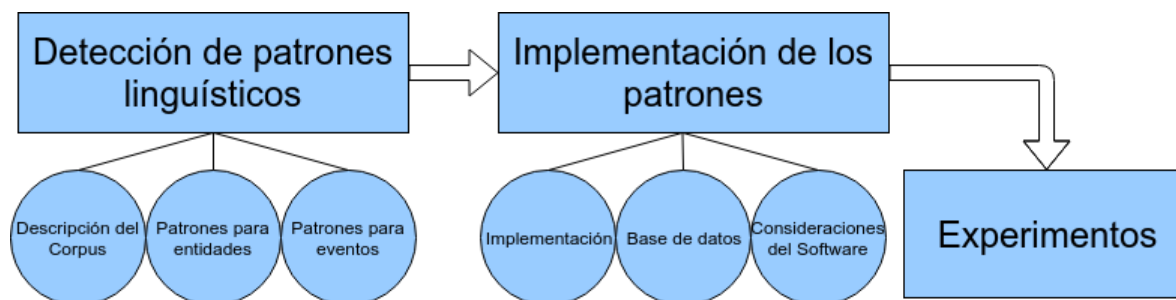
La exhaustividad busca, de la proporción total de entidades que existen, decirnos cuántas fueron encontradas.

Finalmente, la medida F1 busca ser un indicador claro y único del comportamiento, con base en las dos medidas anteriores. Es decir, sirve como un resumen del desempeño obtenido.

Conociendo este marco teórico, es momento de entrar a la sección de *Metodología* donde explicaremos, paso a paso, cómo se han desarrollado los sistemas que dan solución a la problemática planteada en esta tesis, utilizando el marco teórico expuesto.

5. Metodología

En esta sección hablaremos sobre los pasos que se han seguido para resolver la problemática planteada, cumplir el objetivo, y comprobar la hipótesis. De forma general, podemos resumir la metodología con el siguiente diagrama:



5.1 Diagrama de la metodología implementada.

Cada rectángulo es un paso a seguir, primero se detectan los patrones para detectar entidades y eventos, después se implementan en Python, y finalmente se hacen experimentos.

Los círculos son especificaciones dentro de cada paso

El primer recuadro utiliza el marco teórico explicado en los temas *Corpus Lingüísticos* y *Reconocimiento de Patrones Lingüísticos*. El segundo recuadro utiliza los temas *Procesamiento de Lenguaje Natural*, *Reconocimiento de entidades nombradas*, *Extracción de eventos y actantes*, *Bases de datos no estructuradas* y *Lenguaje de programación Python*.

Empecemos entonces hablando sobre el corpus con el que trabajaremos y los patrones que definirán la aparición de entidades y de eventos.

a) Detección de patrones lingüísticos

Los patrones que definirán qué cosa es una entidad y qué cosa es un evento dependen de los documentos específicos con los que trabajaremos. Así, en esta sección, comenzaremos con describir el corpus que utilizamos para trabajar, y después hablaremos sobre los patrones lingüísticos detectados para encontrar entidades nombradas, así como para encontrar eventos y actantes.

Empecemos entonces hablando sobre cómo se ha recopilado el corpus, y las características de los documentos que lo componen.

I. Recopilación y descripción del corpus utilizado

Para la correcta recopilación del corpus es necesario delimitar qué se pretende realizar y el dominio del mismo. En este sentido, el análisis que queremos realizar es la detección de entidades y eventos en resoluciones hechas en sesiones ordinarias del Instituto Federal de Telecomunicaciones, las cuales pueden ser extraídas desde su sitio web¹⁶.

Los criterios que seguimos para descargar documentos de este sitio fueron los siguientes:

- a) Documentos formato .doc o .docx
- b) Documentos del 2015 hasta el 2018

el inciso a) es para asegurarnos de poder extraer correctamente el texto del documento. Si se hubiesen descargado en PDF, no podríamos asegurar que la conversión a texto se ha hecho sin errores.

el inciso b) asegura que haya variedad en el corpus, y que los documentos sean representativos del dominio. Así, el corpus consta de 134 archivos, de los cuales 10 son del 2015, 44 el 2016, 44 del 2017, y 36 del 2018.

Existen más documentos que cumplen el inciso a) para los años 2016 y 2017, por lo que probablemente el sistema se desempeñe mejor para documentos de estos años. Sin embargo, eso será evaluado al final. Probablemente no hay disponibles archivos del 2015 hacia atrás, debido a la lenta digitalización de los mismos.

Año	2015	2016	2017	2018
Número de documentos	10	44	44	36

5.2 Distribución de los documentos por año

El dominio elegido es bastante acotado, pues dentro de las búsquedas que podemos hacer dentro del sitio del IFT, para tipo de documento existen 10 tipos: Acta, consultas públicas, lista de asistencia, orden del día, resoluciones, resoluciones de carácter regulatorio, criterio del pleno, versión estenográfica, voto particular y voto por escrito. De estos 10 tipos sólo elegimos el tipo “resoluciones”. Así mismo, existen dos opciones para tipo de sesión: ordinaria y extraordinaria. En este campo elegimos “ordinaria”.

Con base en este corpus, es momento de hablar de los patrones lingüísticos que permiten reconocer entidades nombradas.

¹⁶ <http://apps.ift.org.mx/cumplimientoStp/secured/adminficum.faces>

II. Patrones lingüísticos para entidades nombradas

Recordemos que un patrón lingüístico es un método, o un proceso, que nos permite tomar una decisión respecto a un fenómeno lingüístico. En este caso, necesitamos, para cada palabra (o conjunto de palabras) tomar la decisión de si dicha palabra es una entidad o no.

Para encontrar patrones lingüísticos fue necesario revisar los documentos manualmente y hacer un análisis sobre qué patrones existen alrededor de las entidades nombradas. El patrón más notorio en este tipo de documentos es el que permite diferenciar las entidades en dos clasificaciones:

- a) entidades con alias.
- b) entidades sin alias.

Un alias es un nombre que se le pone a una entidad para referirse a ella a lo largo del documento, sin tener que decir el nombre completo de la entidad. Por ejemplo:

“Con fecha 11 de junio de 2013 se publicó en el Diario Oficial de la Federación (el “DOF”) [...]”

La entidad es “Diario Oficial de la Federación”, y su alias es “DOF”. Es decir, a lo largo del texto, cada vez que se tenga que referir al Diario Oficial de la Federación, se dirá DOF.

Así, es posible encontrar entidades nombradas buscando primero el alias, y después encontrando la entidad que dicho alias representa. La ventaja de este acercamiento, es que se encontró que los alias generalmente se encuentran dentro de paréntesis, por lo que su identificación es de nivel léxico.

Así mismo, los alias tienen dos formas principales de aparecer, en forma de palabras o palabra, y en forma de siglas. Primero explicaremos los patrones para encontrar entidades con alias, tanto en el caso de tener siglas, como en el caso de tener palabras, y después explicaremos los patrones para encontrar entidades que no contienen alias.

i. Patrones para entidades con alias

Una forma de asegurarnos de que la entidad nombrada es efectivamente una entidad, es si la misma tiene alias. Es decir, toda entidad con alias es, efectivamente, una entidad. Así, de este tipo de entidades se espera tener precisión muy alta en los resultados.

Para encontrar los alias, partimos de que los mismos siempre se encuentran encerrados entre paréntesis. Sabemos que no todo lo que esté entre paréntesis es un alias, pero sí un candidato a alias. Así, a todo lo que esté entre paréntesis, le denominaremos “candidato a alias”. Por ejemplo, en el extracto:

“1.- Refrendo de la Concesión. El 26 de enero del 2000, de conformidad con los artículos 1° y 16 y demás relativos a la Ley Federal de Radio y Televisión (en lo sucesivo, la “LFRTV”), la Secretaría de Comunicaciones y Transportes (en lo sucesivo, la “SCT”) otorgó a favor de Lucia Pérez Medina Vda. de Mondragón (en lo sucesivo el “Concesionario”), el Refrendo de la Concesión para continuar usando comercialmente el canal 2 (-) (54-60 MHz), a través de la estación de televisión con distintivo de llamada XHKG-TV, en Tepic, Nayarit, con vigencia de 10 (diez) años, contados a partir del día 3 de julio de 1999 y vencimiento el 2 de julio de 2009 (en lo sucesivo, la “Concesión”).”

Observamos los siguientes candidatos a alias:

- 1) en lo sucesivo, la “LFRTV”
- 2) en lo sucesivo, la “SCT”
- 3) en lo sucesivo el “Concesionario”
- 4) -
- 5) 54-60 MHz
- 6) diez

de los seis candidatos, sólo los primeros tres son alias, pues están asociados a las entidades Ley Federal de Radio y Televisión, Secretaría de Comunicaciones y Transportes, y Lucia Pérez Medina Vda. de Mondragón, respectivamente. Esta asociación puede hacerse mediante patrones, que se describen a continuación, en términos de reglas.

a. Regla 0

Primero hay que determinar si el alias son siglas o es una palabra. Se consideran siglas si la cantidad de mayúsculas en la palabra es mayor que la cantidad de minúsculas. Así, alias como “LFTyR” y “INEGI” son encontrados como siglas.

Después, se asume siempre que la entidad se encuentra en algún punto entre el último salto de línea (inicio del párrafo), y el candidato a alias. A este fragmento del texto le denominaremos “contexto”.

Después, se sigue una lista de criterios para encontrar la entidad que las siglas representan.

- 1) Buscar mayúsculas en orden
- 2) Buscar minúsculas en orden
- 3) Buscar mayúsculas en desorden
- 4) Buscar con expresión regular

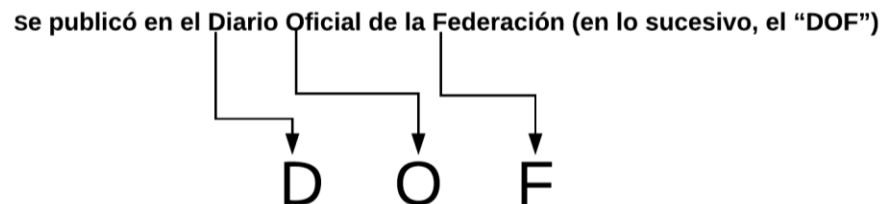
Estas cuatro búsquedas se realizan siempre en el contexto. Los primeros tres puntos se refieren a buscar las letras específicas de las siglas, y el último se refiere a utilizar una expresión regular en el contexto. Hablaremos de cada caso uno por uno a continuación.

1) Buscar mayúsculas en orden

Primero se buscan en el contexto las letras que componen a las siglas en orden y en mayúsculas. Por ejemplo, en el siguiente contexto:

“se publicó en el **Diario Oficial de la Federación** (en lo sucesivo, el “**DOF**”)”

Se busca una palabra que empiece con “D”, luego una con “O” y finalmente una con “F”. Si se encuentra, se toma como entidad desde la primera letra encontrada, hasta el inicio del candidato a alias. En este caso, “Diario Oficial de la Federación”.



Esta correspondencia no siempre es tan obvia, por ejemplo en el contexto:

“la **Comisión Federal de Telecomunicaciones** (en lo sucesivo, la “**COFETEL**”)”

la “C” puede corresponder a “Comisión” pero la “O” ya no corresponde a ninguna palabra. Esto es porque realmente “CO” corresponde a “Comisión”. Para estos casos, patrones comunes en siglas como “CO” para “Comisión” se sustituyen sólo por “C”. A continuación se muestra una tabla donde aparecen todas las contracciones realizadas.

Forma completa	Contracción
NA	N
DI	D
TV	T
CO	C
FE	F
TEL	T
MEX	M

5.3 Contracciones a siglas.

Es decir, si encontramos la cadena “NA”, se sustituye con “N”, y así sucesivamente con cada registro en la tabla.

Así, siguiendo el ejemplo anterior, las siglas “COFETEL”, se contraen a “CFT”. De esta forma, la correspondencia con “Comisión Federal de Telecomunicaciones” se da de la misma forma que el primer ejemplo con el “Diario Oficial de la Federación” (DOF).

2) Buscar minúsculas en orden

Si utilizando el punto anterior no se encontró ninguna entidad, entonces se buscan letras que correspondan con las siglas, aunque no estén en mayúscula. Por ejemplo, en el contexto:

“mil millones de personas en todo el mundo razón por la cual la Unión Internacional de Telecomunicaciones (“UIT”) considera que las tecnologías de información y comunicación (“TIC”)”

Se buscó la “T”, “I” y “C” como inicio de palabra, en orden, y en minúscula. Encontrando así “tecnologías de información y comunicación” como entidad.



Si también falla, se buscan las mayúsculas aunque sea en desorden.

3) Buscar mayúsculas en desorden

Dado el contexto:

“Cabe mencionar que G3ict es un ente creado por la **A**ianza **G**lobal para el **D**esarrollo de las Tecnologías de la Información y Comunicación (**GAID**, por sus siglas en inglés)”

Se buscan las letras “G”, “A”, “I”, y “D”, en el contexto aunque no se encuentren en orden. Cuando se encuentren las cuatro letras, entonces se tomará desde la última letra encontrada, hasta el candidato a alias como una entidad.



Finalmente, si no se han podido resolver las siglas utilizando los elementos de las mismas, entonces se busca una expresión regular.

4) Buscar con expresión regular

La expresión regular devuelve palabras que empiezan en mayúscula, seguidas o separadas por una o varias de las palabras “la”, “el”, “los”, “las”, “un”, “una”, “uno”, “unas”, “unos”, “y”, “con”, “de”, “del”.

Un ejemplo de este caso se observa en el siguiente contexto:

“Por otra parte a nivel internacional, se cuenta con un estándar de accesibilidad web desarrollado por la World Wide Web Consortium (W3C)”

Al haber fallado los otros algoritmos al no encontrar el número 3, la expresión regular devuelve “World Wide Web Consortium” la cual corresponde con la entidad.

Finalmente, si las siglas no se han podido resolver, no se revisarán minúsculas en desorden, pues podría traer muchos falsos positivos. Así, se considerará que las siglas representan a una entidad nombrada si se ha encontrado alguna entidad usando las reglas mencionadas.

A continuación, hablaremos de los casos donde el alias no son siglas.

b. Regla 1

Si el candidato a alias tiene un artículo seguido de una palabra que empieza en mayúsculas, y el candidato tiene menos de diez palabras, entonces se busca la entidad asociada al alias.

La entidad se puede definir por dos patrones:

- a) artículo + entidad + candidato a alias
- b) entidad + candidato a alias

El artículo será aquel que esté seguido de una palabra que empieza en mayúsculas, y que esté contenido en el candidato a alias.

Hablemos primero de cómo encontrar el patrón indicado en el inciso a) y de algunos ejemplos, y después de cómo encontrar el patrón del inciso b).

1) artículo + entidad + candidato a alias

Primero, del candidato a alias se obtiene una cadena conformada por un artículo seguido de una palabra que empieza en mayúsculas. Por ejemplo, si el candidato a alias es “en lo sucesivo, el Concesionario”, se obtendrá la cadena “el Concesionario”, pues es un artículo seguido de una palabra en mayúscula.

Después, se buscará esa cadena hacia atrás en el contexto para determinar a la entidad nombrada. Por ejemplo, con base en el contexto: “conforme a lo establecido en los artículos 15 fracción IV y 17 fracción I de [la Ley Federal de Telecomunicaciones y Radiodifusión \(la “Ley”\)](#)”

Del candidato a alias “la “Ley”” obtenemos la cadena (subrayado) “la Ley”, donde “la” es el artículo, y “Ley” la primer palabra de la entidad a encontrar. Después, buscamos dicha cadena hacia atrás en el contexto. Una vez encontrada, la entidad será la que corresponda con el patrón artículo + entidad (en color verde) + candidato a alias (en color rojo). En este caso, la entidad es “Ley Federal de Telecomunicaciones y Radiodifusión”, pues es lo que se encuentra entre el artículo y el candidato a alias.

Otro ejemplo, con base en otro contexto: “[el “Decreto por el que se reforman y adicionan diversas disposiciones de los artículos 6o., 7o., 27, 28, 73, 78, 94 y 105 de la Constitución Política de los Estados Unidos Mexicanos, en materia de telecomunicaciones”](#) (en lo sucesivo, [el “Decreto de Reforma Constitucional”](#))”

Se obtiene del candidato a alias la cadena “el Decreto”, y se busca hacia atrás. Una vez encontrado, se determina que la entidad es “Decreto por el que se reforman y adicionan diversas disposiciones de los artículos 6o., 7o., 27, 28, 73, 78, 94 y 105 de la Constitución Política de los Estados Unidos Mexicanos, en materia de telecomunicaciones”, pues es lo que se encuentra entre el artículo y el candidato a alias.

Existirán casos donde no se encontrará la cadena conformada por el artículo y la palabra tal cual como viene escrita. Es decir, no siempre la palabra en el contexto se encuentra con mayúscula inicial, así como lo está dentro del candidato a alias. Por ejemplo en el contexto: “y de manera

particular el "PROGRAMA DE TRABAJO PARA REORGANIZAR EL ESPECTRO RADIOELÉCTRICO A ESTACIONES DE RADIO Y TELEVISIÓN" (el "Programa"). La palabra "Programa" no se encontrará tal cual, se encuentra toda en mayúscula. Este caso, tanto aquel en el que la palabra se encuentra toda en minúscula, están contemplados.

En el caso donde la cadena "el Decreto" o "la Ley" no pueda encontrarse en el contexto, se buscará sólo la palabra "Decreto" o la palabra "Ley". A continuación lo explicamos.

2) entidad + candidato a alias

En el contexto: "La Secretaría de Comunicaciones y Transportes (la "Secretaría"), otorgó a favor de el C. Gerardo Abraham Carreño López un título de concesión para instalar, operar y explotar una red pública de telecomunicaciones para prestar el servicio de televisión restringida, con cobertura en Calimaya de Díaz González, México, con una vigencia de 10 (diez) años contados a partir de su otorgamiento (la "Concesión")."

No se encuentra la cadena "la Concesión", así que ahora sólo buscamos la palabra "Concesión" hacia atrás en el contexto. También se busca en mayúsculas y minúsculas. Una vez encontrada, se toma desde esa palabra, hasta el candidato a alias como entidad. En este caso, la entidad se ha marcado con color verde.

c. Regla 2

Si dentro del candidato a alias existe una sola palabra con mayúscula inicial, o varias palabras, pero la primer palabra tiene mayúscula inicial, entonces se considera dicha palabra como la que se buscará hacia atrás en el contexto para determinar la entidad. Es decir, se aplicará el patrón:

a) palabra + entidad + candidato a alias

Por ejemplo, en el contexto: "Tercera Ampliación de cobertura. Con fecha 27 de mayo de 2013, la Secretaría autorizó a la C. Margarita Ciprés Cruces, la ampliación de cobertura de la concesión, hacia las localidades de San Pedro Tlanixco, Municipio de Tenango de el Valle; San José, San Diego, Santa María Aranzazú (Santa María)".

Se detecta la palabra "Santa" y se busca hacia atrás en el contexto. Así, se ha marcado la entidad con color verde, el candidato a alias en color rojo y la palabra buscada se ha subrayado.

d. Regla 3

Si dentro del candidato a alias existe la palabra "siglas", es probable que dentro del candidato a alias se encuentren siglas, por lo que se buscan las mismas palabra por palabra, y luego se resuelven usando el algoritmo de la regla 0.

Por ejemplo, en el contexto: “Cabe mencionar que G3ict es un ente creado por la Alianza Global para el [Desarrollo de las Tecnologías de la Información y Comunicación](#) (GAID, por sus siglas en inglés)”

Se detecta la palabra “siglas” en el candidato a alias, así que se buscan y se encuentran en la palabra “GAID”. Después, con el algoritmo de la regla 0 se resuelve la entidad marcada en color verde.

e. Regla 4

Si no se ha encontrado con ninguna de las reglas anteriores a la entidad que corresponde con el candidato a alias, se procede a buscar dentro del contexto algún elemento presente en un diccionario el cual contiene casos particulares observados en el corpus.

Por ejemplo, en el contexto “En tanto, la G3ict (Iniciativa Global para TIC incluyente)” se detecta que el alias de la entidad “G3ict” es “Iniciativa Global para TIC incluyente” utilizando el diccionario.

Si ninguna de las cuatro reglas ha encontrado una entidad que corresponda con el alias identificado dentro del candidato a alias, se determina entonces que dicho candidato no contiene un alias y que por lo tanto no hay una entidad a encontrar en dicho caso.

A continuación presentaremos el caso para detectar entidades que no vienen acompañados de alias, para lo cual se usan reglas distintas a las expuestas anteriormente.

ii. Patrones para entidades sin alias

Estas entidades son las más frecuentes en el corpus, y representan el mayor reto, pues a diferencia del candidato a alias, su aparición no está definida por un elemento en concreto como los paréntesis, sino por varios, como lo pueden ser palabras específicas que indiquen una persona, empresa, ley, organización o documento.

Así, para detectar una entidad, primero buscamos candidatos a entidades, y después pasamos a los candidatos por una serie de filtros para determinar si efectivamente son entidades o no.

Para el proceso de detección de candidatos se utilizan tres reglas:

- 1) Palabras que empiezan en mayúscula separadas, o no, por alguna de las palabras “la”, “el”, “los”, “las”, “un”, “una”, “uno”, “unas”, “unos”, “y”, “con”, “de”, “del”, “a”, una o más veces. Incluyendo, o no, al inicio, las siguientes cadenas: “CC.”, “C.”, “Mtro.”, “Mtra.”, “Sra.”, “Sr.”, “Lic.”, “Dr.”, “Dra.”, “Pdte.”, y al final, las cadenas “S.A. de C.V.”, “S. de R.L. de C.V.”, “S.A.B. de C.V.”, “S.A.”.

- 2) Cadenas que empiezan con las palabras “oficio”, “expediente” o “acuerdo”, seguidos de un patrón alfanumérico separado por diagonales o guiones.
- 3) Cadenas que empiecen con un número, o con las palabras “artículo”, “fracción”, o “inciso” tanto en su forma normal, como en plural, seguidas por las palabras “del” o “de la”, seguidas del patrón descrito en el inciso 1).

El primer patrón busca encontrar nombres de instituciones, organizaciones, leyes y personas. El segundo patrón busca encontrar documentos, y el tercer patrón busca encontrar artículos y contenidos de leyes en general.

A simple vista, se deduce que el patrón número dos y tres traerán menos falsos positivos que el patrón número uno. Para resolver esto, utilizamos tres criterios que determinan si lo devuelto por los patrones son efectivamente entidades:

- 1) Si tiene menos de tres palabras, no es entidad.
- 2) Si la primer palabra es un artículo, no es una entidad.
- 3) Si ya existe en la base de datos como un alias, entonces no es una entidad.

Finalmente, todas las entidades encontradas como tal, tanto con alias como sin alias, se clasifican buscando palabras dentro de la entidad. Con la entidad convertida en minúsculas, a continuación presentamos las palabras con las que se realiza la clasificación.

- 1) Para clasificar como **Persona**, se busca alguna de las siguientes palabras: "dr.", "mtro.", "mtra.", "lic.", "presidente", "c.", "cc.", "comisionado", "comisionada", "coordinador", "coordinadora", "defensor", "defensora", "delegado", "delegada", "director", "directora", "responsable", "secretario", "secretaria", "subsecretaria", "subsecretario", "titular".
- 2) Para clasificar como **Ley**, se busca alguna de las siguientes palabras: “ley”, “artículo”, “párrafo”, “sección”.
- 3) Para clasificar como **Organización**, se busca alguna de las siguientes palabras: "banco", "instituto", "organización", "s.a.", "c.v.", "dirección", "administración", "administradora", "censo", "comisión", "competencia", "consejo", "consorcio", "coordinación", "departamento", "disposición", "ejecutivo", "ejecutoria", "gobierno", "grupo", "junta", "juzgado", "oficialía", "organismo", "pleno", "sala", "secretaría", "sistema", "subsecretaría", "tribunal", "tribunales", "unidad", "unidades", "universidad".
- 4) Para clasificar como **Documento**, se busca alguna de las siguientes palabras: "acuerdo", "expediente", "constitución", "acta", "resolución", "oficio", "anexo", "documento", "concesión", "constancia", "contrato", "código", "decreto", "derechos", "diario", "escrito", "escritura", "estatuto", "formato", "lineamiento", "lineamientos", "lista", "manual", "otorgamiento", "presupuesto", "prórroga", "publicación", "registro", "reglamento", "reglas", "requerimiento", "resolutivo", "solicitud", "solicitudes".

- 5) Para clasificar como **Lugar**, se busca alguna de las siguientes palabras: "municipio", "estado", "colonia", "calle", "delegación", "oficina".

La jerarquía de la clasificación es el orden en el que se han expuesto. Es decir, si una entidad tuviese palabras que pertenecen a más de una clasificación, la categoría que se elige es la de mayor jerarquía. Además, si existiese una entidad que no entre en alguna de estas clasificaciones, entonces se asigna **Otro**, como clase.

Se ha elegido este orden pues es el que ha arrojado menos errores a la hora de hacer la clasificación. Esto se ha medido durante el desarrollo del programa, por lo que no poseemos una medida específica al respecto.

Es de esta manera como buscaremos detectar las entidades nombradas en nuestros documentos. Ahora, con base en dichas entidades, hablaremos sobre los patrones que usaremos para encontrar eventos.

III. Patrones lingüísticos para detección de eventos

Es importante mencionar que los patrones descritos en toda la sección de eventos, así como su implementación, corresponden, en cierta medida, con lo descrito en Sierra et al. (2018), pues esta tesis ha sido parte de dicho esfuerzo.

Encontrar eventos en los documentos legales es más sencillo después de haber obtenido las entidades nombradas y su clasificación, sin embargo, sigue siendo una tarea compleja. Para disminuir dicha complejidad, acotamos el universo de eventos que podemos obtener, eligiendo sólo cinco verbos: emitir, otorgar, presentar, publicar y solicitar.

Después, obtendremos del documento, utilizando Freeling 4.0, sólo las oraciones que contengan una fecha, y uno de los cinco verbos mencionados. Sobre estas oraciones se responderán cuatro preguntas, las cuales son las siguientes:

	Who	What	To Whom	Where
emitir	YES	YES	NO	NO
otorgar	YES	YES	YES	NO
presentar	YES	YES	YES	NO
publicar	YES	YES	NO	YES
solicitar	YES	YES	YES	NO

5.4 Tabla para obtención de actantes.

**Describimos, para cada verbo, qué información buscaremos obtener
(Sierra et al. 2018)**

Para responder cada pregunta, se utilizarán varios patrones que describiremos a continuación, y que se buscarán en la oración obtenida.

i. Patrón para ¿Quién?

Se buscarán los siguientes patrones:

- 1) representante legal de + * + entidad
- 2) por medio de + * + entidad
- 3) a través de + * + entidad
- 4) fecha + * + mediante + * + entidad + * + verbo
- 5) fecha + * entidad + * + verbo

Dónde “entidad” es una entidad o alias presente en la base de datos, el asterisco simboliza cualquier cantidad de caracteres (incluyendo ningún carácter), “fecha” es la fecha detectada en la oración y “verbo” es el verbo que haya sido encontrado en la oración.

La respuesta a la pregunta “¿Quién?” corresponde en el patrón a la palabra “entidad”, siendo todo lo demás sólo el contexto en el que dicha entidad se vuelve la respuesta correcta a la pregunta.

Además, sólo serán válidas las entidades de clase “Organización” o “Persona”, pues sería irracional que la respuesta fuese una entidad del tipo “Documento” o “Lugar”. Así, la clasificación de las entidades nos ayuda a que los resultados tengan menos falsos positivos.

ii. Patrón para ¿Qué?

Para responder esta pregunta, definiremos tres patrones:

- a) (opinión + * + entidad)
- b) (opinión + *)
- c) verbo + * + entidad

En los dos primeros patrones, la respuesta a la pregunta “¿Qué?” es todo lo que se encuentre entre paréntesis. En el último patrón, la respuesta es “entidad”.

Las entidades válidas para estos patrones son las de tipo “Otro” y “Documento”.

iii. Patrón para ¿Dónde?

Este caso es sencillo, pues sólo lo hemos asociado a un verbo. Así, el único patrón a buscar es:

- a) en + * + entidad.

Este patrón puede traer muchos falsos positivos, además, sabemos que la palabra “en” no sólo marca lugar. Para solucionarlo, acotamos las entidades a las de tipo “Otro” y “Documento”, pues sabemos que el verbo “publicar” generalmente refiere a la publicación dentro de un documento, y no a un lugar físico como un estado o un municipio.

iv. Patrón para ¿A quién?

Se definen tres patrones para responder esta pregunta:

- a) ante + * + entidad
- b) a quien + * + entidad
- c) en favor de + * + entidad
- d) verbo + * + entidad

Las clases válidas para “entidad” son “Persona”, “Organización” y “Documento”.

Notemos que, a diferencia de los patrones para encontrar entidades, los patrones para encontrar eventos han sido muy sencillos, pues parten de la premisa de que ya contamos con las entidades en una base de datos. Esto podría ser una desventaja, pues los errores en el reconocimiento de entidades se traspasa automáticamente al sistema de detección de eventos.

Ahora que tenemos los patrones, para detectar entidades y eventos, pasaremos a implementarlos en algoritmos computacionales que permitan la solución al problema planteado.

b) Implementación de los patrones

Hablaremos en esta sección sobre la implementación de los patrones para encontrar entidades sin alias, la implementación para encontrar entidades con alias, la implementación para encontrar eventos, la base de datos que hemos hecho, y algunas consideraciones que debemos tener para cada parte del software (detector de entidades nombradas y detector de eventos).

Primero, hablaremos sobre el reconocedor de entidades nombradas, el cual se ha escrito para funcionar en Python 3 y se ha dividido en dos grandes partes:

- 1) Detección de entidades con alias
- 2) Detección de entidades sin alias

Para la primera parte, se han definido las siguientes funciones:

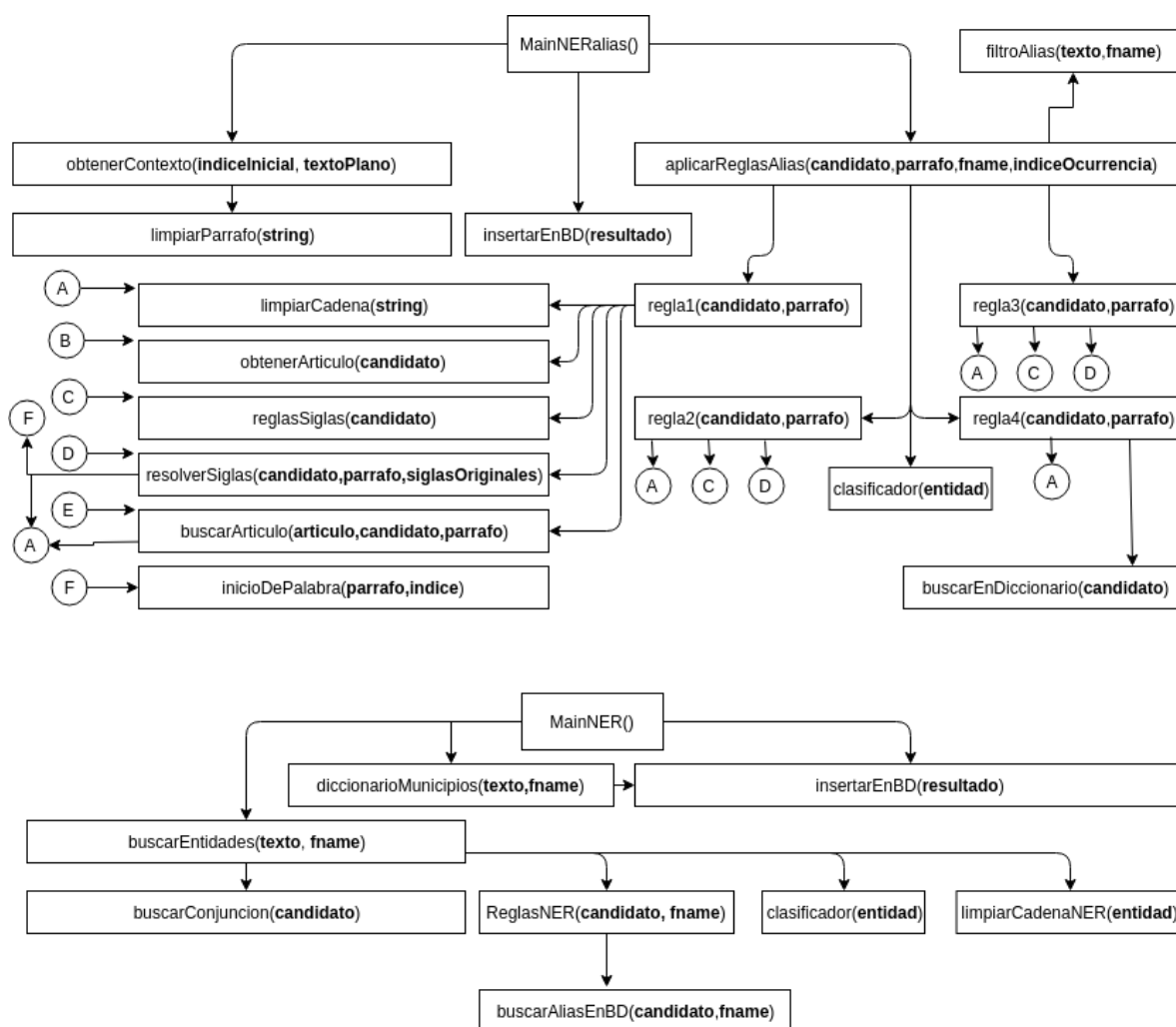
- 1) MainNERalias()
- 2) aplicarReglasAlias(candidato,parrafo,fname,indiceOcurrencia)
- 3) insertarEnBD(resultado)
- 4) obtenerContexto(indiceInicial, textoPlano)

- 5) limpiarParrafo(string)
- 6) regla1(candidato,parrafo)
- 7) regla2(candidato, parrafo)
- 8) regla3(candidato, parrafo)
- 9) regla4(candidato)
- 10) clasificador(entidad)
- 11) filtroAlias(candidatoAlias)
- 12) buscarEnDiccionario(candidato)
- 13) limpiarCadena(string)
- 14) obtenerArticulo(candidato)
- 15) reglasSiglas(candidato)
- 16) resolverSiglas(candidato, parrafo, siglasOriginales)
- 17) buscarArticulo(articulo,candidato,parrafo)
- 18) inicioDePalabra(parrafo, indice)

Para la segunda parte, se han definido las funciones:

- 1) MainNER()
- 2) buscarEntidades(texto, fname)
- 3) insertarEnBD(resultado)
- 4) buscarConjuncion(candidato)
- 5) ReglasNER(candidato, fname)
- 6) clasificador(entidad)
- 7) limpiarCadenaNER(string)
- 8) buscarAliasEnBD(candidato, fname)
- 9) diccionarioMunicipios(texto,fname)

A continuación, se presenta un diagrama que indica cómo se relacionan las funciones mencionadas. Es decir, se indica qué función utiliza qué otra función, dentro del programa, con el fin de establecer un mapa preliminar del programa en cuestión.



5.5 Diagrama general de interconexión entre funciones

El diagrama tiene dos funciones principales: MainNERalias() y MainNER(), la primera busca entidades con alias, y la segunda entidades sin alias.

En este diagrama se visualiza la división del programa en dos partes distintas; la superior encargada de implementar los patrones para reconocer entidades nombradas con alias, y la parte inferior para reconocer entidades nombradas sin alias, en concordancia con el análisis realizado de los patrones en el corpus.

A continuación, se muestran diagramas de flujo para casi cada función, así como la interconexión de todas las funciones que describen el programa realizado. Existe código dentro de las funciones que podrían no aparecer en los diagramas, pues se han obviado algunos pasos.

Esto ocasiona que los diagramas no sean réplicas exactas de lo que se ha programado, sino descriptores de la lógica general que siguen los programas, la cual, en cada caso, podrá

apreciarse implementada en el código original al que pueden pedir acceso a través de mi información de contacto.

Además, dentro de los diagramas se ha usado lenguaje natural para explicar los procesos, con el fin de facilitar la lectura, y se han ocupado, cuando se ha creído conveniente, expresiones propias del lenguaje Python 3. Por ejemplo, cuando se llama a una función que se describe en otro diagrama, o cuando se utiliza explícitamente una expresión regular. Además, se han marcado los nombres de variables con negritas, para diferenciarlas de texto común.

Como última consideración, no todas las funciones tienen diagrama de flujo. Para las que no tienen, se ha considerado prudente explicar su funcionamiento al mismo tiempo que se explica el mecanismo de otra función, pues se consideró las operaciones realizadas por la función son un tanto triviales.

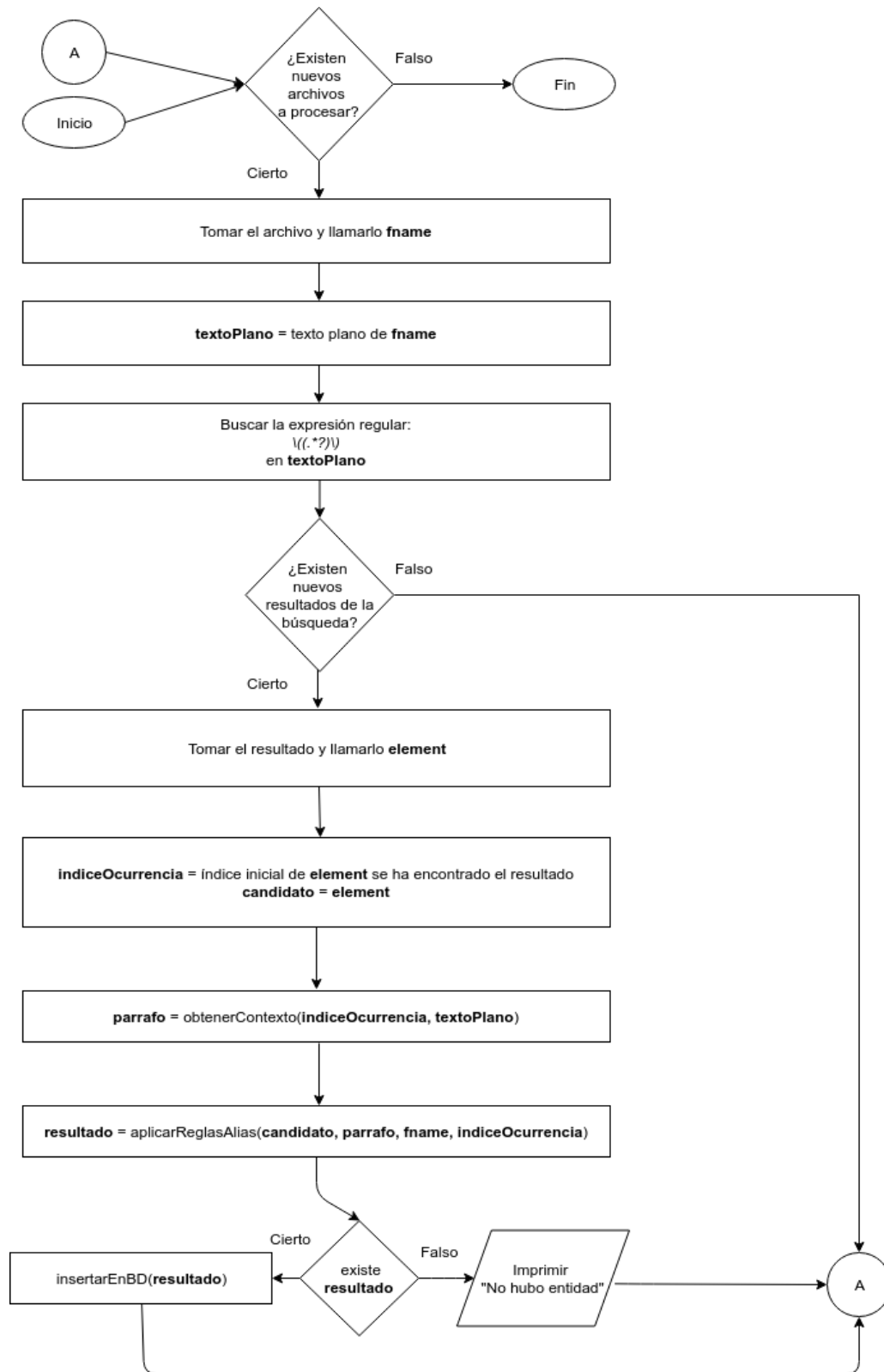
Hablaremos sobre las dos grandes divisiones del programa, es decir, la implementación para reconocer entidades con alias, y la implementación para reconocer entidades sin alias. A su vez, hablaremos específicamente, dentro de cada sección, de las funciones que han servido para cada fin. También sobre la base de datos que hemos implementado, y la estructura específica con la que se ha decidido guardar la información. Finalmente, hablaremos de la implementación de los patrones para reconocer eventos, y consideraciones generales respecto al software utilizado en todas las implementaciones.

Empecemos entonces analizando el programa que implementa los patrones para detectar entidades con alias.

I. Implementación de patrones para entidades con alias

Primero hablaremos sobre las funciones `MainNERalias()` y `aplicarReglasAlias(candidato, parrafo, fname, indiceOurrencia)`, las cuales se encargan de empezar los algoritmos que aplican las reglas descritas en la sección *Patrones para entidades con alias*. Más tarde, hablaremos sobre `regla1(candidato, parrafo)`, `resolverSiglas(siglas, parrafo, candidato)`, `buscarArticulo(articulo, candidato, parrafo)`, `regla2(candidato, parrafo)`, `regla3(candidato, parrafo)` y `regla4(candidato, parrafo)`.

i. Función MainNERalias()



En esta función se destaca que el programa está hecho para leer varios archivos al mismo tiempo. Es decir, se colocan los documentos legales en una carpeta, a la cual el programa ingresa, y aplica el algoritmo a cada archivo.

Después, la expresión regular `((.*?))` devuelve todo lo que se encuentre entre paréntesis. Es decir, conforme a los patrones analizados, es aquí donde se obtienen los “candidatos a alias”.

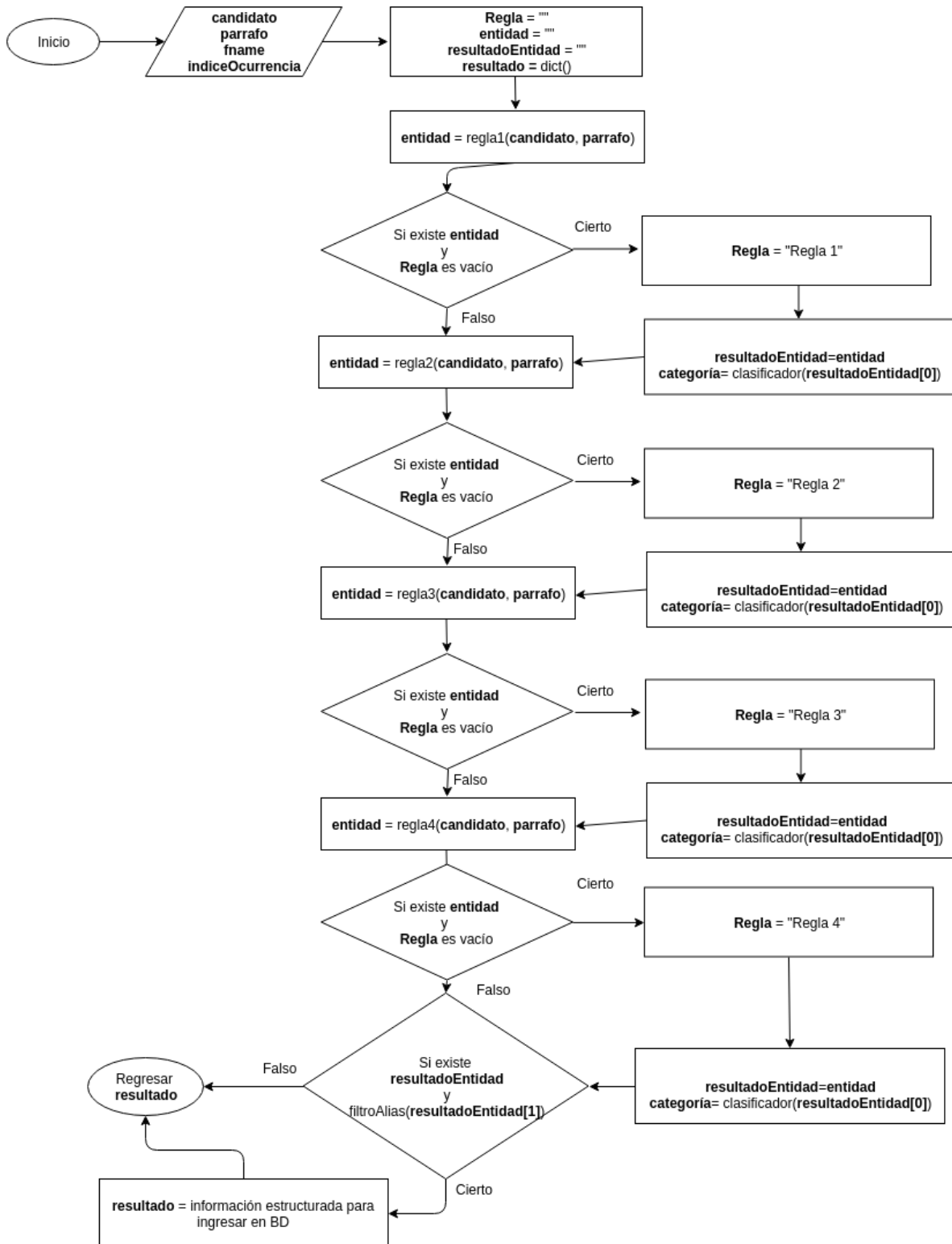
Más adelante, la función `obtenerContexto(indiceOcurriencia, textoPlano)` devuelve lo que se definió en la sección correspondiente como “contexto”, y que en la implementación llamamos “párrafo”.

Finalmente, se aplican las reglas usando la función `aplicarReglasAlias(candidato, párrafo, fname, indiceOcurriencia)`, y si hubo resultado, se guarda en la base de datos, si no lo hubo, pasamos al siguiente archivo a analizar.

Esta función es la abstracción máxima de esta primera sección, pues abre los archivos, aplica las reglas, y guarda los resultados. Conforme vayamos entrando en las demás funciones, la complejidad aumentará.

A continuación, procedemos a explicar la función `aplicarReglasAlias`.

ii. Función aplicarReglasAlias(candidato, parrafo, fname, indiceOcurrencia)



Primero, aclaro que el paralelogramo al inicio representa las entradas que la función recibe al momento de ser llamada.

También se observa que las llamadas a las respectivas reglas se hacen de forma secuencial, y que una vez que una regla ha dado un resultado afirmativo, la variable “Regla” toma un valor que impide que se verifique el resto de las reglas. Así, podemos decir que hemos establecido una jerarquía entre las cuatro reglas ya explicadas en la sección correspondiente, donde la primer regla es la más importante, y la cuarta la menos importante.

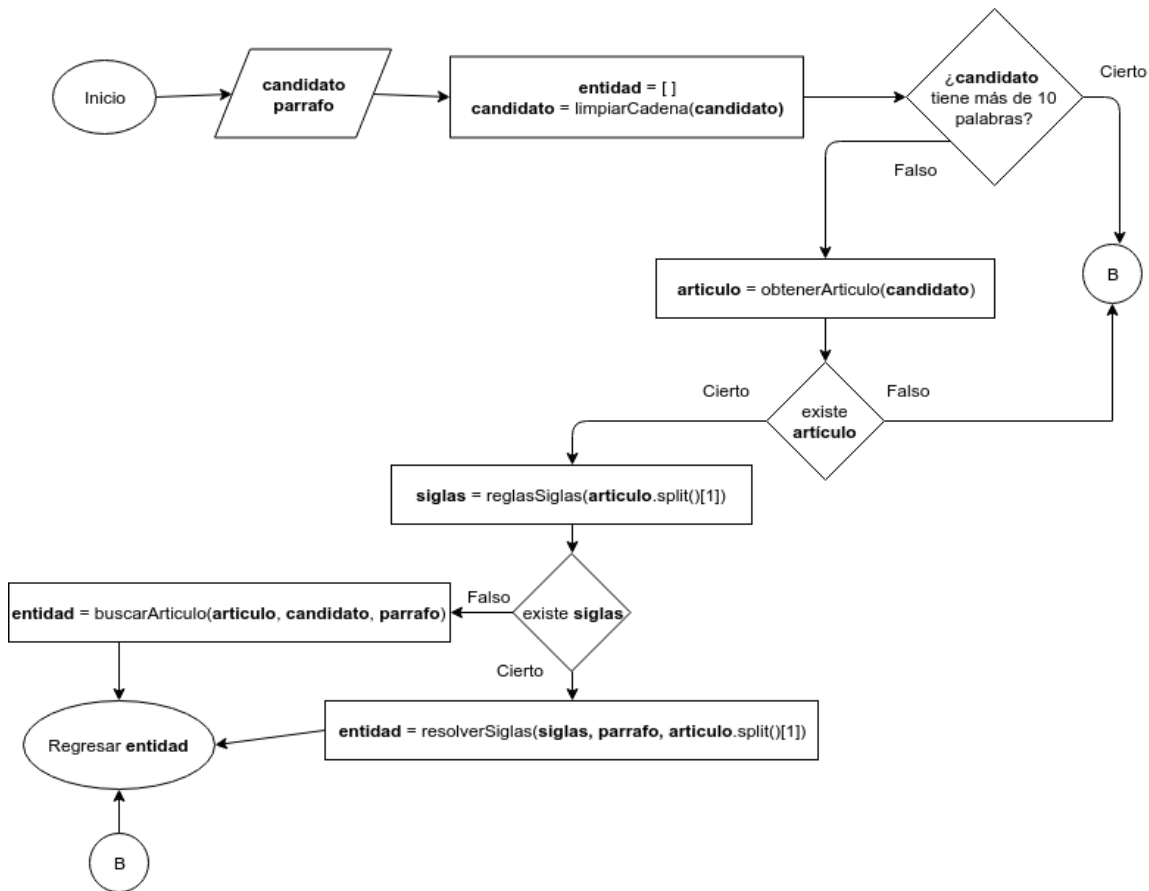
Finalmente, observamos la presencia de una función “filtroAlias(candidatoAlias)”. Esta revisa tres criterios para ver si se descarta el Alias, o no.

- 1) Si el alias es sólo un caracter, se descarta.
- 2) Si el alias son puntos suspensivos, se descarta.
- 3) Si el alias contiene la frase “en los sucesivo”, se descarta.

En esta función se podrían aplicar más filtros para evitar falsos positivos. De hecho, los falsos positivos producto de los patrones, han motivado estos tres primeros filtros, con posibilidad de agregar más en el futuro. Es decir, se ha elegido abstraer esta función por motivos de escalabilidad.

Ahora, pasaremos a analizar los diagramas de cada una de las reglas, para apreciar cómo se han implementado cada uno de los patrones descritos.

iii. Función regla1(candidato, parrafo)



Notaremos que la regla 0 se ha incluido en la programación de la regla 1. Primero, nos aseguramos de que el candidato no tenga más de 10 palabras, y después llamamos a la función `obtenerArticulo(candidato)` que devuelve una cadena con un artículo, y la primera palabra en mayúscula dentro del candidato a alias.

Así, si partimos artículo en dos elementos, el primero (índice 0) corresponde al artículo, y el segundo (índice 1) corresponde a la palabra en sí. Es por esto que a la función `reglasSiglas`, se le pasa como parámetro la palabra, y no el artículo.

`reglasSiglas()` nos dirá si la palabra son siglas o no. Es decir, revisaremos si en la palabra hay mayor cantidad de mayúsculas que de minúsculas. Así, podremos decidir qué hacer, si aplicar la regla 0, o continuar con los patrones de la regla 1; en ambos casos, regresamos la entidad encontrada.

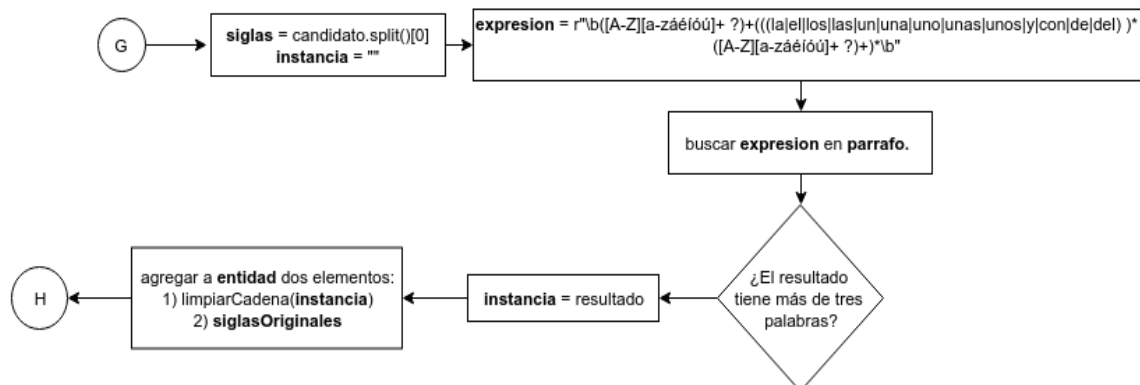
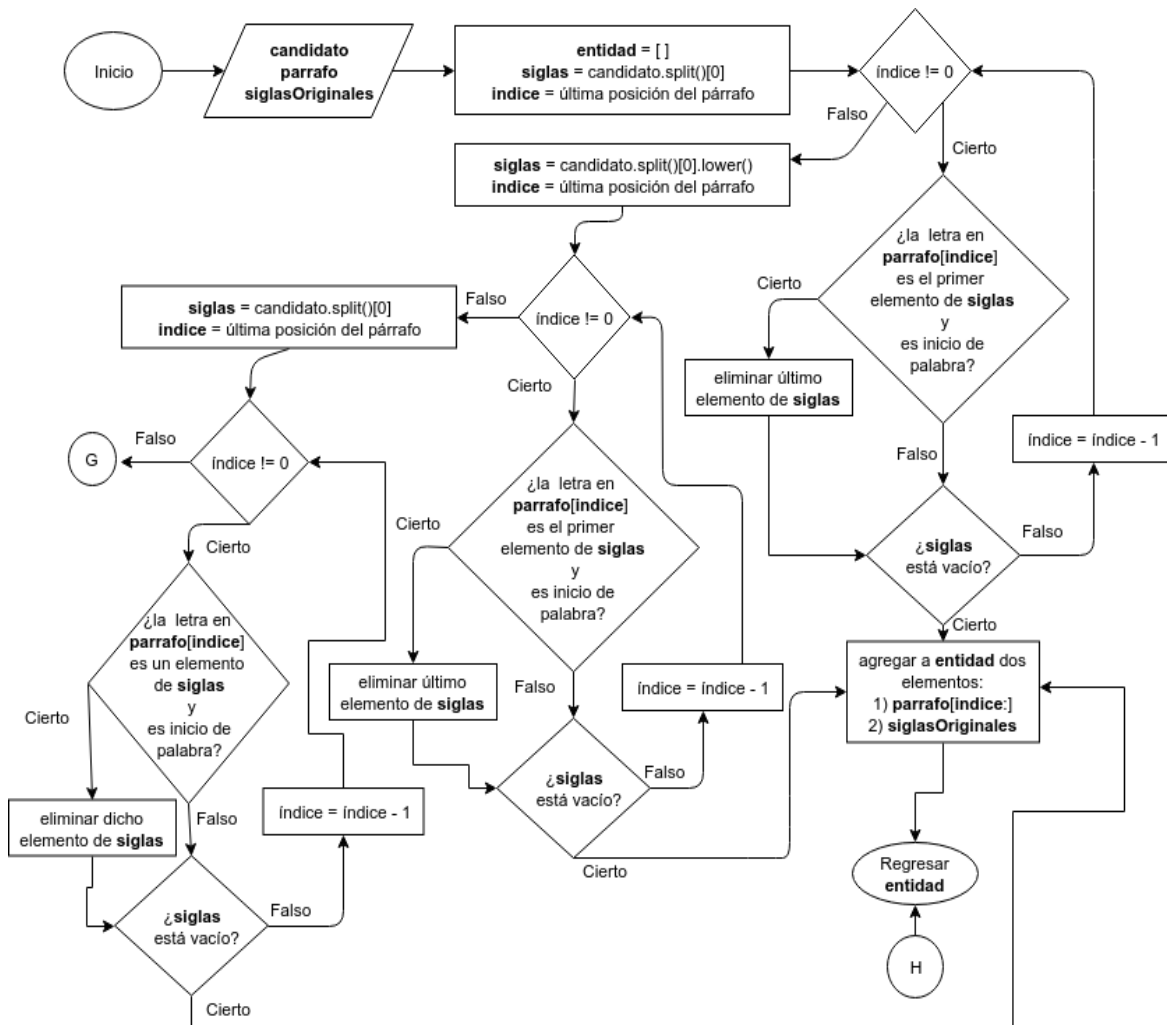
la función `resolverSiglas(siglas,parrafo,siglasOriginales)` contiene la implementación de la regla 0. En general, busca hacia atrás, los cuatro casos ya descritos:

- 1) Buscar mayúsculas en orden

- 2) Buscar minúsculas en orden
- 3) Buscar mayúsculas en desorden
- 4) Buscar con expresión regular

y su diagrama es el siguiente:

iv. Función resolverSiglas(siglas,parrafo,siglasOriginales)



Observamos en el diagrama, a primera vista, tres rombos pequeños de decisión que verifican el valor de índice distinto de 0. Esta decisión marca el inicio de la búsqueda hacia atrás de cada una de las tres búsquedas. El rombo superior, así como lo que de él se desencadena, corresponde a la búsqueda 1), el rombo de enmedio corresponde a la búsqueda 2), el tercer rombo corresponde a la búsqueda 3). Finalmente, el conector G, y lo que desencadena, corresponde con la búsqueda del punto 4).

Para los primeros dos casos, cada letra de las siglas se introduce, de izquierda a derecha, en una cola donde el primero en entrar es el primero en salir (siglas). Conforme se realiza la búsqueda hacia atrás, y se van encontrando las letras correspondientes, se van sacando de la cola, hasta que la misma se vacía. Es entonces donde se considera que se han encontrado las siglas y se devuelve el resultado.

Para la implementación de la búsqueda en desorden se utiliza sólo una lista con cada letra de las siglas, y para cada letra inicial en mayúscula encontrada hacia atrás en el contexto se busca su correspondiente en las siglas; si se encuentra, se elimina. Así, cuando la lista queda vacía, se considera que se encontraron las siglas y se regresa el resultado.

Finalmente, la búsqueda con expresión regular se realiza sobre el contexto, con la siguiente expresión:

```
\b([A-Z][a-záéíóú]+ ?)+(((la|el|los|las|un|una|uno|unas|unos|y|con|de|del) )*([A-Z][a-záéíóú]+ ?)+)*\b
```

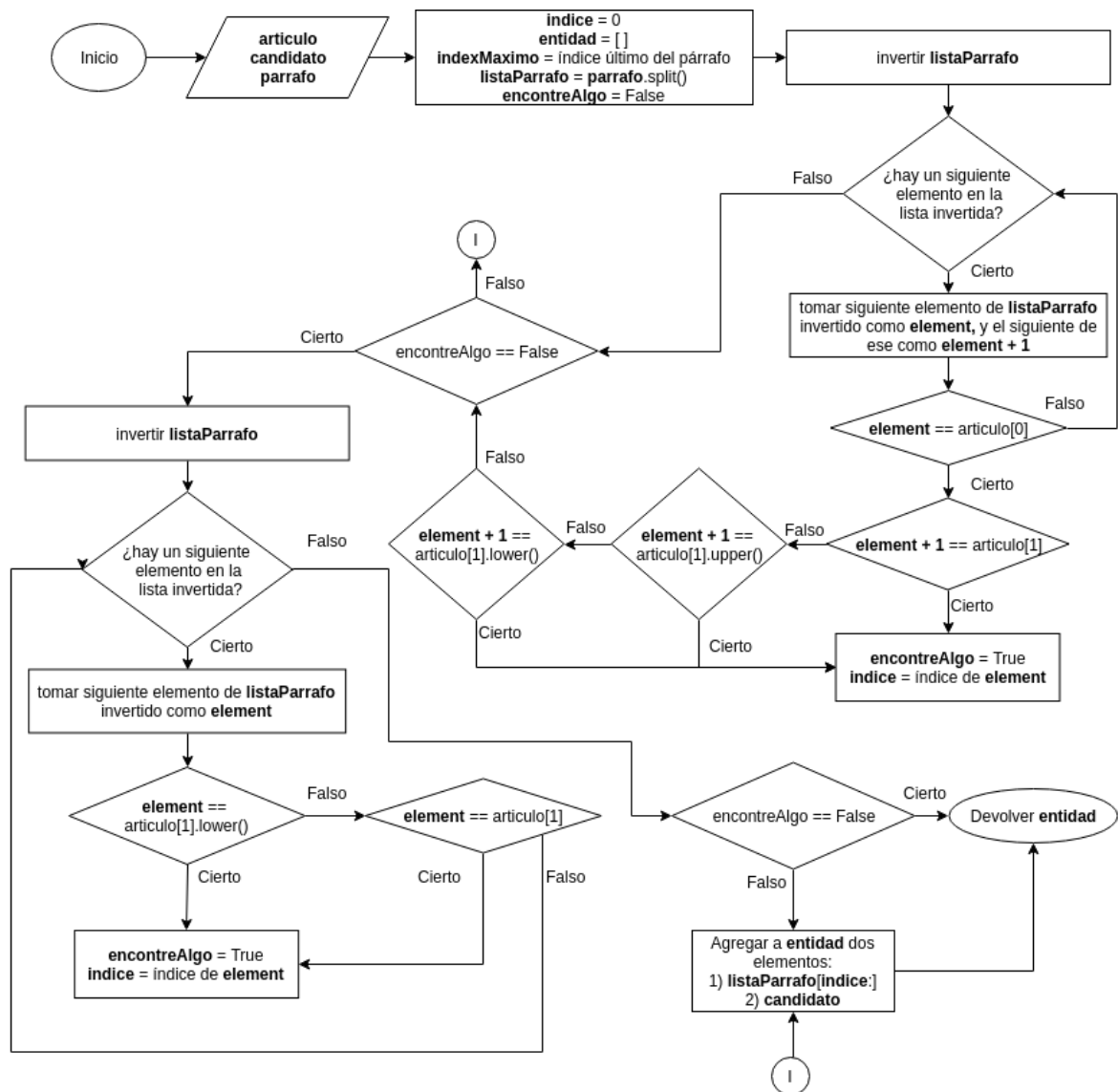
la cual devuelve lo requerido según el patrón establecido en la sección correspondiente. Es así como esta función implementa a la regla 0.

Ahora, la función buscarArticulo(**artículo, candidato, parrafo**) implementa los patrones:

- 1) artículo + entidad + candidato a alias
- 2) entidad + candidato a alias

ya descritas por la regla 1 anteriormente. A continuación se muestra su respectivo diagrama de flujo.

v. Función buscarArticulo(articulo,candidato,parrafo)



En este diagrama, lo primero que debemos notar son los dos rectángulos que dicen “invertir **listaParrafo**”, lo que viene después de cada uno de ellos, representa al patrón 1) y 2) respectivamente.

En el primer recuadro, se parte el párrafo (contexto) en una lista de palabras, la cual se invierte, y se recorre, elemento por elemento, buscando el artículo. Al encontrarse, se revisa si la palabra anterior (es decir, en el texto, la siguiente) es la palabra que acompaña al artículo, y que se obtuvo de la función obtenerArticulo(**candidato**). Si esto no es así, entonces se ve si la palabra es la misma pero en mayúscula; si tampoco es así, se revisa si es la misma en minúscula. De ser cierto cualquiera de estos tres casos, el índice donde ha sido cierto se guarda con el fin de saber dónde empieza la entidad, dentro del párrafo.

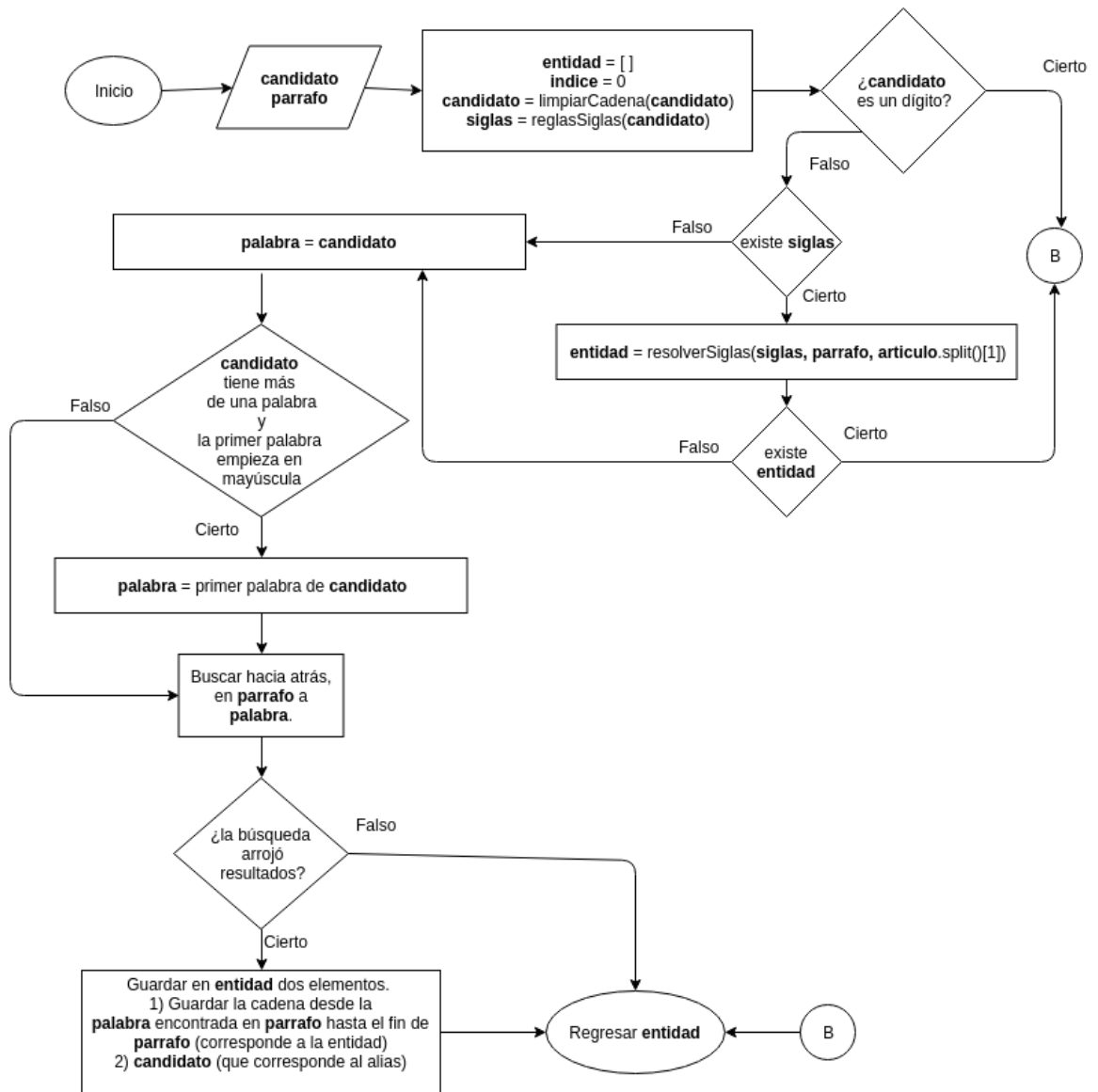
Si estos criterios fueron falsos, se vuelve a recorrer el párrafo invertido, pero ahora sólo se busca la palabra, omitiendo el artículo. Se revisa si la palabra está en minúscula, o tal cual como viene escrita en **articulo[1]** (palabra que acompaña al artículo); de ser cierto, se guarda el índice para indicar el inicio de la entidad.

Si el primer proceso resulta verdadero, una bandera llamada **encontreAlgo** se levanta, y hace las asignaciones a **entidad** correspondientes. De igual forma, el segundo proceso levanta la misma bandera para al final poder validar qué se regresará al usuario, si el valor de **entidad** vacío como se declaró al inicio, o el valor de **entidad** dado por el índice.

Hasta ahora, hemos visto que la función `MainNERAlias()` y `aplicarReglasAlias(candidato, parrafo, fname, indiceOcuencia)` han servido para iniciar el procesamiento de las reglas. Después, las funciones `regla1(candidato,parrafo)`, `resolverSiglas(siglas,parrafo,siglasOriginales)` y `buscarArticulo(articulo,candidato,parrafo)`, de forma general, han implementado los patrones de la regla 0 y la regla 1.

A continuación veremos el funcionamiento de la función `regla2(candidato,parrafo)`, la cual implementa el patrón de la Regla 2.

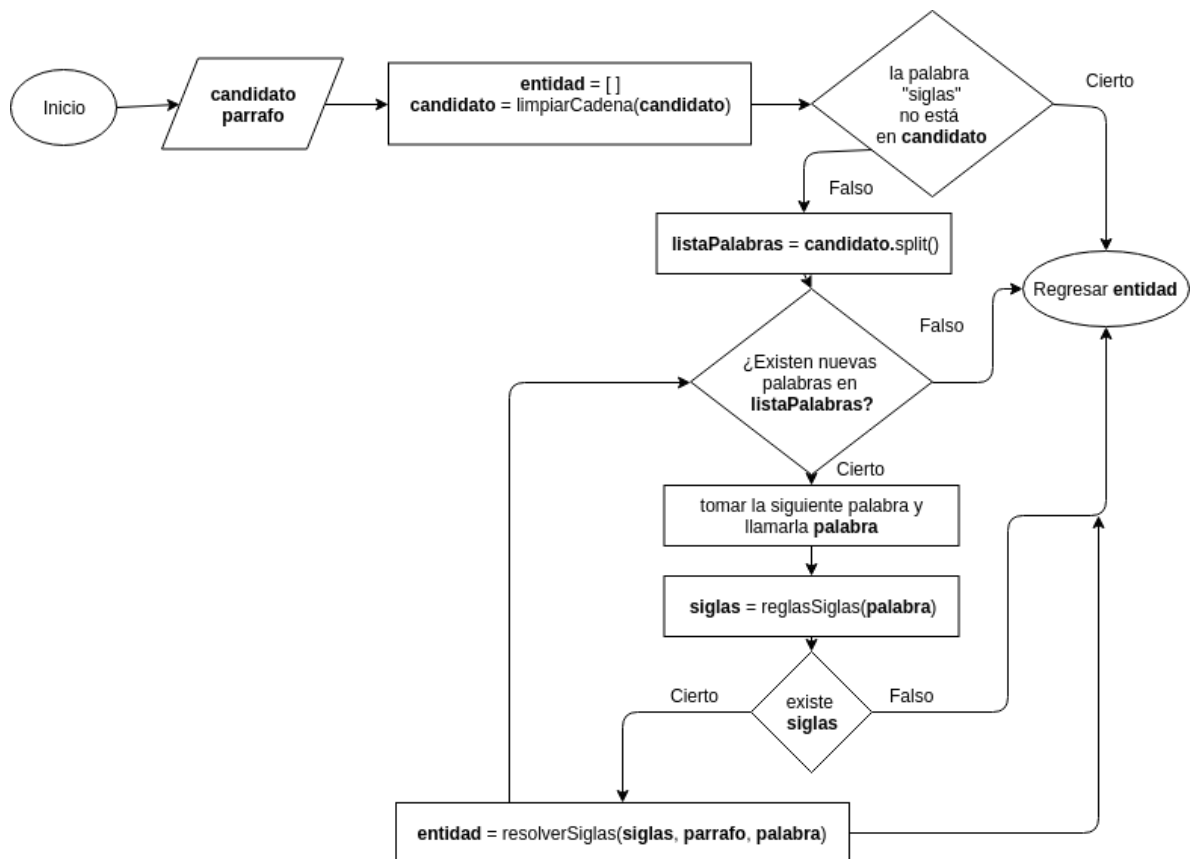
vi. Función regla2(candidato,parrafo)



En esta función, lo primero que hacemos es filtrar que el candidato no sea un dígito, pues al ser el mismo, una sola palabra, a veces se encontraban numeraciones. Después, se valida si la palabra son siglas o no, si sí, entonces se utiliza `resolverSiglas(siglas, parrafo, siglasOriginales)` que aplica la regla 0. Si no son siglas, se toma la primer palabra del candidato como la palabra a buscar hacia atrás en el párrafo, y si se encuentra, se guarda la entidad y su respectivo alias.

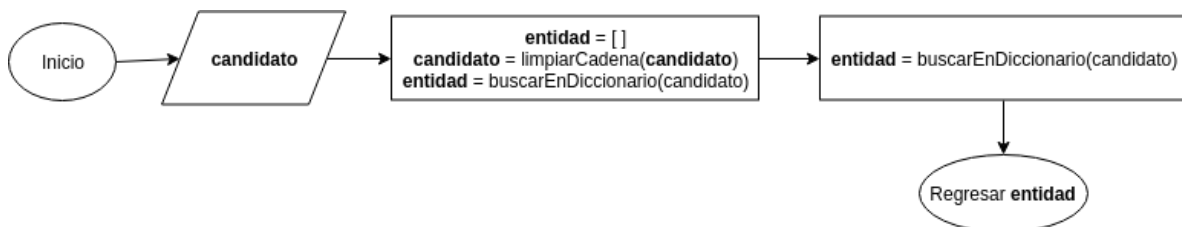
Pasemos ahora a revisar la tercera regla.

vii. Función regla3(candidato,parrafo)



De este diagrama se observa el algoritmo para, primero, verificar si existe la palabra "siglas" en el candidato, y luego buscar dichas siglas para resolverlas mediante la regla 0. Es decir, la regla 3 es, realmente, una forma de encontrar casos específicos donde sabemos que la regla 0 puede actuar correctamente.

viii. Función regla4(candidato,parrafo)



Finalmente, la regla 4 es implementada por una función llamada buscarEnDiccionario(candidato), la cual lo único que hace es abrir un documento en formato .csv, que contiene entidades y alias, para buscar al candidato en ella y saber si se trata de una entidad.

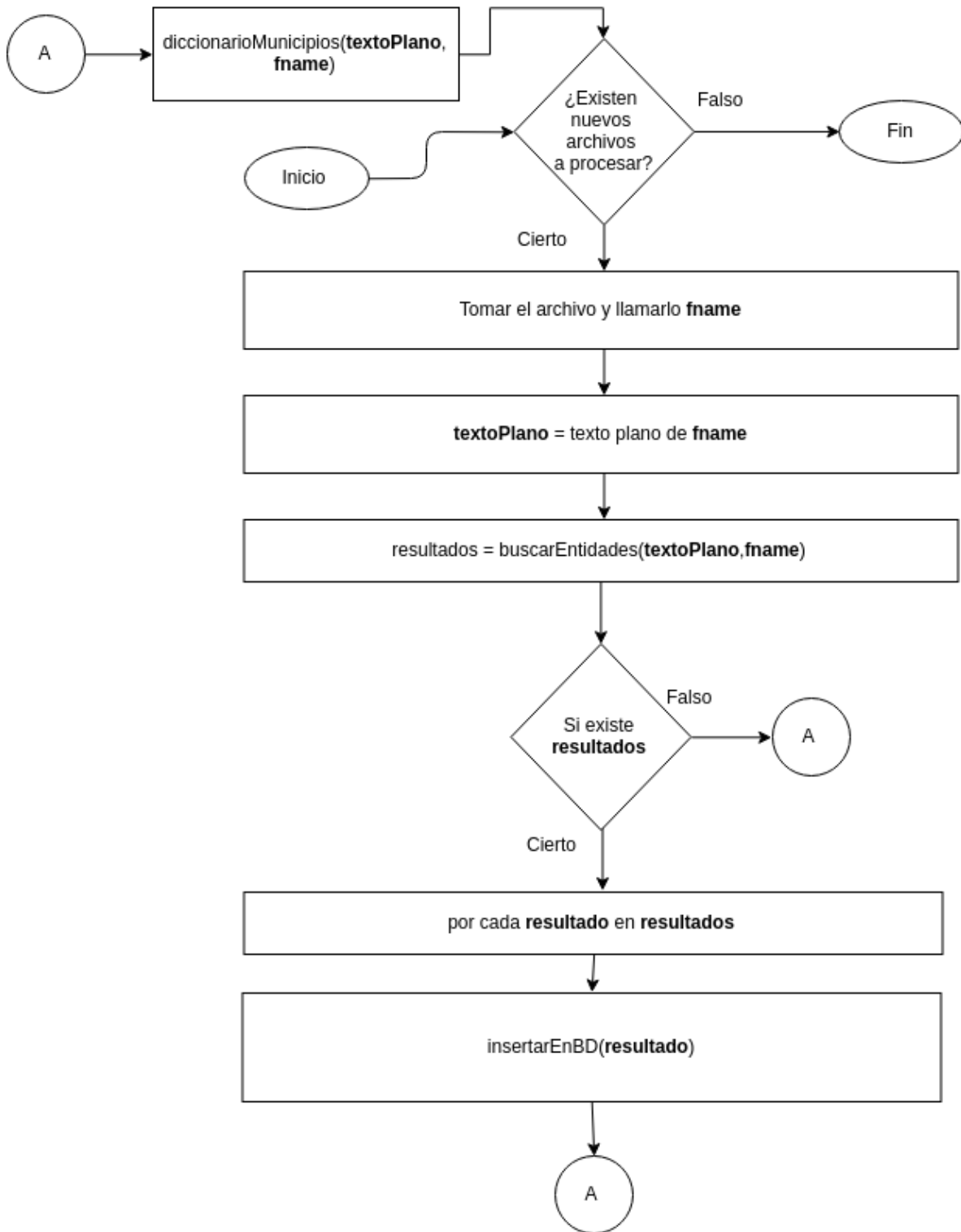
Se quiso abstraer esta función, previendo la posibilidad de usar dicha función más adelante, sin tener que llamar explícitamente a la función “regla4”, así como la posibilidad de agregar más diccionarios.

A continuación hablaremos sobre la implementación de los patrones para encontrar entidades que no tienen alias.

II. Implementación de patrones para entidades sin alias

Esta labor fue descrita en términos de tres reglas para encontrar candidatos a entidad, y luego en algunos criterios para reconocer los candidatos que sí son entidades. Primero, la función `MainNER()` dispara todo el proceso, luego la función `buscarEntidades(texto, fname)` verifica las 3 reglas para encontrar candidatos, y la función `ReglasNER(candidato, fname)` aplica los criterios para filtrar a los candidatos, cuando se ha creído necesario.

Veamos entonces, los diagramas para estas tres funciones, empezando por `MainNER()`.

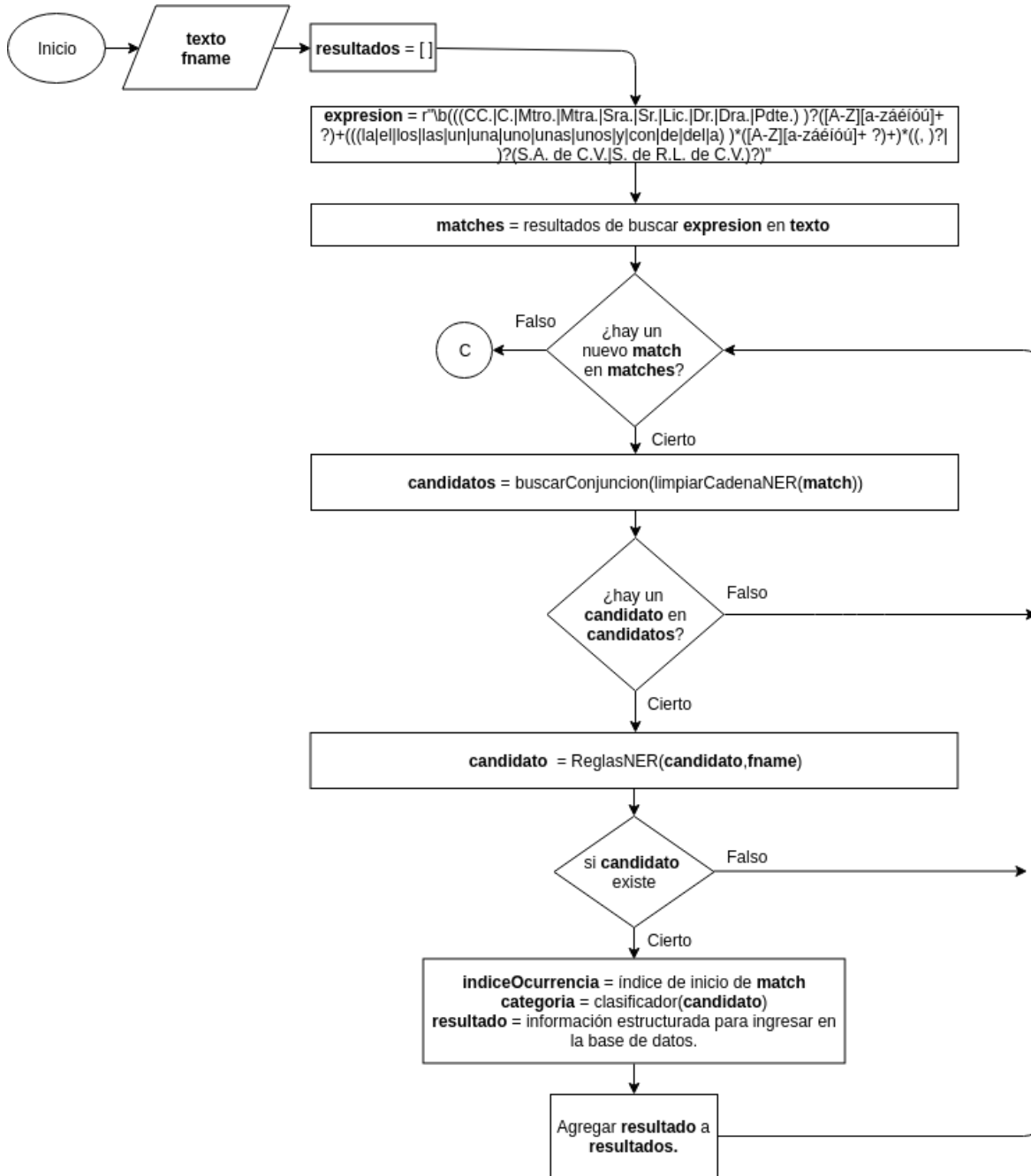
i. Función MainNER()

De esta función se destaca lo mismo que en `MainNERalias()`, esto es, que los archivos que abre están contenidos en una carpeta, y en una sola corrida, se encarga de analizarlos todos. También podemos ver hasta el final la labor de insertar en la base de datos, de la cual también hablaremos más adelante.

Después, siempre antes de regresar a buscar si hay un archivo nuevo, se busca al candidato en un diccionario que contiene municipios y estados de la república.

Ahora veamos cómo encontrar candidatos a entidad con la función `buscarEntidades(texto, fname)`

ii. Función buscarEntidades(texto,fname)



Esta función se divide en dos diagramas, en este primero se aplica el primer criterio para encontrar candidatos, y utiliza la expresión regular:

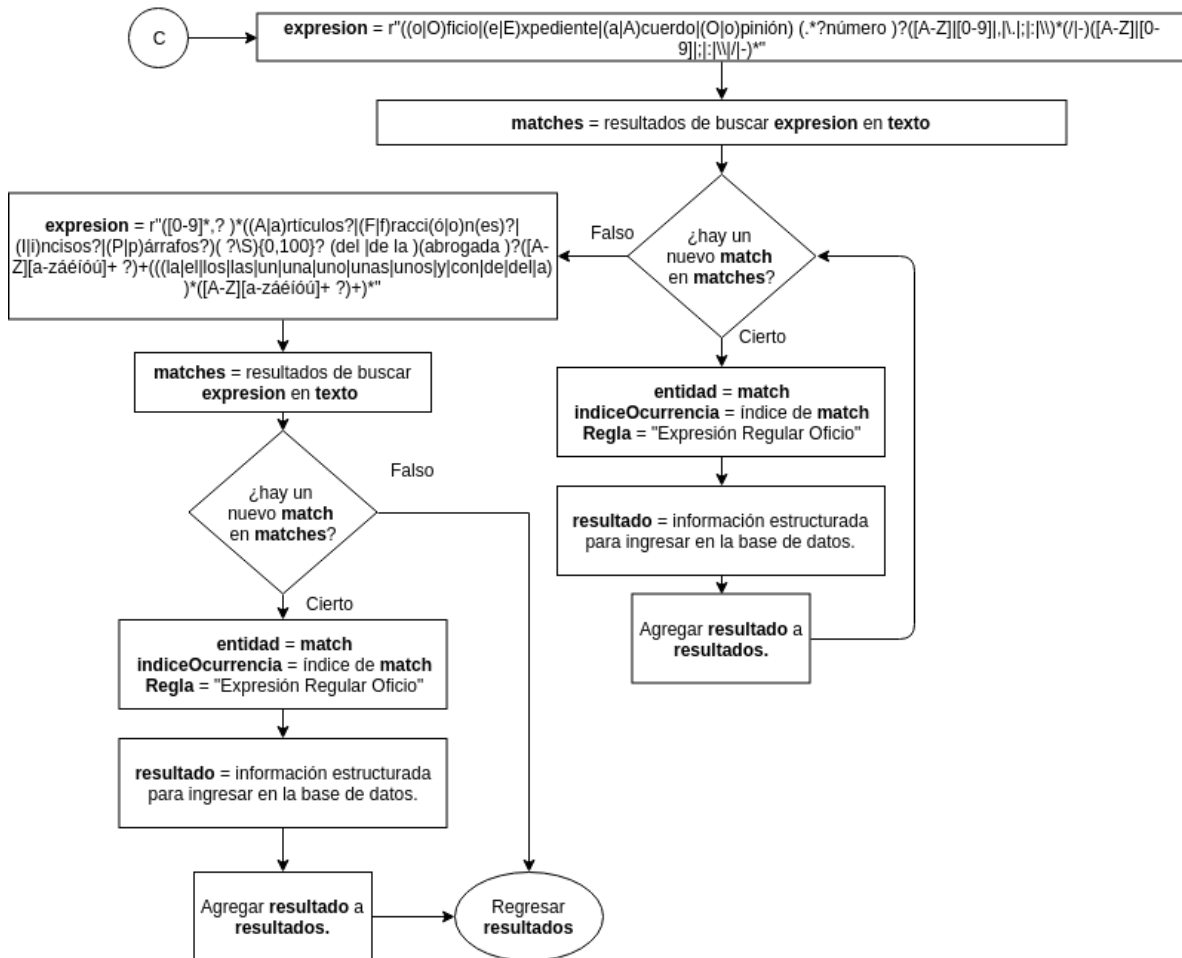
```

\b(((CC.|C.|Mtro.|Mtra.|Sra.|Sr.|Lic.|Dr.|Dra.|Pdte.) ?)[A-Z][a-záéíóú]+
?)+(((|a|e|l|os|las|un|una|uno|unas|unos|y|con|de|del|a) *)*[A-Z][a-záéíóú]+
?)+*((,| ) ?) ?(S.A. de C.V.|S. de R.L. de C.V.)?)

```

Los resultados de esta expresión podrían traer dos entidades distintas separadas por las cadenas “y”, “y la”, “y el” y “de la”, por lo que es necesaria una función buscarConjuncion(**candidato**) que busque dichos casos, y determinar si el candidato contiene realmente dos candidatos a entidad en lugar de uno. Después, para cada candidato a entidad encontrada, se determina si es o no, una entidad, con la función ReglasNER(**candidato**,**fname**).

Para aplicar el criterio 2) y 3) para encontrar candidatos, esta función continua en el siguiente diagrama:



Se busca ahora candidatos a entidad con la expresión regular:

`((o|O)fficio|(e|E)xpediente|(a|A)cuerdo|(O|o)pinión) (. *?número) ?([A-Z][0-9],|\.,|:|\|)*(/-)([A-Z][0-9];:|\|/|)*`

Se ha considerado que todo lo que devuelve la misma, es una entidad, pues el patrón definido que viene después de las palabras “oficio”, “expediente”, “acuerdo” es muy específico, al devolver patrones alfanuméricos separados por diagonales o guiones medios.

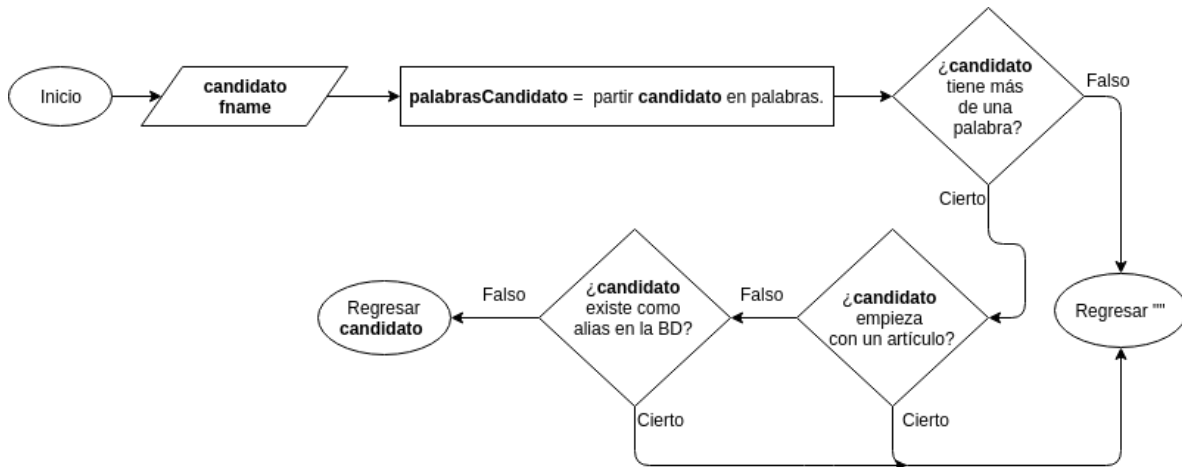
Después, para implementar el criterio número 3, se aplica la expresión regular:

$([0-9]^*, ?)^*((A|a)rtículos?(F|f)racci(ó|o)n(es)?|(I|i)ncisos?(P|p)árrafos?)(?\S){0,100}? (del |de la)(abrogada)?([A-Z][a-záéíóú]+ ?)^+(((a|e|l|los|las|un|una|uno|unas|unos|y|con|de|del|a))^*([A-Z][a-záéíóú]+ ?)^+)^*$

Al igual que la expresión regular anterior, se ha determinado que esta devuelve pocos falsos positivos, por lo que no se aplica ningún otro filtro.

Analicemos ahora la última función `ReglasNER(candidato, fname)`

iii. Función `ReglasNER(candidato, fname)`



Observamos que existen tres criterios que filtran la mayor parte de basura de nuestros candidatos a entidad. El primero elimina todos los candidatos de una palabra, lo cual elimina municipios, y algunos resultados correctos. Sin embargo, se consideró que la ausencia de esta regla traía más falsos positivos que su aplicación.

Después, eliminamos aquellos candidatos que empiezan con un artículo, y finalmente, buscamos si lo que hemos encontrado aquí no se ha encontrado ya como un alias por el reconocedor de patrones implementados anteriormente.

Para mayor información sobre el programa, puede consultarse el código fuente del mismo, el cual está acompañado de comentarios en cada función, como apéndice en esta tesis.

A continuación hablaremos sobre la estructura de datos que hemos usado para guardar la información en la base de datos.

III. Implementación de la base de datos

Primero, el manejador que hemos usado es MongoDB, lo que nos permite utilizar una estructura basada en archivos JSON.

La función `insertarEnBD(resultado)` se encarga de hacer la inserción de cada resultado en la base de datos, utilizando la librería PyMongo que conecta a mongoDB con Python.

La estructura de cada documento en la base de datos es la siguiente:

```
{
  "Nombre": resultadoEntidad[0],
  "Clase": categoria,
  "Archivos":
  {
    "Nombre":fname.replace(".docx",""),
    "indiceOcurrencia": indiceOcurrencia,
    "Alias": resultadoEntidad[1],
    "Regla": Regla
  }
}
```

En el campo "Nombre" se guarda el nombre de la entidad en sí.

El campo "clase" puede contener una de las siguientes categorías: Persona, Organización, Ley, Documento, Lugar, u Otro.

Después, el campo "Archivo" se compone de cuatro campos. El primero, el nombre del archivo donde fue encontrado el alias. Después, el índice, dentro del documento, donde se encontró la entidad. Luego, el alias que se le ha puesto a la entidad, y finalmente el número de regla que lo ha encontrado.

Si se encuentra la misma entidad en varios archivos, al campo "Archivos" se le agrega otra 4-tupla con la nueva información, de modo que de una sola entidad, podemos consultar en cuantos archivos aparece, cuales son los alias con los que se le denomina, la regla que lo encontró, y, lo más importante, el índice respecto a cada documento, del lugar donde se encuentra cada entidad.

Así mismo, el campo "indiceOcurrencia" guarda todos los índices de las veces que ocurre la entidad, dentro del documento. En el caso de entidades con alias, sólo se guarda el índice donde se ha visto a la entidad y no a los alias. A pesar de esto, la información provista es suficiente para resolver la referencia, si así se requiriera.

Para las entidades que se encuentran sin alias, simplemente se deja el campo vacío.

Ahora pasemos a la implementación realizada del detector de eventos.

IV. Implementación de patrones para reconocimiento de eventos

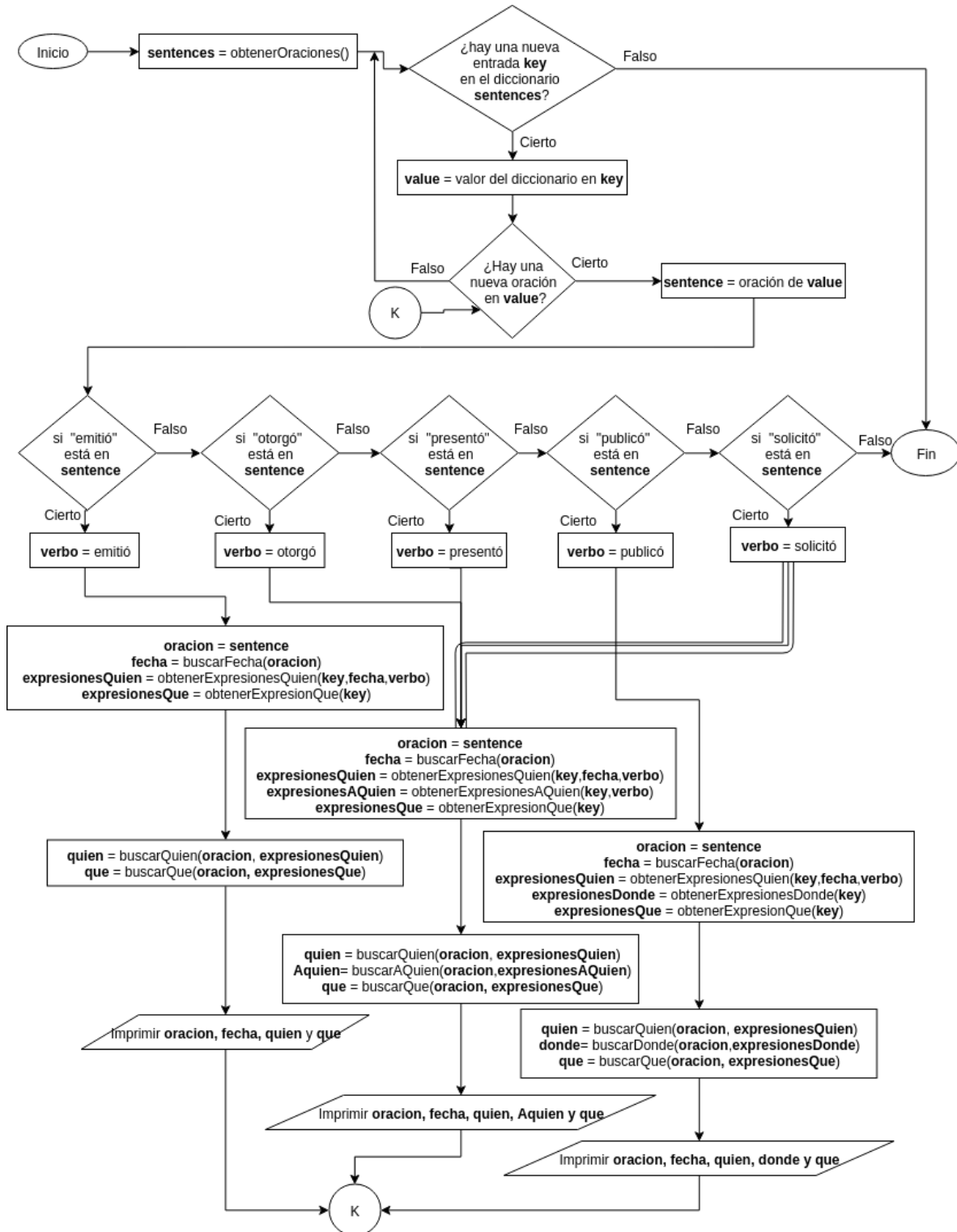
El programa fue escrito, al igual que el reconocedor de entidades, en lenguaje Python, y es significativamente menos complejo que el reconocedor de entidades. Por esta razón, en esta sección nos limitaremos a mostrar un diagrama general, hacer descripciones breves de cada una de las funciones, y mencionar las expresiones regulares utilizadas.

i. Funciones

Las funciones en el programa son:

- 1) `freeling()`
- 2) `obtenerOraciones()`
- 3) `buscarQuien(sentence,expresionesQuien)`
- 4) `buscarFecha(sentence)`
- 5) `buscarAQuien(sentence,expresionesAQuien)`
- 6) `buscarQue(sentence,expresionesQue)`
- 7) `buscarDonde(sentence,expresionesDonde)`
- 8) `obtenerExpresionQuien(fname,fecha,verbo)`
- 9) `obtenerExpresionAQuien(fname,verbo)`
- 10) `obtenerExpresionQue(fname)`
- 11) `obtenerExpresionDonde(fname)`

El diagrama de flujo del programa, y el accionar de estas funciones, es el siguiente:



5.6 Diagrama de flujo para detectar eventos

De este diagrama se rescata que las oraciones que se obtienen son diccionarios, donde la llave es el nombre del documento que se está analizando, y el valor, una lista con las oraciones de ese documento que contienen una fecha y alguno de los cinco verbos.

Después, dependiendo del verbo que esté presente en la oración, se procede a obtener expresiones regulares para cada actuante, y después a aplicarlas e imprimir los resultados.

A continuación daremos una descripción de cada función:

freeling(): Esta función accede a los documentos en una carpeta, y los analiza utilizando FreeLing 4.0, con el fin de poner a disposición del resto del programa archivos analizados que contienen información sintáctica sobre el contenido.

obtenerOraciones(): Se encarga de obtener un diccionario que contiene, para cada archivo analizado, una lista de oraciones que tienen fecha y uno de los cinco verbos elegidos.

obtenerExpresionQuien(fname,fecha,verbo): Accede, en la base de datos, a todas las entidades y alias detectados en el documento **fname** cuya clase sea “Persona” u “Organización”. Dichas entidades y alias se concatenan con el operador *or* (*|*) en una expresión regular **cadenaEntidades**, la cual a su vez se utiliza en las siguientes expresiones regulares:

- 1) "\brepresentante legal de .*? ?" + "(" + **cadenaEntidades** + ")"
- 2) "\bpor medio de .*? ?" + "(" + **cadenaEntidades** + ")"
- 3) "\ba través de .*? ?" + "(" + **cadenaEntidades** + ")"
- 4) "\b" + **fecha** + " .*? " + "mediante .*? ?" + "(" + **cadenaEntidades** + ")" + " .*? " + **verbo**
- 5) "\b" + **fecha** + " .*? ?" + "(" + **cadenaEntidades** + ")" + " .*? " + **verbo**

La función devuelve una lista **expresionesQuien** con cada una de las cinco expresiones regulares obtenidas.

obtenerExpresionAQuien(fname,verbo): Accede, en la base de datos, a todas las entidades y alias detectados en el documento **fname** cuya clase sea “Persona”, “Organización” o “Documento”. Dichas entidades y alias se concatenan con el operador *or* (*|*) en una expresión regular **cadenaEntidades**, la cual a su vez se utiliza en las siguientes expresiones regulares:

- 1) "\bante .*? ?" + "(" + **cadenaEntidades** + ")"
- 2) "\ba quien .*? ?" + "(" + **cadenaEntidades** + ")"
- 3) "\ben favor de .*? ?" + "(" + **cadenaEntidades** + ")"
- 4) "\b" + **verbo** + " a .*? " + "(" + **cadenaEntidades** + ")"

La función devuelve una lista **expresionesAQuien** con cada una de las cuatro expresiones regulares obtenidas.

obtenerExpresionQue(fname): Accede, en la base de datos, a todas las entidades y alias detectados en el documento **fname** cuya clase sea “Persona” o “Documento”. Dichas entidades y alias se concatenan con el operador *or* (/) en una expresión regular **cadenaEntidades**, la cual a su vez se utiliza en las siguientes expresiones regulares:

- 1) "\b(opinión .*? ?" + **cadenaEntidades** + ")"
- 2) "\b(opinión .*?)" + **cadenaEntidades**
- 3) "\b" + **verbo** + " .*?" + "(" + **cadenaEntidades** + ")"

La función devuelve una lista **expresionesQue** con cada una de las tres expresiones regulares obtenidas.

btenerExpresionDonde(fname): Accede, en la base de datos, a todas las entidades y alias detectados en el documento **fname** cuya clase sea “Persona” o “Documento”. Dichas entidades y alias se concatenan con el operador *or* (/) en una expresión regular **cadenaEntidades**, la cual a su vez se utiliza en la expresión regular:

- 1) "\ben .*? ?" + "(" + **cadenaEntidades** + ")"

La función devuelve una lista **expresionesDonde** que sólo contiene la expresión regular expuesta.

buscarFecha(sentence): Utiliza FreeLing 4.0 para analizar la oración contenida en **sentence** y devuelve la fecha detectada en la variable **fecha**

buscarAQuien(sentence,expresionesAQuien): Busca las expresiones regulares de **expresionesAQuien** en la oración **sentence** y devuelve el resultado.

buscarQue(sentence,expresionesQue): Busca las expresiones regulares de **expresionesQue** en la oración **sentence** y devuelve el resultado.

buscarQuien(sentence,expresionesQuien): Busca las expresiones regulares de **expresionesQuien** en la oración **sentence** y devuelve el resultado.

buscarDonde(sentence,expresionesDonde): Busca las expresiones regulares de **expresionesDonde** en la oración **sentence** y devuelve el resultado.

También, cada vez que de la oración analizada se extrae la respuesta a una pregunta, dicha entidad, o dicha cadena, es eliminada antes de volver a ser procesada para buscar otro actuante, así nos aseguramos de no encontrar la misma entidad como respuesta a distintas preguntas, lo cual no tendría sentido.

V. Consideraciones del software

Respecto a todo el software desarrollado, se ha probado su funcionamiento en el sistema operativo Ubuntu 16.04.4 LTS, corriendo Python 3.5.2 con las siguientes librerías de terceros debidamente instaladas haciendo uso del comando *pip install*.

- a) docx2txt
- b) pymongo

Para la base de datos, se utilizó el manejador MongoDB shell version 3.6.5.

Además de la instalación de las librerías de terceros, y el manejador de la base de datos, no se requiere hacer alguna instalación del software que se ha desarrollado aquí.

Para correr el programa, es necesario colocarse dentro de la carpeta que lo contiene y correr el comando:

```
python main.py
```

el cual ejecuta un script que pone en marcha `MainNERalias()` y `MainNER()`. Es importante estar corriendo la versión 3 de Python, de lo contrario se pueden encontrar errores en la codificación de caracteres, así como en la sintaxis del programa.

Para la visualización de los datos dentro de la base de datos, recomendamos Robo 3T - 1.2. Este permite, de forma amigable, hacer queries y tener representaciones de las bases de datos en MongoDB guardadas en la computadora.

Finalmente, para correr el reconocedor de eventos, se utiliza el comando:

```
python eventos.py
```

Esto ejecuta el script mostrado en el diagrama de flujo de la sección *Implementación de patrones para reconocimiento de eventos*. Es importante estar corriendo la versión 3 de Python, igual que en el reconocedor de entidades nombradas.

Es necesario, tener una instancia de FreeLing 4.0 corriendo como servidor en el puerto 50005. Para ello, primero debe instalarse FreeLing, lo cual puede hacerse siguiendo su documentación¹⁷ y después corriendo el comando:

```
analyze -f es.cfg --server --port 50005
```

¹⁷ <https://talp-upc.gitbooks.io/freeling-4-0-user-manual/content/installation.html>

el cual inicia freeling, con el archivo de configuración para español (es.cfg), en modo servidor, esperando instrucciones en el puerto 50005.

Esto es debido a que en el script de eventos, utilizamos FreeLing para detectar las fechas en las oraciones.

Teniendo estas consideraciones, pasemos ahora a probar las implementaciones, y observemos sus resultados, probados en el corpus.

c) Experimentos

En esta sección daremos cuatro ejemplos para la implementación de entidades con alias, tres ejemplos para la implementación de entidades sin alias y cinco ejemplos de eventos obtenidos.

Cada ejemplo contiene, al menos, una breve explicación y la entrada en la base de datos obtenida. Empecemos entonces con las entidades sin alias.

I. Entidades con alias

Observaremos y hablaremos sobre cuatro ejemplos. donde cada uno corresponde a cada regla expuesta. El ejemplo tres, expone la regla tres y la regla 0, pues la regla 3 es una forma, como ya explicamos, de encontrar casos concretos donde se aplica la regla 0. Empecemos.

i. Ejemplo 1

“Por su parte, cabe destacar que para este tipo de solicitudes deben acatarse los requisitos de procedencia establecidos en el Acuerdo de Pleno número P/IFT/EXT/131114/228 de fecha 13 de noviembre de 2014, “Acuerdo mediante el cual el Pleno del Instituto Federal de Telecomunicaciones fija el monto de los aprovechamientos que deberán cobrarse por la prestación de diversos servicios públicos en el ejercicio de sus funciones de derecho público por los que no se establece monto específico en la Ley Federal de Derechos” (el “Acuerdo de Pago de Aprovechamientos”).”

En este ejemplo, se encontró “Acuerdo de Pago de Aprovechamientos” como alias de “Acuerdo mediante el cual el Pleno del Instituto Federal de Telecomunicaciones fija el monto de los aprovechamientos que deberán cobrarse por la prestación de diversos servicios públicos en el ejercicio de sus funciones de derecho público por los que no se establece monto específico en la Ley Federal de Derechos”, por medio de la Regla 1. Es decir, se buscó la cadena “el Acuerdo” hacia atrás en el párrafo, y donde se encontró, se tomó de ahí en adelante para encontrar la entidad.

La entrada en la base de datos es la siguiente:

```
"_id" : ObjectId("5b22eddd219ef328ff95a92e"),
  "Archivos" : [
    {
      "indiceOcurrecencia" : 8634,
      "Alias" : "Acuerdo de Pago de Aprovechamientos",
      "Regla" : "Regla 1",
      "Nombre" : "P_IFT_170316_109_Acc"
    },
    {
      "Regla" : "Regla 1",
      "Alias" : "Acuerdo de Pago de Aprovechamientos",
      "indiceOcurrecencia" : 9352,
      "Nombre" : "P_IFT_170216_49_Acc"
    },
    {
      "indiceOcurrecencia" : 8620,
      "Alias" : "Acuerdo de Pago de Aprovechamientos",
      "Regla" : "Regla 1",
      "Nombre" : "P_IFT_111215_562_Acc"
    },
    {
      "Regla" : "Regla 1",
      "Alias" : "Acuerdo de Pago de Aprovechamientos",
      "indiceOcurrecencia" : 3588,
      "Nombre" : "P_IFT_170216_50_Acc"
    },
    {
      "indiceOcurrecencia" : 3817,
      "Alias" : "Acuerdo de Pago de Aprovechamientos",
      "Regla" : "Regla 1",
      "Nombre" : "P_IFT_170216_51_Acc"
    },
    {
      "Regla" : "Regla 1",
      "Alias" : "Acuerdo de Pago de Aprovechamientos",
      "indiceOcurrecencia" : 8490,
```

```

    "Nombre" : "P_IFT_170316_106_Acc"
  },
  {
    "indiceOcurrencia" : 8695,
    "Alias" : "Acuerdo de Pago de Aprovechamientos",
    "Regla" : "Regla 1",
    "Nombre" : "P_IFT_170216_53_Acc"
  },
  {
    "Regla" : "Regla 1",
    "Alias" : "Acuerdo de Pago de Aprovechamientos",
    "indiceOcurrencia" : 8312,
    "Nombre" : "P_IFT_090316_81_Acc"
  },
  {
    "indiceOcurrencia" : 8810,
    "Alias" : "Acuerdo de Pago de Aprovechamientos",
    "Regla" : "Regla 1",
    "Nombre" : "P_IFT_170316_110_Acc"
  },
  {
    "Regla" : "Regla 1",
    "Alias" : "Acuerdo de Pago de Aprovechamientos",
    "indiceOcurrencia" : 9160,
    "Nombre" : "P_IFT_090316_79_Acc"
  }
],
  "Nombre" : "Acuerdo de Pleno número P/IFT/EXT/131114/228 de fecha 13 de noviembre de
2014 Acuerdo mediante el cual el Pleno de el Instituto Federal de Telecomunicaciones fija el monto
de los aprovechamientos que deberán cobrarse por la prestación de diversos servicios públicos en
el ejercicio de sus funciones de derecho público por los que no se establece monto específico en la
Ley Federal de Derechos",
  "Clase" : "Documento"
}

```

Observamos que esta entidad se encuentra en varios documentos, en todos ellos se encontró usando la regla 1. Recordemos también que el índice de ocurrencia, en el caso de entidades con

alias, sólo guarda el índice donde se ha visto la entidad, y no el alias. Se recalca también que se ha clasificado como “Documento”, de forma correcta.

ii. Ejemplo 2

“III.- AT&T Comunicaciones Digitales, S. de R.L. de C.V., (en lo sucesivo “AT&T Comunicaciones Digitales”)”

En este caso, el contexto es muy pequeño, pues la entidad se encuentra al inicio del párrafo. La regla 2 eliminó del paréntesis las palabras “en lo sucesivo”, y busco la cadena “AT&T” hacia atrás para encontrar a la entidad. A continuación la entrada en la base de datos:

```
{
  "_id" : ObjectId("5b22ede2219ef328ff95aa08"),
  "Archivos" : {
    "Regla" : "Regla 2",
    "Alias" : "AT&T Comunicaciones Digitales",
    "indiceOcurrencia" : 1754,
    "Nombre" : "P_IFT_310118_31_AccUPR"
  },
  "Nombre" : "AT&T Comunicaciones Digitales S. de R.L. de C.V.",
  "Clase" : "Organización"
}
```

Observamos en la entrada de la base de datos que se ha clasificado como Organización y se ha encontrado sólo en un documento de nuestro corpus.

iii. Ejemplo 3

“Cabe mencionar que G3ict es un ente creado por la Alianza Global para el Desarrollo de las Tecnologías de la Información y Comunicación (GAID, por sus siglas en inglés)”.

En este ejemplo podemos apreciar, además de la aplicación de la regla 3, también la aplicación de la regla 0.

Primero, la regla 3 encuentra que dentro del paréntesis existe la palabra “siglas”, y procede a buscarlas, encontrándolas en “GAID”. Después, las busca hacia atrás en el contexto, usando los procedimientos explicados por la regla 0. Así, usando el criterio de buscar letras mayúsculas en desorden, se logra resolver la entidad, devolviendo la siguiente entrada en la base de datos:

```
{
  "_id" : ObjectId("5b22ede1219ef328ff95a9ec"),
```

```

"Archivos" : {
  "Regla" : "Regla 3",
  "Alias" : "GAID",
  "indiceOcurrecencia" : 18428,
  "Nombre" : "DOF_P_IPT_EXT_241116_36_Acc"
},
"Nombre" : "Alianza Global para el Desarrollo de las Tecnologías de la Información y
Comunicación",
"Clase" : "Otro"
}

```

Observamos que en este caso, la clasificación ha sido “Otro”, lo cual se intuye incorrecto, pues es probable que la entidad sea una Organización.

iv. Ejemplo 4

”En tanto, la G3ict (Iniciativa Global para TIC incluyente)”

En este caso, el contexto también es pequeño por encontrarse al inicio del párrafo, y se tuvo que agregar esta entrada al diccionario, pues no se encontró una forma de resolver el alias planteado por medio de ninguna otra regla.

```

{
  "_id" : ObjectId("5b22ede1219ef328ff95a9eb"),
  "Archivos" : {
    "Regla" : "Regla 4",
    "Alias" : "g3ict",
    "indiceOcurrecencia" : 18097,
    "Nombre" : "DOF_P_IPT_EXT_241116_36_Acc"
  },
  "Nombre" : "Iniciativa Global para TIC incluyente",
  "Clase" : "Otro"
}

```

Observamos también que la entidad ha sido clasificada como “Otro” a pesar de que, probablemente, sea una organización.

Pasemos ahora a analizar los resultados de la implementación para reconocer entidades sin alias.

II. Entidades sin alias

Para estos ejemplos, recordemos que existían tres patrones, o expresiones regulares, para obtener candidatos a alias. Cada ejemplo corresponde con cada expresión regular utilizada.

Además, a diferencia de las entidades con alias, estas entidades no requieren de un contexto, pues la expresión regular se busca en todo el texto del documento. Aún así, se proveerá un poco del párrafo donde se ha encontrado, para explicar cómo la expresión regular ha encontrado la entidad.

i. Ejemplo 1

AMM	Adolfo Merino Medina.
AMX	América Móvil, S.A.B. de C.V.
AUTORIDAD INVESTIGADORA	Autoridad Investigadora del INSTITUTO.

5.7 Tabla extraída de un documento del corpus

Observamos que el contexto donde fue encontrada la entidad de este ejemplo, es una tabla (dentro del documento de Word). Al obtener el texto plano de la misma, se obtiene:

AMM

Adolfo Merino Medina.

AMX

América Móvil, S.A.B. de C.V.

Autoridad Investigadora

Autoridad Investigadora del Instituto.

Después, la expresión regular, que corresponde al primer criterio:

```
\b(((CC.|C.|Mtro.|Mtra.|Sra.|Sr.|Lic.|Dr.|Dra.|Pdte.) )?([A-Z][a-záéíóú]+
?)+(((|a|e|l|los|las|un|una|uno|unas|unos|y|con|de|del|a) )*( [A-Z][a-záéíóú]+ ?)+*((, | ) )?(S.A. de C.V.|S. de
R.L. de C.V.|S.A.B. de C.V.|S.A.)?)
```

Devuelve lo siguiente:

Match result:

```

AMM
Adolfo Merino Medina.
AMX
América Móvil, S.A.B. de C.V.
Autoridad Investigadora
Autoridad Investigadora del Instituto.

```

5.8 Tabla de verificación obtenida con el servicio web de Pythex¹⁸

Primero, "Autoridad Investigadora" será filtrado por tener menos de tres palabras, según el criterio establecido. Lo demás será guardado como entidad. A continuación las tres entradas en las bases de datos para las entidades devueltas:

```

{
  "_id" : ObjectId("5b22ee9d219ef328ff95b884"),
  "Archivos" : [
    {
      "indiceOcurrencia" : [
        3542,
        223044
      ],
      "Alias" : "",
      "Regla" : "Expresión Regular",
      "Nombre" : "VP_P_IFT_230117_8_AccAI"
    }
  ],
  "Nombre" : "América Móvil S.A.B. de C.V.",
  "Clase" : "Organización"
}

{
  "_id" : ObjectId("5b22ee9d219ef328ff95b883"),
  "Archivos" : {
    "Regla" : "Expresión Regular",
    "Alias" : "",
    "indiceOcurrencia" : 3514,
    "Nombre" : "VP_P_IFT_230117_8_AccAI"
  }
}

```

¹⁸ <http://pythex.org>

```

},
"Nombre" : "Adolfo Merino Medina",
"Clase" : "Otro"
}

{
  "_id" : ObjectId("5b22ee9d219ef328ff95b885"),
  "Archivos" : {
    "Regla" : "Expresión Regular",
    "Alias" : "",
    "indiceOcurrencia" : 3598,
    "Nombre" : "VP_P_IFT_230117_8_AccAI"
  },
  "Nombre" : "Autoridad Investigadora del Instituto",
  "Clase" : "Organización"
}

```

La primer entrada fue clasificada correctamente, la segunda entrada no fue detectada como persona, pues no hubo indicios de que fuera tal (como un título personal), y la última entrada es debatible si se trata de una persona o de una organización. Observemos también que América Móvil S.A.B. de C.V fue encontrada dos veces en el documento, tal como fue encontrada por la expresión regular.

ii. Ejemplo 1

En este caso, se utiliza la expresión regular en todo el documento:

```
((o|O)ficio|(e|E)xpediente|(a|A)cuerdo|(O|o)pinión) (. *?número )?([A-Z][0-9],|\,|;|:\|\\)*(/-)([A-Z][0-9];|:\|\\|/)*
```

Obteniendo resultados como

```

{
  "_id" : ObjectId("5b22ee4a219ef328ff95b4ed"),
  "Archivos" : {
    "Regla" : "Expresión Regular Documento",
    "Alias" : "",
    "indiceOcurrencia" : 1397,
    "Nombre" : "DOF_P_IFT_EXT_241116_36_Acc"
  },
  "Nombre" : "Acuerdo P/IFT/120815/306",

```

```

"Clase" : "Documento"
}

{
  "_id" : ObjectId("5b22ee9f219ef328ff95b8fa"),
  "Archivos" : {
    "Regla" : "Expresión Regular Documento",
    "Alias" : "",
    "indiceOcurrecencia" : 2409,
    "Nombre" : "VP_P_IFT_230117_8_AccAI"
  },
  "Nombre" : "expediente administrativo con número E-IFT/UC/DGIPM/PMR/0008/2013",
  "Clase" : "Documento"
}

{
  "_id" : ObjectId("5b22ede5219ef328ff95aabdb"),
  "Archivos" : [
    {
      "indiceOcurrecencia" : [
        6395,
        20479
      ],
      "Alias" : "",
      "Regla" : "Expresión Regular Documento",
      "Nombre" : "P_IFT_140218_91_AccUMCA"
    }
  ],
  "Nombre" : "oficio IFT/226/UCE/DG-COEC/024/2018",
  "Clase" : "Documento"
}

```

Las tres entradas han sido clasificadas correctamente y observamos que la expresión regular detecta los patrones más comunes de ID en estos documentos, así como casos donde el mismo se encuentra alejado de la palabra “Oficio”, “Acuerdo”, “Expediente” u “Opinión”, como en la entrada número 2: “expediente administrativo con número E-IFT/UC/DGIPM/PMR/0008/2013”

iii. Ejemplo 3

Este caso se encontró con la tercera expresión regular:

```
([0-9]*,? )*((A|a)rtículos?(F|f)racci(ó|o)n(es)?|(I|i)ncisos?(P|p)árrafos?)( ?\S){0,100}? (del |de la )(abrogada
)?([A-Z][a-záéíóú]+ ?)+(((l|a|e|l|l|o|s|l|a|s|u|n|u|n|a|u|n|o|u|n|a|s|u|n|o|s|y|c|o|n|d|e|d|e|l|a) )*([A-Z][a-záéíóú]+ ?)+)*
```

Observamos, de la expresión, que esta contiene casi toda la expresión 1, con algunas modificaciones al inicio, pues confía que toda ley, artículo, párrafo, fracción o inciso, siempre esté referido a un lugar de procedencia, el cual se encuentra con la expresión del ejemplo 1. Así, la expresión 1 nos ayudó para saber cuándo detener la búsqueda de entidades. Así, observamos el resultado favorable siguiente:

```
{
  "_id" : ObjectId("5b22edee219ef328ff95acbf"),
  "Archivos" : [
    {
      "indiceOcurriencia" : 48695,
      "Alias" : "",
      "Regla" : "Expresión Regular Ley",
      "Nombre" : "P_IFT_240118_25_AccUPR"
    },
    {
      "Regla" : "Expresión Regular Ley",
      "Alias" : "",
      "indiceOcurriencia" : 204171,
      "Nombre" : "P_IFT_210218_119_AccUPR"
    }
  ],
  "Nombre" : "1 2 6 fracción IV 7 15 fracción LXIII 140 141 143 y 144 de la Ley Federal de
  Telecomunicaciones y Radiodifusión",
  "Clase" : "Ley"
}
```

Observamos primero que se detectan los primeros números, que hacen referencia a artículos, y termina hasta que ha encontrado una cadena que cumple los criterios de la expresión regular 1. Además, vemos que se hace referencia a estos mismos artículos, de la misma ley, en dos documentos distintos.

Una vez que hemos experimentado con las entidades, veamos qué tan útiles son a la hora de querer extraer eventos, en la siguiente sección.

III. Eventos

Busquemos ahora algunos eventos, para cada verbo, con sus respectivos actuantes. Esta implementación, recordemos, no ingresa nada en una base de datos, pues no es su finalidad. Sin embargo, podemos comprobar sus resultados, leyéndolos directamente en la terminal donde esté corriendo.

i. Ejemplo 1

```
A el respecto mediante oficio 2.1.-0222 emitido por la Dirección General de Política de
Telecomunicaciones y de Radiodifusión adscrita a la Secretaría notificó el oficio 1.-55 recibido en
este Instituto el 7 de marzo de 2014 mediante el cual la Secretaría emitió la opinión técnica en
sentido favorable.
Verbo: emitió
Fecha: 7 de marzo de 2014
Quien: Secretaría
Que: opinión técnica
```

5.9 salida en consola de la implementación del programa para generar eventos.

Observamos que se ha detectado, en esa oración obtenida con FreeLing, el verbo en su forma “emitió” y la fecha 7 de marzo del 2014, por lo que se procede a buscar los actuantes, que en este caso son Secretaría, y “opinión técnica”.

La expresión regular:

```
\b"+ fecha +" .*?" + "mediante .*? ?" + "(" + cadenaEntidades + ")" + " .*?" + verbo
```

De expresión, extraída directamente del código fuente, observamos que primero debe encontrarse la fecha contenida en la variable **fecha**, después la palabra “mediante”, luego alguna de las entidades detectadas en ese documento, las cuales están contenidas en “**cadenaEntidades**” y finalmente el verbo “emitió” contenido en la variable **verbo**.

Con esto hemos encontrado la respuesta a ¿Quién?

El “qué” fue encontrado con la expresión:

```
\b(opinión .*? )
```

la cual capturó la palabra opinión seguida de una sola palabra. En este caso, no se hizo uso de las entidades detectadas.

ii. Ejemplo 2

El 14 de marzo de 2012 la Secretaría de Comunicaciones y Transportes la “ Secretaría ” otorgó a favor de Digicable S.A. de C.V. un título de concesión para instalar operar y explotar una red pública de telecomunicaciones para prestar el servicio de televisión restringida en Poza Rica de Hidalgo Municipio de Poza Rica de Hidalgo en el Estado de Veracruz con una vigencia de 30 treinta años contados a partir de su otorgamiento la “ Concesión ”.

Verbo: otorgó

Fecha: 14 de marzo de 2012

Quien: Secretaría de Comunicaciones y Transportes

A Quien: Digicable S.A. de C.V.

Que: concesión para instalar operar y explotar una red pública de telecomunicaciones para prestar el servicio de televisión restringida en Poza Rica de Hidalgo Municipio de Poza Rica de Hidalgo en el Estado de Veracruz con una vigencia de 30 treinta años contados a partir de su otorgamiento

5.10 salida en consola de la implementación del programa para generar eventos.

“Quién” se ha detectado con la expresión:

`\b" + fecha + " .*? ?" + "(" + cadenaEntidades + ")" + " .*?" + verbo`

La cual primero detecta la fecha contenida en **fecha**, luego alguna entidad detectada previamente contenida en **cadenaEntidades** y luego el verbo “otorgó” contenido en **verbo**.

el “A Quien” fue encontrado por:

`\b" + verbo + " a .*?" + "(" + cadenaEntidades + ")`

El cual primero detecta el verbo, seguido de la letra “a”, seguido de una entidad detectada previamente.

En ambos casos, la entidad nombrada es la respuesta a la pregunta buscada.

Finalmente el “Qué” se encuentra con la expresión:

`\b" + verbo + " .*?" + "(" + cadenaEntidades + ")`

La cual busca alguna entidad nombrada después del verbo, y habiendo eliminado antes, a la entidad “A Quien”, la entidad siguiente fue la encontrada, y la correcta.

iii. Ejemplo 3

El 13 de enero de 2014 el representante legal de Impacto Telecomunicaciones S.A. de C.V. presentó ante el Instituto solicitud de prórroga de vigencia de la Concesión la “ Solicitud de Prórroga ”.

Verbo: presentó

Fecha: 13 de enero de 2014

Quien: Impacto Telecomunicaciones S.A. de C.V.

A Quien: Instituto

Que: solicitud de prórroga de vigencia de la Concesión

5.11 salida en consola de la implementación del programa para generar eventos.

En este caso, se encontró “Quién” con la expresión:

```
\brepresentante legal de .*? ?" + "(" + cadenaEntidades + " )
```

la cual busca la cadena “representante legal de” antes de alguna entidad reconocida previamente.

Después, el “A Quién” se encontró con:

```
\bante .*? ?" + "(" + cadenaEntidades + " )
```

la cual busca la cadena “ante” antes de una entidad, que en este caso fue “Instituto”. Notemos que “Instituto” seguramente es un Alias, esto es porque la variable **cadenaEntidades** contiene las entidades de ese documento, así como los alias utilizados.

Finalmente el qué, se encontró con la expresión:

```
\b" + verbo + " .*?" + "(" + cadenaEntidades + " )
```

La cual sólo busca una entidad después del verbo, que en este caso es la correcta, pues previamente eliminamos a “Instituto” por haber sido la respuesta a “A Quién”.

iv. Ejemplo 4

El 3 de octubre de 2016 el Instituto publicó en el DOF el “ ACUERDO MEDIANTE EL CUAL EL PLENO DE EL INSTITUTO FEDERAL DE TELECOMUNICACIONES ESTABLECE LAS CONDICIONES TÉCNICAS MÍNIMAS ENTRE CONCESIONARIOS QUE OPEREN REDES PÚBLICAS DE TELECOMUNICACIONES Y DETERMINA LAS TARIFAS DE INTERCONEXIÓN RESULTADO DE LA METODOLOGÍA PARA EL CÁLCULO DE COSTOS DE INTERCONEXIÓN QUE ESTARÁN VIGENTES DE EL 1 DE ENERO A EL 31 DE DICIEMBRE DE 2017 ” aprobado mediante Acuerdo P/IFT/200916/503 en lo sucesivo el “ Acuerdo de CTM y Tarifas 2017 ”.

Verbo: publicó

Fecha: 3 de octubre de 2016

Quien: Instituto

Donde: DOF

Que: Acuerdo P/IFT/200916/503

5.12 salida en consola de la implementación del programa para generar eventos.

Una vez más, el “Quién” fue encontrado por la expresión:

```
\b" + fecha + " .*? ?" + "(" + cadenaEntidades + ")" + " .*?" + verbo
```

pues, resulta ser que la entidad contenida entre la fecha y el verbo, es la respuesta a dicha pregunta.

Después, el “Donde” se ha encontrado con la expresión:

```
\ben .*? ?" + "(" + cadenaEntidades + " )
```

La cual sólo busca la cadena “en” antes de alguna entidad. Esto es efectivo pues sabemos a priori que el verbo “publicar” es el único que puede responder esta pregunta, y según lo analizado en los documentos, no fue necesario elaborar una expresión más compleja para evitar falsos positivos.

Finalmente, en el “Que” observamos que se ha detectado el alias del acuerdo. Sin embargo, sabemos que ese alias no fue detectado por nuestro software, pues el mismo no viene contenido en paréntesis. Sin embargo, fue detectado como una entidad por sí misma, probablemente por el patrón 2 de entidades sin alias, lo que permitió responder esta pregunta de manera correcta.

v. Ejemplo 5

El 7 de noviembre de 2017 el apoderado legal de Mega Cable presentó ante el Instituto escritos mediante los cuales solicitó su intervención para resolver los términos condiciones y tarifas que no pudo convenir con Pegaso PCS para la interconexión de sus respectivas redes públicas de telecomunicaciones aplicables a el periodo 2018 en lo sucesivo las “ Solicitudes de Resolución ”.
 Verbo: presentó
 Fecha: 7 de noviembre de 2017
 Quien: Mega Cable
 A Quien: Instituto
 Que: Solicitudes de Resolución |

5.13 salida en consola de la implementación del programa para generar eventos.

Finalmente, en este caso, el “Quién” fue encontrado por la expresión:

```
\b" + fecha + " .*? ?" + "(" + cadenaEntidades + ")" + " .*?" + verbo
```

Siendo la respuesta, probablemente, un alias.

Después, el “A Quien” se encontró con:

```
\bante .*? ?" + "(" + cadenaEntidades + " )
```

que encontró la palabra “ante” antes del alias “Instituto” al que, consecuentemente, caracterizó como “A Quien”.

Finalmente el “Qué” fue detectado pasando por alto la entidad en sí, y tomando al alias. Esto es porque nuestro reconocedor de entidades no vio que “Solicitudes de Resolución” era un alias de “escritos mediante los cuales solicitó su intervención para resolver los términos condiciones y tarifas que no pudo convenir con Pegaso PCS para la interconexión de sus respectivas redes públicas de telecomunicaciones aplicables a el periodo 2018” debido a que el alias no se encuentra entre paréntesis. Aún así, al ser detectado el alias como una entidad, se pudo dar una respuesta favorable a la pregunta de “Que”.

En estos ejemplos que hemos analizado, vimos el uso de nuestro reconocedor de entidades nombradas, así como la necesidad de tener una base de datos accesible para que el detector de

eventos funcione de manera óptima. También pudimos observar algunos casos donde deficiencias del reconocedor, se han traducido en deficiencias para el reconocedor de eventos.

Pasemos ahora, en la sección de *Resultados*, a cuantificar lo que se ha realizado.

6. Resultados

Después de vistos los experimentos, pasemos ahora a cuantificar los resultados. Hablaremos aquí de las medidas de desempeño mostradas por el reconocedor de entidades, y daremos algunos ejemplos de eventos encontrados, usando los mismos documentos que hemos usado para evaluar. En la presentación de las tablas, a nuestro software le llamaremos “Software NER Especializado”.

Primero, hemos obtenido del sitio web del IFT cinco documentos. Cuatro de ellos pertenecen a nuestro dominio, y son de los años 2015, 2016, 2017 y 2018 respectivamente; el quinto documento del IFT es ajeno a nuestro dominio específico; ninguno de ellos se encuentra en nuestro corpus original. De estos cinco documentos hemos etiquetado, yo y una compañera licenciada en derecho, manualmente sus entidades y clasificaciones.

Para cada documento obtendremos una tabla, disponibles como apéndices, las cuales tendrán tres columnas:

- 1) Entidades nombradas encontradas manualmente
- 2) Entidades nombradas según nuestro programa
- 3) Entidades nombradas según FreeLing 4.0

Después, para cada documento obtendremos las medidas de desempeño de precisión y exhaustividad, tanto de FreeLing 4.0, como de nuestro programa, para evaluar qué tan bien se ha hecho el reconocimiento de entidades nombradas, y establecer las bases para confirmar la hipótesis en la sección *Conclusión*.

Cabe mencionar que para el cálculo de las medidas de desempeño, se ha considerado un verdadero positivo, cuando, palabra por palabra, las entidades coinciden, ignorando signos de puntuación. Este mismo criterio de decisión aplica para falsos negativos y falsos positivos.

Además, notaremos que en la columna de los resultados de FreeLing, la palabra “del” se separa por “de el”. Lo hemos considerado, y no se ha tomado como un error, pues tiene que ver con la configuración de FreeLing.

Empezaremos mostrando una tabla con las medidas de desempeño promediadas de los cuatro documentos que pertenecen a nuestro dominio:

	FreeLing (%)	Software NER Especializado (%)
Precisión	20.04	80.67
Exhaustividad	22.39	75.62

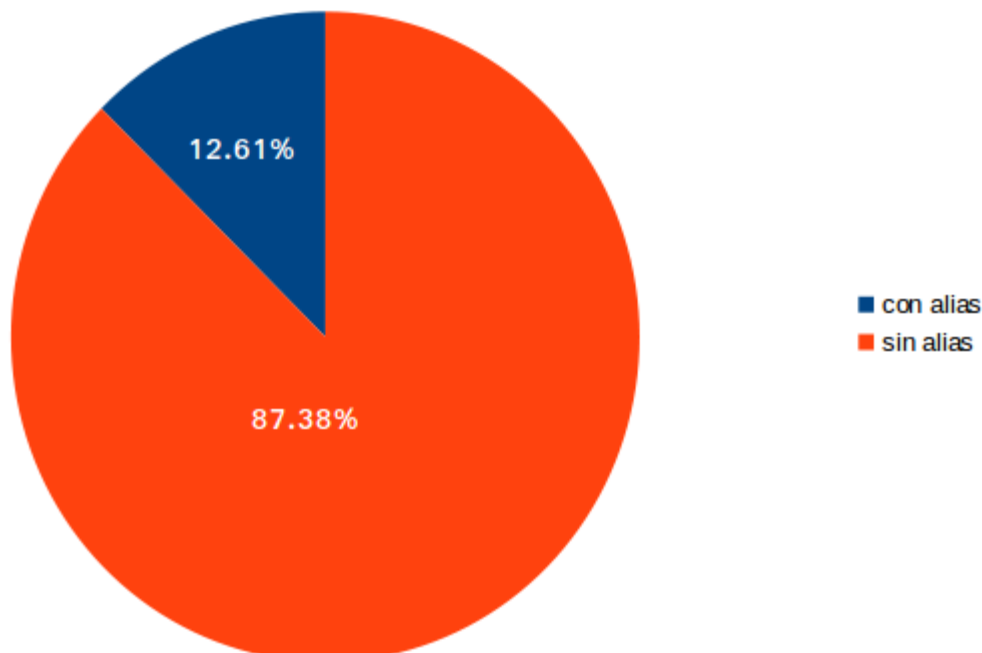
Medida F1	21.14	78.11
-----------	-------	-------

6.1 Medidas de desempeño promediadas de cuatro documentos pertenecientes al dominio, sobre reconocimiento de entidades nombradas

Observamos que se ha podido incrementar significativamente el desempeño, tanto de precisión, como de exhaustividad, respecto al reconocedor de entidades nombradas de FreeLing 4.0.

En el caso de FreeLing, hay mayor exhaustividad que precisión, mientras que en nuestro Software, existe más precisión que exhaustividad. Esto se debe a que los sistemas de reglas tienden a tener menos falsos positivos; Además, conforme más reglas hay, aplican a menos casos, pues siempre se empieza de lo general a lo particular, por lo que cada vez es más complicado aumentar la exhaustividad.

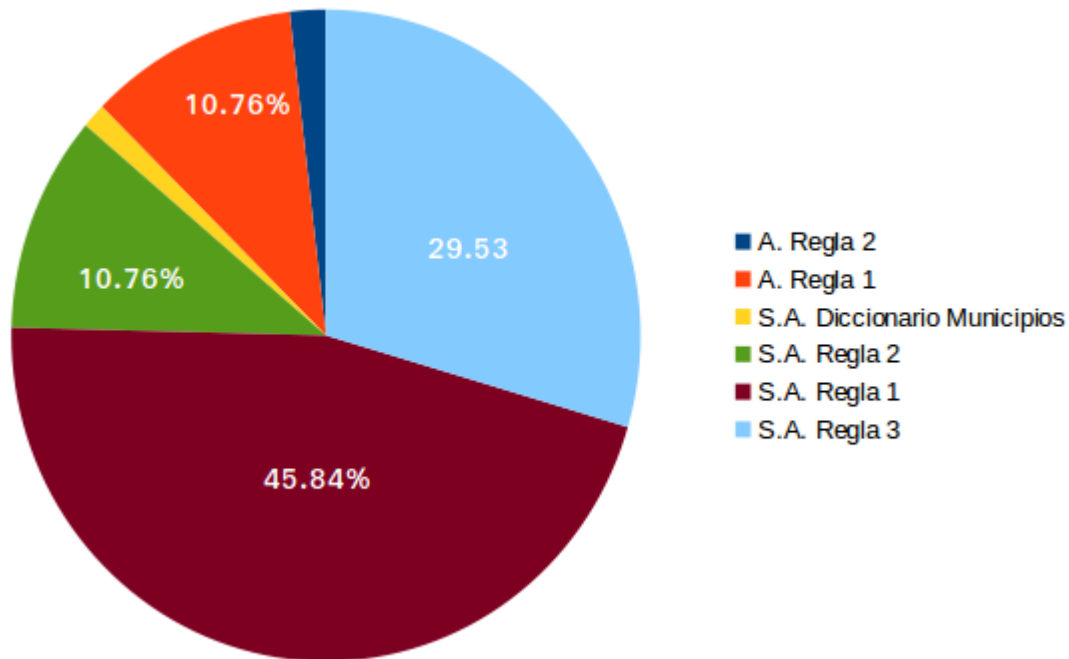
Veamos ahora qué tantas entidades correctas ha devuelto cada regla, en los 4 documentos de nuestro dominio:



6.2 Gráfica que muestra la proporción de entidades nombradas encontradas por cada acercamiento

Primero observamos que los criterios para obtener entidades sin alias, obtienen mucho mayores resultados positivos que los criterios para obtener entidades con alias, representando el 87% del total de entidades, en este corpus de evaluación de cuatro documentos. Esto puede ser debido a que existen muchas más entidades nombradas sin alias que con alias, y no una diferencia entre las precisiones de un método u otro.

Veamos ahora, qué reglas han sido las más efectivas, y cuáles las menos efectivas:



6.3 Gráfica que muestra la proporción de entidades nombradas encontradas por cada regla aplicada

En esta gráfica, la “A.” significa “Alias” y “S.A.” significa “Sin Alias”. Observamos que el criterio 1 de las reglas sin alias, así como el criterio 3, devuelven el mayor número de entidades. En cambio, las reglas para detectar entidades con alias, a pesar de existir cuatro, sólo observamos la presencia de dos, donde la más importante es la regla 1, con 10.76%.

Esto se debe a que la regla 3 con alias resultó resolver un caso demasiado específico. Así mismo, la regla 4 (diccionario) resuelve casos específicos que deben ser encontrados de forma manual, por lo que su presencia en la gráfica podría aumentar si encontráramos en un corpus diferente, o más amplio, entidades que agregar al diccionario.

La regla 1 sin alias ha traído muchos resultados ya que en nuestros documentos existen muchas personas y organizaciones. La regla 2 encuentra documentos, los cuales fueron menor número que las leyes, encontradas por la regla 3.

Para entender mejor por qué hemos obtenido estos números, analicemos las medidas de desempeño de cada documento, de forma individual.

a) Documento del 2015

Empecemos con el documento del año 2015, del cual sólo teníamos 10 muestras en nuestro corpus. Comenzaremos mostrando y explicando una tabla con las medidas de desempeño del

reconocedor de entidades nombradas, y después veremos medidas de desempeño para cada clasificación del clasificador, tanto de FreeLing, como el desarrollado en este trabajo.

	FreeLing 4.0 (%)	Software NERC Especializado (%)
Precisión	29.56	88.88
Exhaustividad	36.17	85.10
Medida F1	32.53	86.94

6.4 Medidas de desempeño para reconocimiento de entidades nombradas documento 2015

En este caso, los valores obtenidos tanto de FreeLing, como de nuestro Software, han sido mayores al promedio. Tratando de dilucidar la razón, he notado que este documento es más pequeño que el resto, con 4058 palabras, y carece de tablas, o estructuras en el texto que pudieran significar ruido a la hora de obtener el texto plano debido a saltos de línea incoherentes.

También podría ser producto de una tendencia propia a encontrar más patrones de documentos del año 2015, que de otros años, al querer compensar el hecho de que los documentos de este año son los de menor cantidad en el corpus. Sin embargo, esto no explica el hecho de que FreeLing haya tenido también un mejor desempeño.

Observando la tabla *Documento 2015* provista en *Apéndices*, notamos primero que FreeLing ha obtenido más entidades, y nuestro software menos entidades, que el etiquetado manual. Esto hace a FreeLing más susceptible a errores de precisión, y mejor exhaustividad, mientras que nuestro programa, al tener menor número de entidades, se garantiza que nunca tendremos 100% de exhaustividad.

Ahora, veamos qué tan bien clasificó FreeLing y nuestro Software las entidades nombradas que reconocieron correctamente. Es decir, sólo tomaremos los verdaderos positivos de los reconocedores de entidades nombradas, ya que queremos evaluar la capacidad de clasificar entidades, independientemente de la capacidad para detectar entidades. Para esta labor, hemos calculado la precisión y exhaustividad para cada clasificación posible.

Recordemos que FreeLing sólo tiene tres clasificaciones, por lo que no se han podido calcular valores para las clasificaciones extra que contiene nuestro Software.

	FreeLing 4.0 (%)			Software NERC Especializado (%)		
	Precisión	Exhaustividad	Medida F1	Precisión	Exhaustividad	Medida F1
Organización	43.75	77.77	55.99	100	84.61	91.66
Persona	77.77	87.5	82.34	100	50.0	66.66
Lugar	75.00	100	85.71	100	100	100
Documento	-	-	-	92.30	75	82.75
Ley	-	-	-	100	86.11	92.53

**6.5 Medidas de desempeño para clasificación de entidades nombradas
documento 2015**

Los valores de 100% significan que todas las clasificaciones que se han hecho, en esa categoría, han sido correctas, sólo tomando en cuenta las entidades nombradas que ya sabemos fueron correctas en primer lugar. Es decir, las entidades que resultaron ser falsos positivos, y que también tienen una clasificación, no fueron tomadas en cuenta, pues se trata de un error de detección, no de clasificación.

Para explicar este comportamiento, basta decir que la clasificación se ha hecho con diccionarios elaborados a partir del análisis de los documentos, como se ha explicado ya en la metodología. Además, el etiquetado manual fue realizado, en su mayoría, por mi, y por una compañera licenciada en derecho, por lo que el mismo criterio que se usó para hacer los diccionarios, se ha usado para hacer el etiquetado, lo cual queda evidenciado en la precisión de los resultados.

Aún considerando esto, vemos que FreeLing clasifica más personas que nuestro Software, el cual tiene problemas para clasificar personas cuando estas no son indicadas por medio de un título como "Sr." o "Lic.", etc.

Además, observamos que, de las entidades encontradas correctamente, los lugares los hemos clasificado todos acertadamente, aunque FreeLing ha clasificado entidades válidas como lugares, sin serlo.

Analizemos ahora el documento del año 2016.

b) Documento del 2016

De este documento tenemos 44 muestras en nuestro corpus. Veamos, al igual que el documento anterior, la tabla sobre reconocimiento de entidades nombradas con sus medidas de desempeño, y después evaluemos el clasificador.

	FreeLing (%)	Software NER Especializado (%)
Precisión	27.65	84.15
Exhaustividad	36.11	79.43
Medida F1	31.31	81.72

**6.6 Medidas de desempeño para reconocimiento de entidades nombradas
documento 2016**

Al igual que el documento anterior, en este caso también tenemos un mejor desempeño que el promedio. Sin embargo, el desempeño es peor respecto al año 2015, a pesar de haber tenido más documentos. Esto pone en evidencia un problema con los sistemas de reglas: son susceptibles a errores humanos, pues las reglas surgen del análisis manual.

En un sistema estadístico, habría una correspondencia clara entre el número de documentos, y el desempeño del programa, pues mientras más documentos, hay más instancias del fenómeno que queremos analizar. En cambio, en los sistemas de reglas, la percepción y criterio del individuo, independientemente del número de documentos, es la que decide qué regla aplicar y qué regla omitir, por lo que la correspondencia entre número de documentos y desempeño, puede no estar presente.

Veamos ahora las medidas de desempeño de las clasificaciones realizadas.

	FreeLing 4.0 (%)			Software NERC Especializado (%)		
	Precisión	Exhaustividad	Medida F1	Precisión	Exhaustividad	Medida F1
Organización	50	80	61.53	100	88.23	93.74
Persona	70	87.5	77.77	0	0	0
Lugar	100	60	75	100	50	66.66
Documento	-	-	-	95.83	69.69	80.69
Ley	-	-	-	100	88.88	94.11

**6.7 Medidas de desempeño para clasificación de entidades nombradas
documento 2016**

En este caso, ambos hemos clasificado correctamente los lugares (Precisión), y hemos tenido mejores medidas para "Organización". Sin embargo, hemos obtenido 0 para "Persona" en precisión y exhaustividad.

Mirando la tabla de resultados para el documento 2016, disponible en *Apéndices*, observamos que sólo hay una entidad clasificada como persona, la cual es “Presidente del Instituto Federal de Telecomunicaciones”. Sin embargo, la entidad correcta según el etiquetado manual, es “Comisionado Presidente del Instituto Federal de Telecomunicaciones”, por lo que no hay ninguna entidad correcta con la clasificación de persona.

A pesar de esto, sí hemos detectado nombres de personas, por ejemplo, “Mario Alberto Radilla Hernández”, pero se ha clasificado como “Otro” en lugar de Persona, por lo que tampoco ayuda a nuestra métrica, a pesar de ser una entidad reconocida correctamente.

Veamos ahora el documento del año 2017

c) Documento del 2017

De estos documentos también tenemos 44 muestras en el corpus. Empecemos con la tabla con las medidas de desempeño del reconocedor de entidades nombradas.

	FreeLing (%)	Software NER Especializado (%)
Precisión	18.91	81.14
Exhaustividad	34.75	74.43
Medida F1	24.49	77.64

6.8 Medidas de desempeño para reconocimiento de entidades nombradas documento 2017

Observamos que la precisión se mantiene arriba del promedio, mientras que la exhaustividad ha bajado del mismo. De igual manera, FreeLing se ha comportado aún peor que antes. Una posible explicación radica en el tamaño de los archivos. El documento del 2015 es el más pequeño, seguido por el del 2016, luego el del 2017, y finalmente el del 2018, lo cual ha coincidido con las bajas en las medidas de desempeño de ambos reconocedores de entidades. Al parecer, mientras más grande el archivo, peores medidas de desempeño.

La injerencia del tamaño también se observa en la tabla correspondiente provista en *Apéndices*, donde la cantidad de entidades encontradas manualmente supera a las encontradas por cualquiera de los dos reconocedores, afectando la exhaustividad de ambos.

Veamos ahora las medidas de los clasificadores.

	FreeLing 4.0 (%)	Software NERC Especializado (%)

	Precisión	Exhaustividad	Medida F1	Precisión	Exhaustividad	Medida F1
Organización	44.44	92.30	59.99	95.65	95.65	95.65
Persona	60	100	75	0	0	0
Lugar	0	0	0	90	60	70
Documento	-	-	-	100	94.87	97.36
Ley	-	-	-	100	94.87	97.36

6.9 Medidas de desempeño para clasificación de entidades nombradas

documento 2017

De nuevo no hemos podido detectar personas, y hemos detectado todos los lugares. Completamente opuesto ha sido el comportamiento de FreeLing, que ha clasificado correctamente a todas las personas, y a ningún lugar.

Recordemos que la mayor cantidad de lugares que se detectan en nuestro programa, son porque poseemos un diccionario de municipios y estados. Así mismo, FreeLing posee un diccionario de nombres, por lo que podríamos concluir que los diccionarios son buenos métodos de clasificación.

Analizemos, el último documento perteneciente al dominio, del año 2018.

d) Documento del 2018

De estos documentos tenemos 36 muestras. Veamos sus medidas de desempeño.

	FreeLing (%)	Software NER Especializado (%)
Precisión	24.24	68.53
Exhaustividad	40.40	63.54
Medida F1	30.3	65.94

6.10 Medidas de desempeño para reconocimiento de entidades nombradas

documento 2018

Siguiendo la tendencia mencionada anteriormente, respecto al tamaño de los documentos, el desempeño ha caído considerablemente, colocándose muy por debajo del promedio, en ambos reconocedores. Sin embargo, en este caso, la cantidad de entidades etiquetadas no sobrepasa por mucho la cantidad de entidades encontradas por nuestro software, por lo que quizá las bajas en desempeño puedan deberse más a un error humano de no haber considerado tanto a los documentos más nuevos para la generación de reglas.

Veamos las medidas de las clasificaciones realizadas.

	FreeLing 4.0 (%)			Software NERC Especializado (%)		
	Precisión	Exhaustividad	Medida F1	Precisión	Exhaustividad	Medida F1
Organización	43.75	63.63	51.84	100	71.42	83.32
Persona	100	75	85.71	0	0	0
Lugar	50	100	66.66	0	0	0
Documento	-	-	-	93.75	65.21	76.91
Ley	-	-	-	95	79.16	86.35

**6.11 Medidas de desempeño para clasificación de entidades nombradas
documento 2018**

Vemos que no hemos clasificado correctamente, ni personas ni lugares. En cambio, FreeLing continúa su buen desempeño clasificando personas, gracias a su diccionario de nombres. Se observa que las clasificaciones más estables en cuanto a resultados en los cuatro documentos, son las de Documento y Ley.

Veamos por último cómo se comportan los programas con un documento ajeno a nuestro dominio.

e) Documento ajeno al dominio

Este documento es el que ha tenido peor desempeño, lo cual es coherente, pues es un dominio que no se ha utilizado para generar reglas. Sin embargo, sigue siendo mejor que el reconocedor de FreeLing 4.0

	FreeLing (%)	Software NER Especializado (%)
Precisión	30.18	67.27
Exhaustividad	50	58.73
Medida F1	37.64	62.71

**6.12 Medidas de desempeño para reconocimiento de entidades nombradas
documento ajeno al dominio (acta)**

Evidentemente, este documento no lo hemos utilizado para generar el promedio mostrado al inicio de esta sección. Además, con estos números se confirma que los sistemas de reglas suelen comportarse de forma indeseable en dominios ajenos.

Veamos ahora las medidas de clasificación.

	FreeLing 4.0 (%)			Software NERC Especializado (%)		
	Precisión	Exhaustividad	Medida F1	Precisión	Exhaustividad	Medida F1
Organización	58.33	100	73.68	100	100	100
Persona	77.77	82.35	79.99	0	0	0
Lugar	100	100	100	100	100	100
Documento	-	-	-	100	50	66.66
Ley	-	-	-	100	80	88.88

6.13 Medidas de desempeño para clasificación de entidades nombradas

documento ajeno al dominio

Como la clasificación se ha hecho con diccionarios, esperaríamos que las medidas de desempeño se comporten de forma similar que los documentos pertenecientes al dominio, y efectivamente es lo que observamos. Seguimos teniendo un mal desempeño clasificando personas, y un buen desempeño en lo demás.

Nuestra clasificación de Leyes ha caído un poco en este documento, respecto a las demás.

Con estos resultados, pasemos ahora a la conclusión, y al análisis sobre el trabajo a futuro.

7. Conclusiones

Con base en los resultados observados, se confirmó la hipótesis de que es posible construir un reconocedor de entidades nombradas utilizando reconocimiento de patrones, a través de un sistema de reglas, que obtenga mejores medidas de desempeño que un etiquetador de propósito general como FreeLing 4.0.

Además, la implementación del sistema creado, implicó la creación de una base de datos que puede ser útil para otras tareas, como la detección de eventos, y cuya existencia no provee FreeLing.

Conociendo los resultados, y vistos los experimentos, me parece que lo realizado en esta tesis puede mejorarse de la siguiente manera:

- 1.- Creación de más reglas y filtros, añadiendo más documentos al corpus: A pesar de que cada vez es más difícil optimizar los resultados, creo que nuevos documentos del dominio podrían evidenciar más reglas que mejoren los resultados.
- 2.- Implementación de una base de datos relacional: Actualmente, las bases de datos no relacionales no son muy usadas en ambientes de producción grandes, particularmente las gubernamentales, las cuales, intuyo, usan SQL, por lo que sería muy útil implementar una base de datos en este lenguaje, con el fin de ofrecer portabilidad a la información provista.
- 3.- Optimización del código: Al ser un corpus pequeño, y no siendo un parámetro el tiempo de ejecución, es probable que haya dentro del código, rutinas que puedan ser abstraídas para mejorar la eficiencia del mismo.
- 4.- Implementación de un sistema estadístico: Podría también invertirse tiempo en etiquetar el corpus, para generar un sistema estadístico de detección, y comparar los resultados. Esto podría ser un primer paso para tener un sistema híbrido todavía mejor que los dos anteriores.

Bibliografía

Aguilar, C. A., Sierra Martínez, G. E. (2017). Tratamiento de información textual y generación de taxonomías. Serie Docencia (ocre). Instituto de Ingeniería de la Universidad Nacional Autónoma de México.

Borsje, J., Hogenboom, F., & Frasincar, F. (2010). Semi-automatic financial events discovery based on lexico-semantic patterns. *International Journal of Web Engineering and Technology*, 6(2), 115-140.

Domínguez, C., & Agelvis, V. (2003). Lingüística: una introducción generalísima. *Universidad de Los Andes. Escuela de Letras. Departamento de Lingüística*.

Dozier, C., Kondadadi, R., Light, M., Vachher, A., Veeramachaneni, S., & Wudali, R. (2010). Named entity recognition and resolution in legal text. In *Semantic Processing of Legal Texts* (pp. 27-43). Springer, Berlin, Heidelberg.

Fukunaga, K. (2013). Introduction to statistical pattern recognition. Academic press.

Galicia-Haro, S. N., Gelbukh, A., & Bolshakov, I. A. (2004, April). Recognition of named entities in Spanish texts. In *Mexican International Conference on Artificial Intelligence* (pp. 420-429). Springer, Berlin, Heidelberg.

Han, J., Haihong, E., Le, G., & Du, J. (2011, October). Survey on NoSQL database. In *Pervasive computing and applications (ICPCA), 2011 6th international conference on* (pp. 363-366). IEEE.

Hogenboom, F., Frasincar, F., Kaymak, U., & De Jong, F. (2011, October). An overview of event extraction from text. In *Workshop on Detection, Representation, and Exploitation of Events in the Semantic Web (DeRiVE 2011) at Tenth International Semantic Web Conference (ISWC 2011)* (Vol. 779, pp. 48-57).

Hopcroft, J. E., Motwani, R., & Ullman, J. D. (2008). *Teoría de autómatas, lenguajes y computación*. Addison Wesley.

Jurado Málaga, E. (2008). *Teoría de autómatas y lenguajes formales*. Universidad de Extremadura. Servicio de Publicaciones.

Kim, J. H., Kang, I. H., & Choi, K. S. (2002, August). Unsupervised named entity classification models and their ensembles. In *Proceedings of the 19th international conference on Computational linguistics-Volume 1* (pp. 1-7). Association for Computational Linguistics.

- Lagos, N., Segond, F., Castellani, S., & O'Neill, J. (2010, October). Event extraction for legal case building and reasoning. In *International Conference on Intelligent Information Processing* (pp. 92-101). Springer, Berlin, Heidelberg.
- Liddy, E. D. (2001). *Natural language processing*.
- Madnani, N. (2007). Getting started on natural language processing with Python. *Crossroads*, 13(4), 5-5.
- Makhoul, J., Kubala, F., Schwartz, R., & Weischedel, R. (1999, February). Performance measures for information extraction. In *Proceedings of DARPA broadcast news workshop* (pp. 249-252).
- Mansouri, A., Affendey, L. S., & Mamat, A. (2008). Named entity recognition approaches. *International Journal of Computer Science and Network Security*, 8(2), 339-344.
- Mattila, H. E. (2016). *Comparative legal linguistics: language of Law, Latin and modern lingua francas*. Routledge.
- Nadeau, D., & Sekine, S. (2007). A survey of named entity recognition and classification. *Lingvisticae Investigationes*, 30(1), 3-26.
- Nishihara, Y., Sato, K., & Sunayama, W. (2009, July). Event extraction and visualization for obtaining personal experiences from blogs. In *Symposium on Human Interface* (pp. 315-324). Springer, Berlin, Heidelberg.
- Oudah, M., & Shaalan, K. (2017). NERA 2.0: Improving coverage and performance of rule-based named entity recognition for Arabic. *Natural Language Engineering*, 23(3), 441-472.
- Polanco, D. (2000). Evaluación y mejora de un sistema automático de análisis sintagmático. *PFC ETSI Telecomunicación UPM, dirigido por JM Montero*.
- Sierra, G. (2017). *Introducción a los corpus lingüísticos*. Instituto de Ingeniería, Universidad Nacional Autónoma de México, México, Ciudad de México.
- Sierra, G., Bel-Enguix, G., López-Velarde, G., Saucedo, R., Rivera, L. (2018). *Event Extraction from Legal Documents*. In *1st Workshop on Language Resources and Technologies for the Legal Knowledge Graph PROCEEDINGS at Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (pp. 36-40)
- Venturi, G. (2010). Legal language and legal knowledge management applications. In *Semantic Processing of Legal Texts* (pp. 3-26). Springer, Berlin, Heidelberg.

Apéndices

Tablas de resultados de entidades nombradas

Documento 2015

Etiquetado Manual Entidad/clasificación	Software NERC Entidad/clasificación	FreeLing 4.0 Entidad/clasificación
artículos 6o. y 7o. de la Constitución / Ley	Secretaría de Comunicaciones y Transportes / Organización	C. Jorge Adalberto Esperón Heredia / Organización
1, 6 fracciones I y XVIII, 32 y 33 fracción II, 41, 42 fracciones I, II y XV y 50 fracción XII del Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Ley	1 6 fracciones I y XVIII 32 y 33 fracción II 41 42 fracciones I II y XV y 50 fracción XII del Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Ley	Acuerdo / Lugar
35 fracción I, 36, 38, 39 y 57 fracción I de la Ley Federal de Procedimiento Administrativo / Ley	35 fracción I 36 38 39 y 57 fracción I de la Ley Federal de Procedimiento Administrativo / Ley	Acuerdo de el Pleno / Lugar
6 fracción IV, 15 fracciones IV y LVII, 16, 17 fracción I, 66, 67 fracción I, 68, 72, 113 y 114 de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	6 fracción IV 15 fracciones IV y LVII 16 17 fracción I 66 67 fracción I 68 72 113 y 114 de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Acuerdo de el Pleno de el Instituto Federal de Telecomunicaciones / Organización
94 fracción III de la Ley Federal de Derechos / Ley	94 fracción III de la Ley Federal de Derechos / Ley	Adolfo Cuevas Teja / Persona
Acuerdo del Pleno del Instituto Federal de Telecomunicaciones número P/IFT/EXT/131114/228 / Documento	Acuerdo del Pleno / Documento	Adriana Sofía Labardini Inzunza / Persona
Acuerdo del Pleno número P/IFT/EXT/131114/228 / Documento	Acuerdo del Pleno del Instituto Federal de Telecomunicaciones / Documento	Análisis de las Solicitudes de Prórroga / Lugar
Acuerdo P/IFT/161215/600 / Documento	Acuerdo del Pleno del Instituto Federal de Telecomunicaciones número P/IFT/EXT/131114/228 / Documento	Antecedentes / Organización

Adolfo Cuevas Teja / Persona	Acuerdo del Pleno número P/IFT/EXT/131114/228 / Documento	ANTECEDENTES Otorgamiento de la Concesión / Otro
Adriana Sofía Labardini Inzunza / Persona	Acuerdo P/IFT/161215/600 / Documento	Apartado B / Persona
Análisis de las Solicitudes de Prórroga / Documento	Adolfo Cuevas Teja / Otro	Asimismo / Persona
artículo 113 de la Ley / Ley	Adriana Sofía Labardini Inzunza / Otro	C. Jorge Adalberto Esperón Heredia / Otro
artículo 14 fracción X del Estatuto Orgánico / Ley	Análisis de las Solicitudes de Prórroga / Documento	Comisionado Presidente de el Instituto Federal de Telecomunicaciones / Organización
artículo 27 de la abrogada Ley Federal de Telecomunicaciones / Ley	artículo 113 de la Ley / Ley	Comisionados / Organización
artículo 28 párrafo décimo séptimo de la Constitución Política de los Estados Unidos Mexicanos / Ley	artículo 14 fracción X del Estatuto Orgánico / Ley	Competencia / Otro
artículo 28 párrafos décimo quinto, décimo sexto y décimo séptimo de la Constitución / Ley	artículo 27 de la abrogada Ley Federal de Telecomunicaciones / Ley	Competencia Económica / Organización
artículo 3 fracción LVIII de la Ley / Ley	artículo 28 párrafo décimo séptimo de la Constitución Política de los Estados Unidos Mexicanos / Ley	Concesión / Organización
artículo 33 fracción II del Estatuto Orgánico / Ley	artículo 28 párrafos décimo quinto décimo sexto y décimo séptimo de la Constitución / Ley	Concesión / Otro
artículo 6 fracciones I y XVIII del Estatuto Orgánico / Ley	artículo 3 fracción LVIII de la Ley / Ley	Concesiones / Organización
artículo 66 / Ley	artículo 33 fracción II del Estatuto Orgánico / Ley	Concesiones de la Unidad de Competencia Económica / Organización
artículo 6o. Apartado B fracción II de la Constitución / Ley	artículo 6 fracciones I y XVIII del Estatuto Orgánico / Ley	CONSIDERANDO Primero / Persona

artículo 94 fracción III de la Ley Federal de Derechos / Ley	artículo 6o. Apartado B fracción II de la Constitución / Ley	Constitución / Otro
artículo Séptimo Transitorio del Decreto de Ley / Ley	artículo 94 fracción III de la Ley Federal de Derechos / Ley	Constitución Política de los Estados Unidos Mexicanos / Organización
artículo Séptimo Transitorio del Decreto de Reforma Constitucional / Ley	artículo Séptimo Transitorio del Decreto de Ley / Ley	Cuichapa / Lugar
Artículos 1, 7, 8 y 12 del Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Ley	artículos 1 7 8 y 12 del Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Ley	Cumplimiento de Obligaciones / Otro
artículos 15 fracciones IV y LVII y 17 fracción I de la Ley / Ley	artículos 15 fracciones IV y LVII y 17 fracción I de la Ley / Ley	Decreto de Ley / Persona
artículos 28 párrafos décimo quinto, décimo sexto y décimo séptimo de la Constitución Política de los Estados Unidos Mexicanos / Ley	artículos 28 párrafos décimo quinto décimo sexto y décimo séptimo de la Constitución Política de los Estados Unidos Mexicanos / Ley	Decreto de Reforma Constitucional / Otro
artículos 6o., 7o., 27, 28, 73, 78, 94 y 105 de la Constitución Política de los Estados Unidos Mexicanos / Ley	artículos 6o. 7o. 27 28 73 78 94 y 105 de la Constitución Política de los Estados Unidos Mexicanos / Ley	Diario Oficial de la Federación / Organización
artículos 7, 16 y 45 de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	artículos 6o. y 7o. de la Constitución / Ley	Dirección General de Adquisiciones / Organización
concesión para instalar, operar y explotar una red pública de telecomunicaciones para prestar el servicio de televisión restringida en Cuichapa, Municipio de Moloacán, en el Estado de Veracruz, con una vigencia de 10 (diez) años contados a partir de su otorgamiento / Documento	artículos 7 16 y 45 de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Dirección General de Concentraciones / Organización
Constitución Política de los Estados Unidos Mexicanos / Ley	concesión para instalar operar y explotar una red pública de telecomunicaciones para prestar el servicio de televisión restringida en Cuichapa Municipio de Moloacán en el Estado de Veracruz con una vigencia de 10 diez años contados a partir de su otorgamiento / Documento	Dirección General de Concesiones de Telecomunicaciones / Organización

Cuichapa / Lugar	Constitución Política de los Estados Unidos Mexicanos / Documento	Dirección General de Política de Telecomunicaciones / Organización
Decreto por el que se expiden la Ley Federal de Telecomunicaciones y Radiodifusión, y la Ley del Sistema Público de Radiodifusión del Estado Mexicano; y se reforman, adicionan y derogan diversas disposiciones en materia de telecomunicaciones y radiodifusión / Documento	Cuichapa / Lugar	Dirección General de Supervisión / Organización
Decreto por el que se reforman y adicionan diversas disposiciones de los artículos 6o., 7o., 27, 28, 73, 78, 94 y 105 de la Constitución Política de los Estados Unidos Mexicanos, en materia de telecomunicaciones / Documento	Cumplimiento de Obligaciones / Otro	Dirección General de Verificación / Organización
Diario Oficial de la Federación / Documento	Decreto por el que se expiden la Ley Federal de Telecomunicaciones y Radiodifusión y la Ley de el Sistema Público de Radiodifusión de el Estado Mexicano; y se reforman adicionan y derogan diversas disposiciones en materia de telecomunicaciones y radiodifusión / Documento	Dish / Persona
Dirección General de Adquisiciones, Recursos Materiales y Servicios Generales / Organización	Decreto por el que se reforman y adicionan diversas disposiciones de los artículos 6o. 7o. 27 28 73 78 94 y 105 de la Constitución Política de los Estados Unidos Mexicanos en materia de telecomunicaciones / Documento	DTH / Organización
Dirección General de Concentraciones y Concesiones / Organización	Diario Oficial de la Federación / Documento	Ernesto Estrada González / Persona
Dirección General de Concesiones de Telecomunicaciones / Organización	Dirección General de Adquisiciones / Organización	Estado / Organización
Dirección General de Política de Telecomunicaciones y de Radiodifusión / Organización	Dirección General de Concentraciones y Concesiones / Organización	Estado de Veracruz / Lugar

Dirección General de Supervisión / Organización	Dirección General de Concesiones de Telecomunicaciones / Organización	Estatuto Orgánico / Lugar
Dirección General de Verificación / Organización	Dirección General de Política de Telecomunicaciones y de Radiodifusión / Organización	Estatuto Orgánico / Organización
Dish / Organización	Dirección General de Política de Telecomunicaciones y de Radiodifusión de la Secretaría / Organización	Estatuto Orgánico de el Instituto Federal de Telecomunicaciones / Organización
Ernesto Estrada González / Persona	Dirección General de Supervisión / Organización	Estatuto Orgánico de el Instituto Federal de Telecomunicaciones / Otro
Estado de Veracruz / Lugar	Dirección General de Verificación / Organización	Gabriel Oswaldo Contreras Saldívar / Persona
Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Documento	Ernesto Estrada González / Otro	I / Otro
expediente 02/1223 / Documento	Estado de Veracruz / Lugar	I de la Ley / Organización
fracción I del artículo 67 de la Ley / Ley	Estatuto Orgánico de el Instituto Federal de Telecomunicaciones / Documento	I de la Ley Federal de Procedimiento Administrativo / Organización
fracción III del artículo 94 de la Ley Federal de Derechos / Ley	Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Documento	II / Otro
Gabriel Oswaldo Contreras Saldívar / Persona	expediente 02/1223 / Documento	II de el Estatuto Orgánico / Otro
Instituto Federal de Telecomunicaciones / Organización	fracción I del artículo 67 de la Ley / Ley	II de la Constitución / Otro
Jorge Adalberto Esperón Heredia / Persona	fracción III del artículo 94 de la Ley Federal de Derechos / Ley	III / Otro
ley del sistema público de radiodifusión del estado mexicano / Ley	Gabriel Oswaldo Contreras Saldívar / Otro	III de la Ley Federal de Derechos / Organización
Ley Federal de Competencia Económica / Ley	González / Lugar	III de la Ley Federal de Derechos / Otro

Ley Federal de Derechos / Ley	Instituto Federal de Telecomunicaciones / Organización	Instituto / Organización
ley federal de procedimiento administrativo / Ley	Jorge Adalberto Esperón Heredia / Otro	Instituto / Otro
Ley Federal de Telecomunicaciones / Ley	Ley del Sistema Público de Radiodifusión del Estado Mexicano / Ley	Instituto Federal de Telecomunicaciones / Organización
Ley Federal de Telecomunicaciones en su artículo 3 fracción X / Ley	Ley Federal de Competencia Económica / Ley	IV / Organización
Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Ley Federal de Derechos / Ley	JORGE ADALBERTO ESPERÓN HEREDIA / Organización
Luis Fernando Borjón Figueroa / Persona	Ley Federal de Procedimiento Administrativo / Ley	Ley / Otro
Marco normativo general aplicable a prórrogas de vigencia de concesiones en materia de telecomunicaciones / Documento	Ley Federal de Telecomunicaciones / Ley	Ley de el Sistema Público de Radiodifusión de el Estado Mexicano / Otro
María Elena Estavillo Flores / Persona	Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Ley Federal de Competencia Económica / Otro
Mario Germán Fromow Rangel / Persona	Luis Fernando Borjón Figueroa / Otro	Ley Federal de Derechos / Otro
Municipio de Moloacán / Lugar	María Elena Estavillo Flores / Otro	Ley Federal de Telecomunicaciones / Otro
oficio 1.-132 / Documento	Mario Germán Fromow Rangel / Otro	Luis Fernando Borjón Figueroa / Persona
oficio 2.1.-0823 / Documento	Municipio de Moloacán / Otro	LVII / Organización
oficio IFT/223/UCS/649/2015 / Documento	oficio 1.-132 / Documento	LVII / Otro
Oficio IFT/223/UCS/DG-CTEL/1242/2015 / Documento	oficio 2.1.-0823 / Documento	LVIII de la Ley / Organización
oficio IFT/223/UCS/DG-CTEL/1243/2015 / Documento	oficio IFT/223/UCS/649/2015 / Documento	María Elena Estavillo Flores / Persona

oficio IFT/225/UC/DG-SUV/2634/2015 / Documento	oficio IFT/223/UCS/DG-CTEL/1242/2015 / Documento	Mario Germán Fromow Rangel / Persona
oficio IFT/225/UC/DG-VER/1934/2015 / Documento	oficio IFT/223/UCS/DG-CTEL/1243/2015 / Documento	Municipio de Moloacán / Lugar
oficio IFT/226/UCE/DG-CCON/124/2015 / Documento	oficio IFT/225/UC/DG-SUV/2634/2015 / Documento	Opinión / Organización
Opinión en materia de Competencia Económica / Documento	oficio IFT/225/UC/DG-VER/1934/2015 / Documento	Opinión / Otro
Opinión en materia de Cumplimiento de Obligaciones / Documento	oficio IFT/226/UCE/DG-CCON/124/2015 / Documento	Opinión Técnica de la Secretaría / Organización
Opinión Técnica de la Secretaría / Documento	Opinión Técnica de la Secretaría / Documento	Pleno / Organización
Otorgamiento de la Concesión / Otro	Otorgamiento de la Concesión / Documento	Pleno de el Instituto / Organización
párrafo cuarto del artículo 114 de la Ley / Ley	párrafo cuarto del artículo 114 de la Ley / Ley	Pleno de el Instituto Federal de Telecomunicaciones / Organización
párrafos vigésimo, fracciones I y III; y vigésimo primero, del artículo 28 de la Constitución Política de los Estados Unidos Mexicanos / Ley	párrafo y Segundo el Presidente del Instituto Federal de Telecomunicaciones / Ley	Radiodifusión / Organización
Pleno del Instituto Federal de Telecomunicaciones / Organización	párrafos vigésimo fracciones I y III; y vigésimo primero del artículo 28 de la Constitución Política de los Estados Unidos Mexicanos / Ley	Radiodifusión de la Secretaría / Organización
presidente del instituto federal de telecomunicaciones / Persona	Pleno del Instituto / Organización	Recursos Materiales / Persona
Registro Público de Concesiones / Documento	Pleno del Instituto Federal de Telecomunicaciones / Organización	Registro Público de Concesiones / Organización
RESOLUCIÓN POR LA CUAL EL PLENO DEL INSTITUTO FEDERAL DE TELECOMUNICACIONES PRORROGA LA VIGENCIA DE LA CONCESIÓN DEL C. JORGE ADALBERTO ESPERÓN HEREDIA Y OTORGA UN TÍTULO DE CONCESIÓN ÚNICA PARA USO COMERCIAL / Documento	Presidente del Instituto Federal de Telecomunicaciones / Persona	Resolución / Otro

Resolutivo Segundo / Documento	Reforma / Lugar	RESOLUCIÓN POR LA CUAL EL PLENO DE EL INSTITUTO FEDERAL DE TELECOMUNICACIONES PRORROGA LA VIGENCIA DE LA CONCESIÓN DE EL C. JORGE ADALBERTO ESPERÓN HEREDIA Y OTORGA UN TÍTULO DE CONCESIÓN ÚNICA PARA USO COMERCIAL / Organización
Resolutivos Primero cuarto párrafo / Documento	Registro Público de Concesiones / Documento	Resolutivo / Organización
Secretaría de Comunicaciones y Transportes / Organización	Séptimo Transitorio del Decreto de Ley / Documento	Resolutivo Primero / Lugar
Séptimo Transitorio de "Decreto por el que se expiden la Ley Federal de Telecomunicaciones y Radiodifusión, y la Ley del Sistema Público de Radiodifusión del Estado Mexicano; y se reforman, adicionan y derogan diversas disposiciones en materia de telecomunicaciones y radiodifusión" / Ley	Solicitud de Opinión Técnica / Documento	Resolutivo Segundo / Organización
Sky / Organización	Solicitud de Prórroga de Vigencia / Documento	Resolutivos Primero / Organización
Solicitud de Opinión Técnica / Documento	solicitud de prórroga de vigencia de la Concesión / Documento	RESOLUTIVOS PRIMERO / Persona
Solicitud de Prórroga de Vigencia / Documento	Unidad de Competencia Económica / Organización	Secretaría / Organización
solicitud de prórroga de vigencia de la Concesión / Documento	Unidad de Concesiones y Servicios / Organización	Secretaría de Comunicaciones / Organización
Título de concesión única para uso comercial / Documento	Unidad de Cumplimiento / Organización	Segundo / Otro
Unidad de Competencia Económica / Organización		Séptimo Transitorio / Persona
Unidad de Concesiones y Servicios / Organización		Séptimo Transitorio de el Decreto de Ley / Persona

Unidad de Cumplimiento / Organización		Servicios / Organización
		Servicios Generales / Persona
		Sky / Persona
		Solicitante / Persona
		Solicitud de Opinión Técnica / Otro
		Solicitud de Prórroga / Lugar
		Solicitud de Prórroga / Organización
		Solicitud de Prórroga / Otro
		Solicitud de Prórroga / Persona
		Solicitud de Prórroga de Vigencia / Organización
		Transportes / Organización
		TV / Organización
		Unidad de Competencia Económica / Organización
		Unidad de Concesiones / Organización
		Unidad de Cumplimiento / Organización
		Veracruz / Persona
		X / Otro
		X de el Estatuto Orgánico / Otro

		XII de el Estatuto Orgánico de el Instituto Federal de Telecomunicaciones / Organización
		XV / Organización
		XVIII / Organización
		XVIII de el Estatuto Orgánico / Otro
		XXIX Sesión Ordinaria / Otro
		C. Jorge Adalberto Esperón Heredia / Organización

Documento 2016

Etiquetado Manual Entidad/clasificación	Software NERC Entidad/clasificación	FreeLing 4.0 Entidad/clasificación
Comisionado Presidente del Instituto Federal de Telecomunicaciones / Persona	Secretaría de Comunicaciones y Transportes / Organización	Estatuto Orgánico de el Instituto Federal de Telecomunicaciones / Organización
1, 6 fracciones I, XVIII y XXXVII, 32 y 33 fracción II, 41, 42 fracciones I, II y XV y 50 fracción XII del Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Ley	1 6 fracciones I XVIII y XXXVII 32 y 33 fracción II 41 42 fracciones I II y XV y 50 fracción XII del Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Ley	Acatlán / Lugar
27 de la Ley Federal de Telecomunicaciones / Ley	35 fracción I 36 38 39 y 57 fracción I de la Ley Federal de Procedimiento Administrativo / Ley	Acuerdo / Lugar
35 fracción I, 36, 38, 39 y 57 fracción I de la Ley Federal de Procedimiento Administrativo / Ley	6 fracción IV 15 fracciones IV y LVII 16 17 fracción I 66 67 fracción I 68 y 72 de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Acuerdo de Pago de Aprovechamientos / Otro

6 fracción IV, 15 fracciones IV y LVII, 16, 17 fracción I, 66, 67 fracción I, 68 y 72 de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	94 fracción III de la Ley Federal de Derechos / Ley	Acuerdo de Pago de Aprovechamientos / Persona
8, segundo párrafo, del Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Ley	Acuerdo mediante el cual el Pleno de el Instituto Federal de Telecomunicaciones fija el monto de los aprovechamientos que deberán cobrarse por la prestación de diversos servicios públicos en el ejercicio de sus funciones de derecho público por los que no se establece monto específico en la Ley Federal de Derechos / Documento	Adolfo Cuevas Teja / Persona
94 fracción III de la Ley Federal de Derechos / Ley	Acuerdo P/IFT/090316/82 / Documento	Adriana Sofía Labardini Inzunza / Persona
Acuerdo de Pago de Aprovechamientos / Documento	Adolfo Cuevas Teja / Otro	Análisis de la Solicitud de Prórroga / Lugar
Acuerdo mediante el cual el Pleno del Instituto Federal de Telecomunicaciones fija el monto de los aprovechamientos que deberán cobrarse por la prestación de diversos servicios públicos en el ejercicio de sus funciones de derecho público por los que no se establece monto específico en la Ley Federal de Derechos / Documento	Adriana Sofía Labardini Inzunza / Otro	Antecedentes / Organización
acuerdo p/ift/090316/82 / Documento	artículo 14 fracción X del Estatuto Orgánico / Ley	ANTECEDENTES Otorgamiento de la Concesión / Otro
Adolfo Cuevas Teja / Persona	artículo 174-B fracción I inciso a de la Ley Federal de Derechos / Ley	Apartado B / Persona
Adriana Sofía Labardini Inzunza / Persona	artículo 27 de la abrogada Ley Federal de Telecomunicaciones / Ley	Asimismo / Persona
Ampliación de cobertura / Otro	artículo 28 párrafo décimo séptimo de la Constitución / Ley	C. Mario Alberto Radilla Hernández / Organización
artículo 14 fracción X del Estatuto Orgánico / Ley	artículo 28 párrafo décimo séptimo de la Constitución Política de los Estados Unidos Mexicanos / Ley	C. Mario Alberto Radilla Hernández / Otro

artículo 174-B / Ley	artículo 28 párrafos décimo quinto décimo sexto y décimo séptimo de la Constitución / Ley	C. Mario Alberto Radilla Hernández / Persona
artículo 174-B fracción I inciso a) de la Ley Federal de Derechos / Ley	artículo 3 fracción LVIII de la Ley / Ley	C. Sandra Mireia Meraz Ponce / Persona
artículo 27 de la abrogada Ley Federal de Telecomunicaciones / Ley	artículo 6 fracciones I y XVIII del Estatuto Orgánico / Ley	Capítulo / Persona
artículo 27 de la LFT / Ley	artículo 6o. Apartado B fracción II de la Constitución / Ley	Capítulo IX de el Título I / Otro
artículo 28 párrafo décimo séptimo de la Constitución / Ley	artículo Séptimo Transitorio del Decreto de Ley / Ley	CC / Organización
artículo 28 párrafo décimo séptimo de la Constitución Política de los Estados Unidos Mexicanos / Ley	artículo Séptimo Transitorio del Decreto de Reforma Constitucional / Ley	Comisionada Adriana Sofía Labardini Inzunza / Organización
artículo 28 párrafos décimo quinto, décimo sexto y décimo séptimo de la Constitución / Ley	artículo Sexto Transitorio del Decreto de Ley / Ley	Comisionado Presidente de el Instituto Federal de Telecomunicaciones / Organización
artículo 3 fracción LVIII de la Ley / Ley	artículos 1 7 8 y 12 del Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Ley	Comisionados Gabriel Oswaldo Contreras Saldívar / Organización
artículo 6 fracciones I y XVIII del Estatuto Orgánico / Ley	artículos 15 fracción IV y 17 fracción I de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Competencia / Otro
artículo 66 / Ley	artículos 28 párrafos décimo quinto décimo sexto y décimo séptimo de la Constitución Política de los Estados Unidos Mexicanos / Ley	Competencia Económica / Organización
artículo 67 de la Ley / Ley	artículos 32 y 33 fracción II del Estatuto Orgánico / Ley	Concesión / Lugar
artículo 6o. Apartado B fracción II de la Constitución / Ley	artículos 45 tercer párrafo de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Concesión / Organización
artículo Séptimo Transitorio del Decreto de Ley / Ley	artículos 6o. 7o. 27 28 73 78 94 y 105 de la Constitución Política de los Estados Unidos Mexicanos / Ley	Concesión / Otro

artículo Sexto Transitorio del Decreto de Ley / Ley	artículos 6o. y 7o. de la Constitución / Ley	Concesiones / Organización
artículos 1, 7, 8 y 12 del Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Ley	artículos 7 16 y 45 de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Concesiones de la Unidad de Competencia Económica / Organización
artículos 15 fracción IV y 17 fracción I de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	artículos 91 93 94 94-A 95 96 97 98 99 100 101 102 y 105 de la Ley Federal de Derechos / Ley	CONSIDERANDO Primero / Persona
artículos 28 párrafos décimo quinto, décimo sexto y décimo séptimo de la Constitución Política de los Estados Unidos Mexicanos / Ley	concesión para instalar operar y explotar una red pública de telecomunicaciones para prestar el servicio de televisión restringida en Naolinco [debiendo ser Naolinco de Victoria] en el Estado de Veracruz con una vigencia de 10 diez años contados a partir de su otorgamiento / Documento	Constitución / Otro
artículos 32 y 33 fracción II del Estatuto Orgánico / Ley	Constitución Política de los Estados Unidos Mexicanos / Documento	Constitución Política de los Estados Unidos Mexicanos / Organización
artículos 45, tercer párrafo de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Cumplimiento de Obligaciones / Otro	Cumplimiento de Obligaciones / Otro
artículos 6o. y 7o. de la Constitución / Ley	Decreto por el que se expiden la Ley Federal de Telecomunicaciones y Radiodifusión y la Ley de el Sistema Público de Radiodifusión de el Estado Mexicano; y se reforman adicionan y derogan diversas disposiciones en materia de telecomunicaciones y radiodifusión / Documento	Decreto de Ley / Persona
artículos 6o., 7o., 27, 28, 73, 78, 94 y 105 de la Constitución Política de los Estados Unidos Mexicanos / Ley	Decreto por el que se reforman y adicionan diversas disposiciones de los artículos 6o. 7o. 27 28 73 78 94 y 105 de la Constitución Política de los Estados Unidos Mexicanos en materia de telecomunicaciones / Documento	Decreto de Reforma Constitucional / Organización
artículos 7, 16 y 45 de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Diario Oficial de la Federación / Documento	Decreto de Reforma Constitucional / Otro

artículos 91, 93, 94, 94-A, 95, 96, 97, 98, 99, 100, 101, 102 y 105 de la Ley Federal de Derechos / Ley	Dirección General de Adquisiciones / Organización	Diario Oficial de la Federación / Organización
Capítulo IX del Título I / Documento	Dirección General de Concentraciones y Concesiones / Organización	Dirección General / Organización
concesión para instalar, operar y explotar una red pública de telecomunicaciones para prestar el servicio de televisión restringida en Naolinco [debiendo ser "Naolinco de Victoria"], en el Estado de Veracruz, con una vigencia de 10 (diez) años contados a partir de su otorgamiento / Documento	Dirección General de Concesiones de Telecomunicaciones / Organización	Dirección General de Adquisiciones / Organización
Constitución Política de los Estados Unidos Mexicanos / Documento	Dirección General de Licitaciones de Espectro Radioeléctrico y Servicios / Organización	Dirección General de Concentraciones / Organización
Decreto por el que se expiden la Ley Federal de Telecomunicaciones y Radiodifusión, y la Ley del Sistema Público de Radiodifusión del Estado Mexicano; y se reforman, adicionan y derogan diversas disposiciones en materia de telecomunicaciones y radiodifusión / Documento	Dirección General de Política de Telecomunicaciones y de Radiodifusión / Organización	Dirección General de Concesiones de Telecomunicaciones / Organización
Decreto por el que se reforman y adicionan diversas disposiciones de los artículos 6o., 7o., 27, 28, 73, 78, 94 y 105 de la Constitución Política de los Estados Unidos Mexicanos, en materia de telecomunicaciones / Documento	Dirección General de Política de Telecomunicaciones y de Radiodifusión de la Secretaría / Organización	Dirección General de Licitaciones de Espectro Radioeléctrico / Organización
Decreto por el que se reforman, adicionan y derogan diversas disposiciones de la Ley Federal de Derechos / Documento	Dirección General de Supervisión / Organización	Dirección General de Política de Telecomunicaciones / Organización
Diario Oficial de la Federación / Documento	Dirección General de Verificación / Organización	Dirección General de Supervisión / Organización
Dirección General de Adquisiciones, Recursos Materiales y Servicios Generales / Organización	Ernesto Estrada González / Otro	Dirección General de Verificación / Organización
Dirección General de Concentraciones y Concesiones / Organización	Estado de Veracruz / Lugar	Dish / Persona
Dirección General de Concesiones de Telecomunicaciones / Organización	Estatuto Orgánico de el Instituto Federal de Telecomunicaciones /	DTH / Organización

	Documento	
Dirección General de Licitaciones de Espectro Radioeléctrico y Servicios / Organización	Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Documento	Ernesto Estrada González / Persona
Dirección General de Política de Telecomunicaciones y de Radiodifusión / Organización	expediente 02/0799 / Documento	Estado / Organización
Dirección General de Supervisión / Organización	fracción I del artículo 67 de la Ley / Ley	Estado de Veracruz / Lugar
Dirección General de Verificación / Organización	Gabriel Oswaldo Contreras Saldívar / Otro	Estatuto Orgánico / Lugar
Dish / Organización	González / Lugar	Estatuto Orgánico / Organización
Ernesto Estrada González / Persona	Instituto Federal de Telecomunicaciones / Organización	Estatuto Orgánico de el Instituto Federal de Telecomunicaciones / Otro
Estado de Veracruz / Lugar	José Luis Oliva Cano / Otro	I / Otro
Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Documento	Ley del Sistema Público de Radiodifusión del Estado Mexicano / Ley	I de la Ley Federal de Procedimiento Administrativo / Organización
Expediente 02/0799 / Documento	Ley Federal de Derechos / Ley	I de la Ley Federal de Telecomunicaciones / Organización
fracción I del artículo 67 de la Ley / Ley	Ley Federal de Procedimiento Administrativo / Ley	II / Otro
Gabriel Oswaldo Contreras Saldívar / Persona	Ley Federal de Telecomunicaciones / Ley	II de el Estatuto Orgánico / Otro
Instituto Federal de Telecomunicaciones / Organización	Ley Federal de Telecomunicaciones y Radiodifusión / Ley	II de la Constitución / Otro
José Luis Oliva Cano / Persona	María Elena Estavillo Flores / Otro	III / Otro
Ley del Sistema Público de Radiodifusión del Estado Mexicano / Ley	Mario Alberto Radilla Hernández / Otro	III de la Ley Federal de Derechos / Organización

Ley Federal de Derechos / Ley	Mario Germán Fromow Rangel / Otro	Industria de el Instituto / Organización
ley federal de procedimiento administrativo / Ley	Municipio de Acatlán / Otro	Instituto / Organización
Ley Federal de Telecomunicaciones / Ley	Municipio de Miahuatlán / Otro	Instituto / Otro
Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Naolinco / Lugar	Instituto Federal de Telecomunicaciones / Organización
LFT en su artículo 3 fracción X / Ley	Naolinco de Victoria / Otro	IV / Organización
María Elena Estavillo Flores / Persona	oficio 1.-74 / Documento	IV / Otro
Mario Alberto Radilla Hernández / Persona	oficio 2.1.-0339 / Documento	José Luis Oliva Cano / Persona
Mario Germán Fromow Rangel / Persona	oficio IFT/223/UCS/DG-CTEL/094/2014 / Documento	Ley / Otro
Municipio de Acatlán / Lugar	oficio IFT/225/UC/DG-SUV/6354/2015 / Documento	Ley de el Sistema Público de Radiodifusión de el Estado Mexicano / Otro
Municipio de Miahuatlán / Lugar	oficio IFT/225/UC/DG-VER/5468/2015 / Documento	Ley Federal de Derechos / Otro
Naolinco / Lugar	oficio IFT/226/UCE/DG-CCON/074/2015 / Documento	Ley Federal de Telecomunicaciones / Otro
Naolinco de Victoria / Lugar	oficio IFT/D01/P/153/2014 / Documento	LFT / Organización
Oficio 1.-74 / Documento	oficio IFT/D03/USI/DGLS/74/2013 / Documento	LVII / Otro
Oficio 2.1.-0339 / Documento	Opinión Técnica de la Secretaría / Documento	LVIII de la Ley / Organización
oficio IFT/223/UCS/DG-CTEL/094/2014 / Documento	Otorgamiento de la Concesión / Documento	María Elena Estavillo Flores / Persona
oficio IFT/225/UC/DG-SUV/6354/2015 / Documento	párrafo del Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Ley	Mario Alberto Radilla Hernández / Persona

oficio IFT/225/UC/DG-VER/5468/2015 / Documento	párrafo segundo del artículo Séptimo Transitorio del Decreto de Reforma Constitucional / Ley	Mario Germán Fromow Rangel / Persona
oficio IFT/226/UC/DG-CCON/074/2015 / Documento	párrafo y Segundo el Presidente del Instituto Federal de Telecomunicaciones / Ley	Miahuatlán / Lugar
oficio IFT/D01/P/153/2014 / Documento	párrafos vigésimo fracciones I y III; y vigésimo primero del artículo 28 de la Constitución Política de los Estados Unidos Mexicanos / Ley	Miahuatlán / Persona
oficio IFT/D03/USI/DGLS/74/2013 / Documento	Pleno del Instituto / Organización	Municipio de Acatlán / Lugar
opinión técnica de la secretaría / Documento	Pleno del Instituto Federal de Telecomunicaciones / Organización	Municipio de Miahuatlán / Persona
Otorgamiento de la Concesión / Otro	Presidente del Instituto Federal de Telecomunicaciones / Persona	Naolinco / Lugar
párrafo segundo del artículo Séptimo Transitorio del Decreto de Reforma Constitucional / Ley	Reforma / Lugar	Naolinco de Victoria / Organización
párrafos vigésimo, fracciones I y III; y vigésimo primero, del artículo 28 de la Constitución Política de los Estados Unidos Mexicanos / Ley	Registro Público de Concesiones / Documento	Opinión / Organización
Pleno del Instituto Federal de Telecomunicaciones / Organización	Sandra Mirea Meraz Ponce / Otro	Opinión / Otro
Registro Público de Concesiones / Documento	Sección Primera del Capítulo / Ley	Opinión Técnica de la Secretaría / Organización
RESOLUCIÓN MEDIANTE LA CUAL EL PLENO DEL INSTITUTO FEDERAL DE TELECOMUNICACIONES PRORROGA LA VIGENCIA DE LA CONCESIÓN DEL C. MARIO ALBERTO RADILLA HERNÁNDEZ, Y OTORGA UN TÍTULO DE CONCESIÓN ÚNICA PARA USO COMERCIAL. / Documento	Séptimo Transitorio del Decreto de Ley / Documento	Pleno / Organización
Resolutivo Segundo / Documento	Séptimo Transitorio del Decreto de Reforma Constitucional / Documento	Pleno de el Instituto / Organización
Resolutivos Primero cuarto párrafo / Documento	Servicios de Telecomunicaciones / Otro	Pleno de el Instituto Federal de Telecomunicaciones / Organización

Sandra Mirea Meraz Ponce / Persona	Sexto Transitorio del Decreto de Ley / Documento	Radiodifusión / Organización
Sección Primera del Capítulo VIII del Título I / Ley	Solicitud de Opinión Técnica / Documento	Radiodifusión de la Secretaría / Organización
Secretaría de Comunicaciones y Transportes / Organización	Solicitud de Prórroga de Vigencia / Documento	Recursos Materiales / Persona
Séptimo Transitorio del "Decreto por el que se reforman y adicionan diversas disposiciones de los artículos 6o., 7o., 27, 28, 73, 78, 94 y 105 de la Constitución Política de los Estados Unidos Mexicanos, en materia de telecomunicaciones" / Ley	solicitud de prórroga de vigencia de la Concesión / Documento	Registro Público de Concesiones / Organización
Sexto y Séptimo Transitorios de "Decreto por el que se expiden la Ley Federal de Telecomunicaciones y Radiodifusión, y la Ley del Sistema Público de Radiodifusión del Estado Mexicano; y se reforman, adicionan y derogan diversas disposiciones en materia de telecomunicaciones y radiodifusión" / Ley	Subsecretaría de Comunicaciones de la Secretaría / Organización	Resolución / Otro
Sky / Organización	Unidad de Competencia Económica / Organización	RESOLUCIÓN MEDIANTE LA CUAL EL PLENO DE EL INSTITUTO FEDERAL DE TELECOMUNICACIONES PRORROGA LA VIGENCIA DE LA CONCESIÓN DE EL C. MARIO ALBERTO RADILLA HERNÁNDEZ / Persona
solicitud de opinión técnica / Documento	Unidad de Concesiones y Servicios / Organización	Resolutivo / Organización
Solicitud de Prórroga de Vigencia / Documento	Unidad de Cumplimiento / Organización	Resolutivo Primero / Lugar
solicitud de prórroga de vigencia de la Concesión / Documento	Unidad de Servicios a la Industria del Instituto / Organización	Resolutivo Segundo / Organización
Subsecretaría de Comunicaciones de la Secretaría / Organización	Unidad de Supervisión y Verificación / Organización	Resolutivos Primero / Organización
Título de concesión única para uso comercial / Documento	Victoria / Lugar	RESOLUTIVOS PRIMERO / Persona

Unidad de Competencia Económica / Organización		Sección Primera de el Capítulo VIII de el Título I / Otro
Unidad de Concesiones y Servicios / Organización		Secretaría / Organización
Unidad de Cumplimiento / Organización		Secretaría de Comunicaciones / Organización
Unidad de Servicios a la Industria del Instituto / Organización		Segundo / Otro
Unidad de Supervisión y Verificación / Organización		Séptimo Transitorio / Persona
		Séptimo Transitorio de el Decreto de Ley / Persona
		Séptimo Transitorio de el Decreto de Reforma Constitucional / Otro
		Séptimo Transitorios / Persona
		Servicios / Organización
		Servicios de Telecomunicaciones / Otro
		Servicios Generales / Persona
		Sesión / Organización
		Sexto / Persona
		Sexto Transitorio de el Decreto de Ley / Otro
		Sky / Persona
		Solicitante / Otro

		Solicitud de Opinión Técnica / Otro
		Solicitud de Prórroga / Lugar
		Solicitud de Prórroga / Organización
		Solicitud de Prórroga / Otro
		Solicitud de Prórroga / Persona
		Solicitud de Prórroga de Vigencia / Organización
		Subsecretaría de Comunicaciones de la Secretaría / Organización
		Transportes / Organización
		TV / Organización
		Unidad de Competencia Económica / Organización
		Unidad de Concesiones / Organización
		Unidad de Cumplimiento / Organización
		Unidad de Servicios / Organización
		Unidad de Supervisión / Organización
		Verificación / Organización
		VII Sesión Ordinaria / Otro
		X / Otro

		X de el Estatuto Orgánico / Otro
		XII de el Estatuto Orgánico de el Instituto Federal de Telecomunicaciones / Organización
		XV / Organización
		XVIII / Organización
		XVIII de el Estatuto Orgánico / Otro
		XXXVII / Organización
		Y OTORGA UN TÍTULO DE CONCESIÓN ÚNICA PARA USO COMERCIAL / Persona

Documento 2017

Etiquetado Manual Entidad/clasificación	Software NERC Entidad/clasificación	FreeLing 4.0 Entidad/clasificación
Opinión de la Secretaría de Hacienda y Crédito Público / Documento	Acuerdo por el que se establecen los requisitos para llevar a cabo el cambio de frecuencias autorizadas para prestar el servicio de radio y que operan en la banda de Amplitud Modulada / Documento	Estudios Técnicos / Organización
1, 3 y 16 fracción X de la Ley Federal de Procedimiento Administrativo / Ley	1 4 fracción V inciso iii 29 fracción VII 32 y 34 fracción XIV del Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Ley	A / Otro
1, 4 fracción V, inciso iii), 29, fracción VII, 32 y 34, fracción XIV del Estatuto Orgánico del Instituto Federal de Telecomunicaciones	16 fracción X de la Ley Federal de Procedimiento Administrativo / Ley	Acuerdo / Lugar

/ Ley		
16 fracción X de la Ley Federal de Procedimiento Administrativo / Ley	20 fracción XI del Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Ley	Acuerdo / Organización
20 fracción xi del estatuto orgánico del instituto federal de telecomunicaciones / Ley	29 fracción VII del Estatuto Orgánico del Instituto / Ley	Acuerdo / Otro
29, fracción VII del Estatuto Orgánico del Instituto / Ley	7 15 fracción VIII 16 y 17 fracción I 99 100 155 y 156 así como Tercero y Sexto Transitorios de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Acuerdo / Persona
3° del Código Fiscal de la Federación / Ley	Acuerdo P/IFT/250117/27 / Documento	Acuerdo de Transición / Otro
41 de la Ley Federal de Radio y Televisión / Ley	Adolfo Cuevas Teja / Otro	Adolfo Cuevas Teja / Persona
7, 15, fracción VIII, 16 y 17, fracción I, 99, 100, 155 y 156, así como Tercero y Sexto Transitorios de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Adriana Sofía Labardini Inzunza / Otro	Adriana Sofía Labardini Inzunza / Persona
Acuerdo P/IFT/250117/27 / Documento	artículo 10 de la Ley de Ingresos de la Federación / Ley	Ahora bien / Organización
Acuerdo por el que se establecen los requisitos para llevar a cabo el cambio de frecuencias autorizadas para prestar el servicio de radio y que operan en la banda de Amplitud Modulada / Documento	artículo 124 fracción II inciso e de la Ley Federal de Derechos / Ley	Altura / Lugar
Adolfo Cuevas Teja / Persona	artículo 134 de la Constitución Política de los Estados Unidos Mexicanos / Ley	Altura / Otro
Adriana Sofía Labardini Inzunza / Persona	artículo 155 de la Ley / Ley	ALTURA DE EL SISTEMA RADIADOR La / Otro
artículo 10 de la Ley de Ingresos de la Federación / Ley	artículo 156 de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	AM / Lugar
artículo 100 de Ley / Ley	artículo 28 de la Constitución Política de los Estados Unidos	AM / Organización

	Mexicanos / Ley	
artículo 124, fracción II, inciso e) de la Ley Federal de Derechos / Ley	artículo 28 párrafo décimo quinto y décimo sexto de la Constitución / Ley	Amplitud Modulada / Persona
artículo 134 de la Constitución Política de los Estados Unidos Mexicanos / Ley	artículo 29 fracción VII del Estatuto Orgánico del Instituto / Ley	Análisis de la Solicitud de Modificación Técnica / Otro
artículo 155 de la Ley / Ley	artículo 32 del Estatuto Orgánico / Ley	Antecedente II / Organización
artículo 156 de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	artículo 6 fracción I del Estatuto Orgánico / Ley	Antecedente VI / Organización
artículo 28 de la Constitución Política de los Estados Unidos Mexicanos / Ley	Artículo 99. Todas las contraprestaciones a que se refiere esta Ley requerirán previa opinión no vinculante de la Secretaría de Hacienda y Crédito Público / Ley	Antecedente VII / Organización
artículo 28 párrafo décimo quinto y décimo sexto de la Constitución / Ley	artículo Séptimo Transitorio del Decreto / Ley	Artículo / Otro
artículo 29, fracción VII del Estatuto Orgánico del Instituto / Ley	artículo Séptimo Transitorio del Decreto de Reforma / Ley	Artículos Cuarto / Organización
artículo 32 del Estatuto Orgánico / Ley	artículo Sexto Transitorio del Decreto / Ley	AS-FM / Lugar
artículo 34, fracción XIV del ordenamiento jurídico en cita / Ley	artículos 1 7 8 y 12 del Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Ley	AS-FM / Organización
artículo 41 de la LFRT / Ley	artículos 10 y 12 de la Ley de Ingresos / Ley	AS-FM-I-II / Lugar
artículo 6 fracción I del Estatuto Orgánico / Ley	artículos 15 fracción VIII 16 y 17 fracción I 99 y 100 de la Ley / Ley	AS-FM-I-II / Organización
artículo Séptimo Transitorio del Decreto / Ley	artículos 15 fracción VIII 99 y 100 de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Asimismo / Otro
artículo Séptimo Transitorio del Decreto de Reforma / Ley	artículos 15 fracción XXVIII de la Ley Federal de Telecomunicaciones y	B / Otro

	Radiodifusión / Ley	
artículo Séptimo Transitorio del Decreto por el que se reforman y adicionan diversas disposiciones de los artículos 6o., 7o., 27, 28, 73, 78, 94 y 105 de la Constitución Política de los Estados Unidos Mexicanos / Ley	artículos 6º. 7o 27 28 73 78 94 y 105 de la Constitución Política de los Estados Unidos Mexicanos / Ley	Capitulo / Organización
artículo sexto transitorio del decreto / Ley	artículos 6o. 7o. 27 28 73 78 94 y 105 de la Constitución Política de los Estados Unidos Mexicanos / Ley	Capítulo / Otro
artículos 1, 7, 8 y 12 del Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Ley	artículos 6o. y 28 de la Constitución / Ley	Características Técnicas de la Estación / Otro
artículos 10 y 12 de la Ley de Ingresos / Ley	artículos 6o. y 7o. de la Constitución / Ley	Cerro Paquimex S / Persona
artículos 15 fracción VIII, 99 y 100 de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	artículos 7 16 y 45 de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Chihuahua / Organización
artículos 15, fracción VIII, 16 y 17, fracción I 99 y 100 de la Ley / Ley	artículos 99 y 100 de la Ley / Ley	Chihuahua / Persona
artículos 15, fracción XXVIII de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	artículos referidos se desprende que para la tramitación y resolución de la Solicitud / Ley	CHIHUAHUA ANTECEDENTES Refrendo de la Concesión / Persona
artículos 25, 26, 27 y 28 que conforman su capítulo económico / Ley	Características Técnicas de la Estación / Otro	Chihuahua Coordenadas Geográficas / Persona
artículos 6 apartado B, fracción III, 28 párrafos décimo quinto y décimo séptimo, así como artículo Séptimo Transitorio del "Decreto por el que se reforman y adicionan diversas disposiciones de los artículos 6o., 7o., 27, 28, 73, 78, 94 y 105 de la Constitución Política de los Estados Unidos Mexicanos, en materia de telecomunicaciones / Ley	Casas / Lugar	Clase de Estación / Lugar
artículos 6º. 7o 27 28 73 78 94 y 105 de la constitución política de los estados unidos mexicanos / Ley	Casas Grandes / Lugar	Cobertura / Otro

artículos 6o. y 28 de la Constitución / Ley	Chihuahua / Lugar	Código Fiscal de la Federación / Organización
artículos 6o. y 7o. de la Constitución / Ley	Clase de Estación / Otro	Cofetel / Organización
artículos 6o., 7o., 27, 28, 73, 78, 94 y 105 de la Constitución Política de los Estados Unidos Mexicanos / Ley	Código Fiscal de la Federación / Documento	COM-FM / Organización
artículos 7, 16 y 45 de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Comisión Federal de Competencia / Organización	COM-FM-I / Organización
artículos 99 y 100 de la Ley / Ley	Comisión Federal de Competencia Económica / Organización	COM-FM-I-II / Organización
Artículos Cuarto y Quinto de dicho Acuerdo / Ley	Comisión Federal de Telecomunicaciones / Organización	Comisión / Organización
autorización para cambiar la ubicación de la antena y planta transmisora de la estación XHTX-FM / Documento	Constitución Política de los Estados Unidos Mexicanos / Documento	Comisión Federal de Competencia / Organización
Capitulo 10, numerales 10.3 y 10.4 de la "Norma Oficial Mexicana, NOM-02-SCT1-93, Especificaciones y requerimientos para la instalación y operación de estaciones de radiodifusión sonora en la banda de 88 a 108 MHz, con portadora principal modulada en frecuencia / Documento	Croquis de Operación Múltiple / Otro	Comisión Federal de Competencia Económica / Organización
Características Técnicas de la Estación / Otro	Decreto por el que se expiden la Ley Federal de Telecomunicaciones y Radiodifusión y la Ley de el Sistema Público de Radiodifusión de el Estado Mexicano; y se reforman adicionan y derogan diversas disposiciones en materia de telecomunicaciones y radiodifusión / Documento	Comisión Federal de Telecomunicaciones / Organización
Cerro Paquimex S/N / Lugar	Decreto por el que se reforman y adicionan diversas disposiciones de los artículos 6o. 7o. 27 28 73 78 94 y 105 de la Constitución Política de los Estados Unidos Mexicanos en materia de telecomunicaciones / Documento	Comisionados Gabriel Oswaldo Contreras Saldívar / Organización

Chihuahua / Lugar	Diario Oficial de la Federación / Documento	Competencia / Otro
Código Fiscal de la Federación / Documento	Dirección General Adjunta de Desarrollo / Organización	Concesión / Otro
Cofetel / Organización	Dirección General Adjunta de Desarrollo de la Radiodifusión / Organización	Concesión Radio Casas Grandes / Organización
Comisión Federal de Competencia / Organización	Dirección General Adjunta de Trámites y Servicios / Organización	Concesionario / Organización
Comisión Federal de Competencia Económica / Organización	Dirección General Adjunta del Registro Público de Telecomunicaciones / Organización	Concesionario / Otro
Comisión Federal de Telecomunicaciones / Organización	Dirección General de Aeronáutica Civil / Organización	Concesionario Factor / Persona
comprobante de pago de derechos con número de folio 665130008561 / Documento	Dirección General de Concesiones de Radiodifusión / Organización	CONSIDERANDO Primero / Persona
Constitución Política de los Estados Unidos Mexicanos / Documento	Dirección General de Economía del Espectro y Recursos Orbitales / Organización	Constitución / Otro
Croquis de Operación Múltiple / Documento	Dirección General de Ingeniería del Espectro y Estudios Técnicos / Organización	Constitución Política de los Estados Unidos Mexicanos / Organización
Decreto por el que se expiden la Ley Federal de Telecomunicaciones y Radiodifusión, y la Ley del Sistema Público de Radiodifusión del Estado Mexicano; y se reforman, adicionan y derogan diversas disposiciones en materia de telecomunicaciones y radiodifusión / Documento	Entidades Federativas de República Mexicana / Otro	Consumidor / Organización
Decreto por el que se reforman y adicionan diversas disposiciones de los artículos 6o., 7o., 27, 28, 73, 78, 94 y 105 de la Constitución Política de los Estados Unidos Mexicanos, en materia de telecomunicaciones / Documento	Ernesto Estrada González / Otro	Contraprestación VR / Organización
Diario Oficial de la Federación / Documento	Estaciones de Radio / Otro	CP / Organización

Dirección General Adjunta de Desarrollo / Organización	Estatuto Orgánico de el Instituto Federal de Telecomunicaciones / Documento	Crédito Público / Organización
Dirección General Adjunta de Desarrollo de la Radiodifusión / Organización	Estatuto Orgánico del Instituto / Documento	Crédito Público / Persona
Dirección General Adjunta de Trámites y Servicios / Organización	Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Documento	Croquis de Operación Múltiple / Lugar
Dirección General Adjunta del Registro Público de Telecomunicaciones / Organización	Estudio de Predicción / Otro	Croquis de Operación Múltiple / Organización
dirección general de aeronáutica civil / Organización	Frecuencia Modulada / Otro	Croquis de Operación Múltiple / Otro
Dirección General de Concesiones de Radiodifusión / Organización	Gabriel Oswaldo Contreras Saldívar / Otro	CTE-FM-II-III / Organización
Dirección General de Economía del Espectro y Recursos Orbitales / Organización	González / Lugar	Decreto / Lugar
Dirección General de Ingeniería del Espectro y Estudios Técnicos / Organización	Instituto a la Secretaría de Hacienda y Crédito Público / Organización	Decreto / Organización
Disposición Técnica IFT-002-2016 / Documento	Instituto Federal de Telecomunicaciones / Organización	Decreto / Otro
Ejercicio Fiscal de 2016 / Otro	Javier Juárez Mojica / Otro	Decreto / Persona
Entidades Federativas de República Mexicana / Otro	Juárez / Lugar	Decreto de Ley / Otro
Ernesto Estrada González / Persona	Ley de Ingresos / Ley	Decreto de Ley / Persona
Estación de Radiodifusión Sonora / Otro	Ley de Ingresos de la Federación / Ley	Decreto de Reforma Constitucional / Otro
estatuto orgánico del instituto / Organización	Ley del Sistema Público de Radiodifusión del Estado Mexicano / Ley	Desarrollo Económico / Organización
Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Documento	Ley Federal de Derechos / Ley	DGADR / Organización

Estudio de Predicción de Áreas de Servicio / Documento	Ley Federal de Procedimiento Administrativo / Ley	Diario Oficial de la Federación / Organización
fracciones I a VI de este artículo / Ley	Ley Federal de Radio y Televisión / Ley	Dictamen / Persona
frecuencia modulada / Otro	Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Dirección General / Organización
Gabriel Oswaldo Contreras Saldívar / Persona	Marco Antonio Delgado Merchan / Otro	Dirección General Adjunta de Desarrollo / Organización
III Sesión Ordinaria / Otro	María Elena Estavillo Flores / Otro	Dirección General Adjunta de Desarrollo de la Radiodifusión / Organización
Instituto Federal de Telecomunicaciones / Organización	Mario Germán Fromow Rangel / Otro	Dirección General Adjunta de el Registro Público de Telecomunicaciones / Organización
Javier Juárez Mojica / Persona	Nacional de Precios / Otro	Dirección General Adjunta de Trámites / Organización
Ley de Ingresos / Ley	Norma Oficial Mexicana / Otro	Dirección General de Aeronáutica Civil de la Secretaría de Comunicaciones / Organización
Ley de Ingresos de la Federación / Ley	Nuevo Casas Grandes / Otro	Dirección General de Concesiones de Radiodifusión / Organización
Ley del Sistema Público de Radiodifusión del Estado Mexicano / Ley	oficio 349-B-278 / Documento	Dirección General de Economía de el Espectro / Organización
Ley Federal de Derechos / Ley	oficio 4.1.2.3.1899/VUS / Documento	Dirección General de Ingeniería de el Espectro / Organización
ley federal de procedimiento administrativo / Ley	oficio CFT/D01/STP/591/12 / Documento	Directividad / Persona

Ley Federal de Radio y Televisión / Ley	oficio IFT/222/UER/DG-IEET/1724/2014 / Documento	Disposición Técnica / Otro
Ley Federal de Telecomunicaciones y Radiodifusión / Ley	oficio IFT/D02/USRT/DGAD/0492/2014 / Documento	DOF / Organización
marco antonio delgado merchan / Persona	oficio número 349-B-141 / Documento	Donde / Persona
María Elena Estavillo Flores / Persona	oficio número IFT/223/UCS/DG-CRAD/2379/2015 / Documento	ELECCIÓN DE EL SITIO Este / Otro
Mario Germán Fromow Rangel / Persona	Operación de las Estaciones de Radiodifusión Sonora / Otro	Entidades Federativas de República Mexicana / Lugar
NOM-02-SCT1-93, Especificaciones y requerimientos para la instalación y operación de estaciones de radiodifusión sonora en la banda de 88 A 108 MHz, con portadora principal modulada en frecuencia / Documento	opinión a la Secretaría de Hacienda y Crédito Público. Con oficio número IFT/222/UER/024/2016 / Documento	Ernesto Estrada González / Persona
Norma Oficial Mexicana / Organización	Opinión de la Dirección General de Aeronáutica Civil. Con oficio número 4.1.2.3.1899/VUS / Documento	Especificaciones / Organización
Nuevo Casas Grandes / Lugar	Opinión de la Secretaría de Hacienda y Crédito Público.- Con oficio número 349-B-141 / Documento	Estaciones de Radio / Organización
oficio 349-B-278 / Documento	Organización para la Cooperación y Desarrollo Económico / Organización	Estado / Organización
oficio 4.1.2.3.1899/vus / Documento	párrafo que antecede la Unidad de Espectro Radioeléctrico con fundamento en lo dispuesto en los artículos 99 y 100 de la Ley / Ley	Estatuto Orgánico / Lugar
oficio CFT/D01/STP/591/12 / Documento	párrafos vigésimo fracciones I y III; y vigésimo primero del artículo 28 de la Constitución Política de los Estados Unidos Mexicanos / Ley	Estatuto Orgánico / Organización
oficio IFT/222/UER/DG-IEET/1724/2014 / Documento	Plan Nacional de Desarrollo / Otro	Estatuto Orgánico de el Instituto Federal de Telecomunicaciones / Organización

oficio IFT/D02/USRT/DGAD/0492/2014 / Documento	Plano de Ubicación / Otro	ESTRUCTURA Para / Persona
oficio no. 349-B-278 / Documento	Pleno del Instituto / Organización	Estudio de Predicción de Áreas de Servicio / Lugar
oficio número 349-B-141 / Documento	Pleno del Instituto Federal de Telecomunicaciones / Organización	Estudio de Predicción de Áreas de Servicio / Organización
oficio número 4.1.2.3.1899/VUS / Documento	Pruebas de Comportamiento de la Estación / Otro	F.M. / Organización
oficio número IFT/222/UER/024/2016 / Documento	Radio Casas Grandes S.A. / Organización	Factor / Persona
oficio número IFT/223/UCS/DG-CRAD/2379/2015 / Documento	Reforma / Lugar	Factor Económico / Lugar
Opinión de la Dirección General de Aeronáutica Civil / Documento	Refrendo de la Concesión / Documento	Factor Técnico / Organización
Organización para la Cooperación y Desarrollo Económico / Organización	Registro Público de Concesiones / Documento	FE / Organización
párrafos vigésimo, fracciones I y III; y vigésimo primero, del artículo 28 de la Constitución Política de los Estados Unidos Mexicanos / Ley	Secretaría de Comunicaciones / Organización	FM / Organización
Plan Nacional de Desarrollo / Documento	Secretaría de Comunicaciones y Transportes / Organización	FM / Otro
Plano de Ubicación / Documento	Secretaría de Hacienda y Crédito Público / Organización	Frecuencia Modulada / Lugar
Pleno de la Suprema Cortes de Justicia de la Nación / Organización	Séptimo Transitorio del Decreto / Documento	Frecuencia Modulada / Otro
Pleno del Instituto Federal De Telecomunicaciones / Organización	Séptimo Transitorio del Decreto de Reforma / Documento	FT / Organización
Política de Ingresos No Tributarios de la Secretaría de Hacienda y Crédito Público / Documento	Sexto Transitorio del Decreto / Documento	FT+FE / Lugar
Pruebas de Comportamiento de la Estación / Otro	Solicitud de Modificación Técnica / Documento	Gobierno Federal / Organización

Radio Casas Grandes S.A. / Lugar	solicitud la documentación legal y técnica correspondiente. / Documento	Horario / Organización
Refrendo de la Concesión / Otro	Sonora / Lugar	I / Otro
Registro Público de Concesiones / Documento	Tercero y Sexto Transitorios / Otro	I de el Estatuto Orgánico / Otro
Resolutivo Primero / Documento	Unidad de Concesiones y Servicios / Organización	I. Banda / Organización
Resolutivo TERCERO / Documento	Unidad de Espectro Radioeléctrico / Organización	IFT / Organización
Secretaría de Comunicaciones / Organización	Unidad de Política de Ingresos No Tributarios / Organización	II / Otro
Secretaría de Comunicaciones y Transportes / Organización	Unidad de Política de Ingresos No Tributarios de la Secretaría de Hacienda y Crédito Público / Organización	III / Organización
Secretaría de Hacienda y Crédito Público / Organización	Unidad de Sistemas de Radio y Televisión / Organización	III / Otro
Solicitud de Modificación Técnica / Documento		III Sesión Ordinaria / Otro
Sonora / Lugar		Inclinación / Otro
Tesis de Jurisprudencia 72/2007 / Documento		Índice Nacional de Precios / Organización
título de refrendo de concesión / Documento		INEGI / Organización
Unidad de Concesiones y Servicios / Organización		Ingeniero Marco Antonio Delgado Merchan / Organización
Unidad de Espectro Radioeléctrico / Organización		INPC / Organización
Unidad de Política de Ingresos No Tributarios / Organización		Inscripción / Organización

Unidad de Política de Ingresos No Tributarios de la Secretaría de Hacienda y Crédito Público / Organización		Instituto / Organización
Unidad de Sistemas de Radio y Televisión / Organización		Instituto / Otro
Vigésimo Transitorio del presente Decreto / Ley		Instituto Federal de Telecomunicaciones / Organización
		IV / Otro
		Javier Juárez Mojica / Persona
		L.N. / Organización
		L.W. / Organización
		La / Persona
		Ley / Otro
		Ley de el Sistema Público de Radiodifusión de el Estado Mexicano / Otro
		Ley de Ingresos de la Federación para el Ejercicio Fiscal / Otro
		Ley de Ingresos para el Ejercicio Fiscal / Otro
		Ley Federal de Derechos / Otro
		Ley Federal de Radio / Otro
		Ley Federal de Telecomunicaciones / Otro
		LFRT / Organización

		M.N. / Organización
		María Elena Estavillo Flores / Persona
		Mario Germán Fromow Rangel / Persona
		MHz / Lugar
		MHz / Otro
		MHz / Persona
		MHz DE NUEVO CASAS GRANDES / Organización
		MHz Distintivo / Organización
		Monto / Lugar
		N / Otro
		Nación / Organización
		ND / Organización
		NO / Organización
		Norma Oficial Mexicana / Persona
		NOVENO / Organización
		Nuevo Casas Grandes / Lugar
		Nuevo Casas Grandes / Organización
		Observaciones / Persona

		OCDE / Organización
		OCTAVO / Organización
		Operación de las Estaciones de Radiodifusión Sonora / Organización
		Opinión de la Dirección General de Aeronáutica Civil / Organización
		Opinión de la Secretaría de Hacienda / Organización
		Organización para la Cooperación / Organización
		PCE-FM-I-II / Organización
		Periodo / Persona
		Perito / Organización
		Perito / Persona
		Plan Nacional de Desarrollo / Otro
		Plano de Ubicación / Otro
		Pleno / Organización
		Pleno de el Instituto / Organización
		Pleno de el Instituto Federal de Telecomunicaciones / Organización
		Población Servida / Lugar

		Potencia / Organización
		PRA / Organización
		Pruebas de Comportamiento de la Estación / Otro
		PU-FM / Organización
		Que / Persona
		Quinto / Otro
		Radio Casas Grandes / Organización
		Radio Casas Grandes / Persona
		Radiodifusión / Organización
		Recursos Orbitales / Organización
		Registro Público de Concesiones / Organización
		Requerimientos para la Instalación / Organización
		Resolución / Otro
		RESOLUCIÓN MEDIANTE LA CUAL EL PLENO DE EL INSTITUTO FEDERAL DE TELECOMUNICACION ES AUTORIZA LA MODIFICACIÓN A LAS CARACTERÍSTICAS TÉCNICAS DE LA ESTACIÓN DE RADIODIFUSIÓN

		SONORA CON DISTINTIVO DE LLAMADA XHTX-FM QUE OPERA LA FRECUENCIA / Organización
		Resolución / Otro
		Resolutivo / Organización
		Resolutivo Primero / Lugar
		Resolutivo PRIMERO / Otro
		Resolutivo Primero / Persona
		Resolutivo Segundo / Lugar
		Resolutivo Tercero / Lugar
		Resolutivo TERCERO / Otro
		RESOLUTIVOS PRIMERO / Persona
		S.A. / Organización
		S.C.T. / Otro
		Secretaría / Organización
		Secretaría de Comunicaciones / Organización
		Secretaría de Hacienda / Organización

		SEGUNDO / Otro
		SÉPTIMO / Otro
		Séptimo Transitorio / Persona
		Séptimo Transitorio de el Decreto / Otro
		Séptimo Transitorio de el Decreto de Reforma / Otro
		Servicios / Organización
		SEXTO / Organización
		Sexto Transitorio / Persona
		Sexto Transitorio de el Decreto / Otro
		Sexto Transitorios de la Ley Federal de Telecomunicaciones / Otro
		SHCP / Organización
		Sistema / Persona
		Solicitud / Otro
		Solicitud / Persona
		Técnicamente Factible Después de / Otro
		Telecomunicaciones / Otro
		Televisión / Organización

		Tercero / Otro
		Tercero / Persona
		Transportes / Organización
		Ubicación / Otro
		UBICACIÓN CERCANA A OTRAS ESTACIONES Cuando / Otro
		UBICACIÓN DE EL SISTEMA RADIADOR El / Otro
		UBICACIÓN IDEAL Es / Persona
		UER / Organización
		Unidad de Concesiones / Organización
		Unidad de Espectro Radioeléctrico / Organización
		Unidad de Política de Ingresos No Tributarios de la Secretaría de Hacienda / Organización
		Unidad de Política de Ingresos No Tributarios de la SHCP / Organización
		Unidad de Sistemas de Radio / Organización
		V / Otro
		V. Referencias / Otro

		Valor / Organización
		VI / Organización
		Vigencia / Organización
		Vigésimo Transitorio / Otro
		VII / Organización
		VII de el Estatuto Orgánico de el Instituto / Otro
		VIII / Organización
		VR / Organización
		X / Otro
		X de la Ley Federal de Procedimiento Administrativo / Organización
		XETX-AM / Otro
		XHNVG-FM de Nuevo Casas Grandes / Organización
		XHTX-FM / Organización
		XHTX-FM / Otro
		XHTX-FM de Nuevo Casas Grandes / Organización
		XHTX-FM Población / Otro
		XI de el Estatuto Orgánico de el Instituto Federal de

		Telecomunicaciones / Otro
		XIV / Otro
		XIV de el Estatuto Orgánico de el Instituto Federal de Telecomunicaciones / Otro
		XXVIII de la Ley Federal de Telecomunicaciones / Otro

Documento 2018

Etiquetado Manual Entidad/clasificación	Software NERC Entidad/clasificación	FreeLing 4.0 Entidad/clasificación
7, primer párrafo de la LFTR / Ley	Instituto Federal de Telecomunicaciones / Organización	Resolución Radiomóvil Dipsa / Persona
10 fracción II, inciso a) de la Ley Orgánica del Poder Judicial de la Federación / Ley	16 fracción X 32 35 fracción I 36 38 39 45 y 57 de la Ley Federal de Procedimiento Administrativo / Ley	A efecto de / Persona
16 fracción X, 32, 35, fracción I, 36, 38, 39, 45 y 57 de la Ley Federal de Procedimiento Administrativo / Ley	6 fracciones IV y VII 15 fracción X 17 fracción I 125 128 y 129 fracciones VII VIII y IX 176 177 fracción VII 312 y 313 de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	A EL / Persona
4 fracción I y 6, fracción XXXVII, del Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Ley	81 fracción I inciso e y 84 de la Ley de Amparo y 10 fracción II inciso a de la Ley Orgánica del Poder Judicial de la Federación / Ley	Acuerdo / Lugar
6, fracciones IV y VII, 15, fracción X, 17, fracción I, 125, 128 y 129, fracciones VII, VIII y IX, 176, 177 fracción VII, 312 y 313 de la Ley Federal de Telecomunicaciones y	4 fracción I y 6 fracción XXXVII del Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Ley	Acuerdo / Organización

Radiodifusión / Ley		
72, 73, 74, 197 y 202 del Código Federal de Procedimientos Civiles / Ley	acuerdo de 21 de octubre de 2016, la Segunda Sala de la Suprema Corte de Justicia de la Nación (en lo sucesivo, "SCJN") asumió la competencia para conocer el recurso de revisión, lo registro con número 1028/2016 / Documento	Acuerdo / Otro
81, fracción I inciso e), y 84 dela Ley de Amparo / Ley	Acuerdo de Tarifas / Documento	Acuerdo / Persona
acuerdo de 21 de octubre de 2016, la Segunda Sala de la Suprema Corte de Justicia de la Nación (en lo sucesivo, "SCJN") asumió la competencia para conocer el recurso de revisión, lo registro con número 1028/2016 / Documento	Acuerdo del Sistema / Documento	Acuerdo de el Sistema / Persona
acuerdo de tarifas 2015 / Documento	ACUERDO mediante el cual el Pleno de el Instituto Federal de Telecomunicaciones determina las tarifas de interconexión resultado de la metodología para el cálculo de costos de interconexión que se utilizará para resolver los desacuerdos de interconexión que se presenten respecto de las condiciones aplicables al año 2015 aprobado mediante Acuerdo P/IFT/EXT/191214/284 / Documento	Acuerdo de Tarifas / Organización
ACUERDO mediante el cual el Pleno del Instituto Federal de Telecomunicaciones determina las tarifas de interconexión resultado de la metodología para el cálculo de costos de interconexión que se utilizará para resolver los desacuerdos de interconexión que se presenten respecto de las condiciones aplicables al año 2015", aprobado mediante Acuerdo P/IFT/EXT/191214/284 / Documento	ACUERDO mediante el cual el Pleno de el Instituto Federal de Telecomunicaciones establece el Sistema Electrónico de Solicitudes de Interconexión / Documento	Acuerdo de Tarifas / Otro

ACUERDO mediante el cual el Pleno del Instituto Federal de Telecomunicaciones emite la metodología para el cálculo de costos de interconexión de conformidad con la Ley Federal de Telecomunicaciones y Radiodifusión / Documento	Acuerdo P/IFT/120815/369 / Documento	Adolfo Cuevas Teja / Persona
ACUERDO mediante el cual el Pleno del Instituto Federal de Telecomunicaciones establece el Sistema Electrónico de Solicitudes de Interconexión / Documento	Acuerdo P/IFT/310118/30 / Documento	Adriana Sofía Labardini Inzunza / Persona
Acuerdo P/IFT/120815/369 / Documento	Acuerdo P/IFT/EXT//161214/277 / Documento	Agente de el Ministerio Público de la Federación / Organización
Acuerdo P/IFT/310118/30 / Documento	Acuerdo P/IFT/EXT/191214/284 / Documento	Antecedentes / Organización
Acuerdo P/IFT/EXT//161214/277 / Documento	Adolfo Cuevas Teja / Otro	ANTECEDENTES Radiomóvil Dipsa / Organización
Acuerdo P/IFT/EXT/191214/284 / Documento	Adriana Sofía Labardini Inzunza / Otro	APLICABLES DE EL / Persona
Adolfo Cuevas Teja / Persona	Agente del Ministerio Público de la Federación / Otro	Arturo Robles Rovalo / Persona
Adriana Sofía Labardini Inzunza / Persona	artículo 129 de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	B / Otro
Agente del Ministerio Público / Persona	artículo 129 fracción VIII de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Ciudad de México / Lugar
agente del ministerio público de la federación / Persona	artículo 17 de la Ley de Amparo / Ley	Código Federal de Procedimientos Civiles / Otro
amparo en revisión 329/2016 / Documento	artículo 28 párrafos décimo quinto y décimo sexto de la Constitución Política de los Estados Unidos Mexicanos / Ley	Comisionado Adolfo Cuevas Teja / Organización
amparo en revisión R.A. 105/2016 / Documento	artículo 6° fracción I del Estatuto / Ley	Comisionados Gabriel Oswaldo Contreras Saldívar / Organización

artículo 129 de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	artículo vigésimo transitorio segundo párrafo de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Como / Persona
artículo 129 fracción VIII de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	artículos 1 7 8 y 12 del Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Ley	Competencia de el Instituto / Organización
artículo 131 de la propia Ley Federal de Telecomunicaciones y Radiodifusión / Ley	artículos 103 y 107 de la Constitución Política de los Estados Unidos Mexicanos / Ley	Competencia Económica / Lugar
artículo 137 de la LFTR / Ley	artículos 107 fracción VIII inciso a de la Constitución General de la República / Ley	Consideraciones / Organización
artículo 17 de la Ley de Amparo / Ley	artículos 128 176 y 177 fracción VII de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	CONSIDERANDO PRIMERO / Organización
artículo 28 de la Constitución Política de los Estados Unidos Mexicanos / Ley	artículos 312 y 313 de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Constitución / Otro
artículo 28, párrafos décimo quinto y décimo sexto de la Constitución Política de los Estados Unidos Mexicanos / Ley	artículos 6° apartado B fracción II 28 párrafos décimo quinto y décimo sexto de la Constitución / Ley	Constitución General de la República / Organización
artículo 6°, fracción I, del Estatuto / Ley	artículos 7 15 fracción X 17 fracción I y 129 de la LFTR el Pleno del Instituto / Ley	Constitución Política de los Estados Unidos Mexicanos / Organización
artículo primero transitorio / Ley	artículos 7 16 y 45 de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Constitución Política de los Estados Unidos Mexicanos / Otro
artículo vigésimo transitorio, segundo párrafo de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	artículos indicados el Pleno del Instituto / Ley	Convenio / Otro
artículos 1, 7, 8 y 12 del Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Ley	Arturo Robles Rovalo / Otro	Convergencia / Organización

artículos 103 y 107 de la Constitución Política de los Estados Unidos Mexicanos / Ley	Ciudad de México / Otro	Convergencia / Persona
artículos 107, fracción VIII, inciso a), de la Constitución General de la República / Ley	Código Federal de Procedimientos Civiles / Documento	Convergencia de México / Organización
artículos 128 y 177 fracción VII de la LFTR / Ley	Competencia del Instituto / Organización	Costo Incremental de Largo Plazo Puro / Lugar
artículos 128, 176 y 177, fracción VII de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Constitución General de la República / Documento	Cuarto / Organización
artículos 129 fracción IX, 176, 177, fracción XV y 178 de la LFTR / Ley	Constitución Política de los Estados Unidos Mexicanos / Documento	CUARTO / Organización
artículos 312 y 313 de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Convergencia de México / Otro	De / Persona
artículos 6°, apartado B fracción II, 28, párrafos décimo quinto y décimo sexto de la Constitución / Ley	Convergencia de México S. A. de C. V. / Persona	Debiendo / Persona
artículos 7, 15, fracción X, 17, fracción I, y 129 de la LFTR / Ley	Convergencia de México S.A. de C.V. / Organización	Decreto / Organización
artículos 7, 16 y 45 de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Costo Incremental de Largo Plazo Puro / Otro	Diario Oficial de la Federación / Organización
Arturo Robles Rovalo / Persona	Diario Oficial de la Federación / Documento	Distrito Federal / Lugar
Ciudad de México / Lugar	Emisión del Acuerdo / Documento	DOF / Organización
Código Federal de Procedimientos Civiles / Documento	Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Documento	Emisión de el Acuerdo / Otro
constitución general de la república / Documento	expediente 1649/2015 / Documento	EMITIDA MEDIANTE ACUERDO / Otro
constitución política de los estados unidos mexicanos / Documento	Gabriel Oswaldo Contreras Saldivar / Otro	EMITIDA POR SEGUNDO TRIBUNAL COLEGIADO EN MATERIA ADMINISTRATIVA ESPECIALIZADO EN

		COMPETENCIA ECONÓMICA / Organización
contradicción de tesis 268/2010 / Documento	Javier Juárez Mojica / Otro	En / Persona
Convergencia de México S. A. de C. V. / Organización	Juárez / Lugar	En consecuencia / Organización
Costo Incremental de Largo Plazo Puro / Otro	Juez Segundo de Distrito / Otro	EN CUMPLIMIENTO A LA EJECUTORIA DE FECHA / Persona
Diario Oficial de la Federación / Documento	Justicia de la Nación / Otro	Estado / Organización
Estatuto Orgánico del Instituto Federal de Telecomunicaciones / Documento	Juzgado Segundo de Distrito / Organización	Estatuto / Otro
expediente 1649/2015 / Documento	Juzgados de Distrito Especializados / Otro	Estatuto Orgánico de el Instituto Federal de Telecomunicaciones / Organización
Gabriel Oswaldo Contreras Saldívar / Persona	Ley de Amparo / Ley	Finalmente / Persona
Instituto Federal de Telecomunicaciones / Organización	Ley del Sistema Público de Radiodifusión del Estado Mexicano / Ley	I / Otro
Javier Juárez Mojica / Persona	Ley Federal de Procedimiento Administrativo / Ley	II / Otro
Juez Segundo de Distrito / Persona	Ley Federal de Telecomunicaciones y Radiodifusión / Ley	III / Organización
Juzgado Segundo de Distrito / Organización	Ley Orgánica del Poder Judicial de la Federación / Ley	III / Otro
juzgados de distrito especializados / Organización	María Elena Estavillo Flores / Otro	III Sesión Ordinaria / Otro
Ley de Amparo / Ley	Mario Germán Fromow Rangel / Otro	Instituto / Organización
Ley del Sistema Público de Radiodifusión del Estado Mexicano / Ley	Materia Administrativa Especializado / Otro	Instituto Federal de Telecomunicaciones / Organización

Ley Federal de Procedimiento Administrativo / Ley	Materia Administrativa Especializados / Otro	IV / Organización
Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Materia de Competencia Económica / Organización	IX / Otro
Ley Orgánica del Poder Judicial de la Federación / Ley	metodología para el cálculo de costos de interconexión de conformidad con la Ley Federal de Telecomunicaciones y Radiodifusión aprobado por el Pleno de el Instituto mediante Acuerdo P/IFT/EXT//161214/277 / Ley	Javier Juárez Mojica / Persona
María Elena Estavillo Flores / Persona	México / Lugar	Juez Segundo de Distrito / Organización
Mario Germán Fromow Rangel / Persona	Modelos de Costos / Otro	Jurisdicción / Otro
Materia Administrativa Especializado en Competencia Económica, Radiodifusión y Telecomunicaciones / Otro	Oficialía de Partes / Organización	Justicia de la Nación / Otro
metodología para el cálculo de costos de interconexión de conformidad con la Ley Federal de Telecomunicaciones y Radiodifusión”, aprobado por el Pleno del Instituto mediante Acuerdo P/IFT/EXT//161214/277 / Otro	Oficina de Correspondencia Común de los Juzgados de Distrito y Tribunales Colegiados / Lugar	Justicia de la Unión AMPARA Y PROTEGE / Organización
Modelos de Costos / Otro	párrafo del artículo 129 de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Juzgado Segundo de Distrito / Organización
Oficialía de Partes / Organización	párrafo del artículo vigésimo transitorio de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Juzgados de Distrito Especializados / Organización
Oficina de Correspondencia Común / Organización	párrafos vigésimo fracciones I y III; y vigésimo primero del artículo 28 de la Constitución Política de los Estados Unidos Mexicanos / Ley	Ley de Amparo / Otro
Oficina de Correspondencia Común de los Juzgados de Distrito y Tribunales Colegiados / Organización	Pleno del Instituto / Organización	Ley de el Sistema Público de Radiodifusión de el Estado Mexicano / Otro

Oficina de Correspondencia Común de los Juzgados de Distrito y Tribunales Colegiados en Materia Administrativa Especializados en Competencia Económica, Radiodifusión y Telecomunicaciones / Organización	Pleno del Instituto Federal de Telecomunicaciones / Organización	Ley Federal de Procedimiento Administrativo / Otro
penúltimo párrafo del artículo 129 de la Ley Federal de Telecomunicaciones y Radiodifusión / Ley	Publicación de Tarifas de Interconexión / Documento	Ley Federal de Telecomunicaciones / Otro
Pleno del Instituto Federal de Telecomunicaciones / Organización	Radiomóvil Dipsa S.A. de C.V. / Organización	Ley Orgánica de el Poder Judicial de la Federación / Otro
Publicación de Tarifas de Interconexión del año 2015 / Otro	Registro Público de Concesiones del Instituto / Documento	LFTR / Organización
R.A. 105/2016 / Documento	Registro Público de Concesiones del Instituto Federal de Telecomunicaciones / Documento	Lo / Persona
Radiomóvil Dipsa S.A. de C.V. / Organización	Registro Público de Telecomunicaciones / Documento	M.N. / Organización
Registro Público de Concesiones / Documento	Resolución Radiomóvil Dipsa S.A. de C.V. / Documento	Magistrado / Persona
registro público de concesiones del instituto federal de telecomunicaciones / Documento	Segunda Sala de la Suprema / Organización	María Elena Estavillo Flores / Persona
Registro Público de Telecomunicaciones / Documento	Segundo Tribunal Colegiado / Organización	Mario Germán Fromow Rangel / Persona
RESOLUCIÓN MEDIANTE LA CUAL EL PLENO DEL INSTITUTO FEDERAL DE TELECOMUNICACIONES DETERMINA LAS CONDICIONES DE INTERCONEXIÓN NO CONVENIDAS ENTRE CONVERGIA DE MÉXICO, S.A. DE C.V. Y RADIOMÓVIL DIPSA, S.A. DE C.V., APLICABLES DEL 1 DE ENERO DE 2015 AL 31 DE DICIEMBRE DE 2016 / Documento	Segundo Tribunal Colegiado de Circuito / Organización	Materia Administrativa / Lugar
Resolución mediante la cual el Pleno del Instituto Federal de Telecomunicaciones determina las condiciones de interconexión no convenidas entre Convergencia de México, S.A. de C.V. y Radiomóvil Dipsa, S.A. de C.V., aplicables del 1 de enero de 2015 al 31	Servicios de Interconexión / Otro	Materia Administrativa Especializado / Lugar

de diciembre de 2016”, contenida en el Acuerdo P/IFT/120815/369, de doce de agosto de dos mil quince / Documento		
RESOLUCIÓN MEDIANTE LA CUAL EL PLENO DEL INSTITUTO FEDERAL DE TELECOMUNICACIONES MODIFICA LA “RESOLUCIÓN MEDIANTE LA CUAL EL PLENO DEL INSTITUTO FEDERAL DE TELECOMUNICACIONES DETERMINA LAS CONDICIONES DE INTERCONEXIÓN NO CONVENIDAS ENTRE CONVERGIA DE MÉXICO, S.A. DE C.V. Y RADIOMÓVIL DIPSA, S.A. DE C.V., APLICABLES DEL 1 DE ENERO DE 2015 AL 31 DE DICIEMBRE DE 2016” EMITIDA MEDIANTE ACUERDO P/IFT/120815/369, EN CUMPLIMIENTO A LA EJECUTORIA DE FECHA 14 DE DICIEMBRE DE 2017 EMITIDA POR SEGUNDO TRIBUNAL COLEGIADO EN MATERIA ADMINISTRATIVA ESPECIALIZADO EN COMPETENCIA ECONÓMICA, RADIODIFUSIÓN Y TELECOMUNICACIONES CON RESIDENCIA EN LA CIUDAD DE MÉXICO Y JURISDICCIÓN EN TODA LA REPÚBLICA CORRESPONDIENTE AL AMPARO EN REVISIÓN 105/2016 / Documento	Sistema Electrónico de Solicitudes de Interconexión / Organización	Materia Administrativa Especializado / Persona
Segunda Sala de la Suprema Corte de Justicia de la Nación / Organización	Solicitud de resolución de condiciones de interconexión no convenidas. El 5 de mayo de 2015 el representante legal de Telcel presentó ante el Instituto escrito mediante el cual solicitó su intervención para resolver los términos tarifas y condiciones que no pudo convenir con Convergía para la interconexión de sus respectivas redes públicas de telecomunicaciones que aplicarán para los ejercicios de el 1 de enero al 31 de diciembre de el 2015 y de el 1 de enero al 31 de diciembre de el 2016 / Documento	Materia Administrativa Especializados / Organización
SEGUNDO TRIBUNAL COLEGIADO / Organización	Suprema Corte de Justicia / Otro	Máximo Tribunal / Organización

Segundo Tribunal Colegiado de Circuito / Organización	Suprema Corte de Justicia de la Nación / Otro	Metodología / Organización
Sistema Electrónico de Solicitudes de Interconexión / Organización	Tribunal Colegiado de Circuito / Organización	Metodología / Persona
Solicitud de Resolución / Documento	Vigésimo Transitorio del Decreto / Documento	Metodología de Costos / Organización
Solicitud de resolución de condiciones de interconexión no convenidas. El 5 de mayo de 2015, el representante legal de Telcel presentó ante el Instituto, escrito mediante el cual solicitó su intervención para resolver los términos, tarifas y condiciones que no pudo convenir con Convergía para la interconexión de sus respectivas redes públicas de telecomunicaciones que aplicarán para los ejercicios del 1 de enero al 31 de diciembre del 2015 y del 1 de enero al 31 de diciembre del 2016 / Documento		Metodología de Costos / Otro
Suprema Corte de Justicia de la Nación / Organización		Metodología de Costos / Persona
Tribunal Colegiado / Organización		Modelos de Costos / Lugar
Vigésimo Transitorio del Decreto / Ley		Modelos de Costos / Otro
Vigésimo Transitorio del Decreto por el que se expiden la Ley Federal de Telecomunicaciones y Radiodifusión, y la Ley del Sistema Público de Radiodifusión del Estado Mexicano; y se reforman, adicionan y derogan diversas disposiciones en materia de telecomunicaciones y radiodifusión / Ley		Oficialía de Partes / Otro
XVI Sesión Ordinaria / Otro		Oficina de Correspondencia Común de los Juzgados de Distrito / Lugar
		Órgano Jurisdiccional / Otro
		Pleno / Organización
		Pleno de el Instituto / Organización

		Pleno de el Instituto Federal de Telecomunicaciones / Organización
		Por / Organización
		Publicación de Tarifas de Interconexión / Otro
		QUINTO / Organización
		R.A. / Otro
		Radiodifusión / Lugar
		Radiodifusión / Organización
		RADIODIFUSIÓN Y TELECOMUNICACIONES CON RESIDENCIA EN LA CIUDAD DE MÉXICO Y JURISDICCIÓN EN TODA LA REPÚBLICA CORRESPONDIENTE A EL AMPARO EN REVISIÓN / Otro
		Radiomóvil Dipsa / Organización
		Radiomóvil Dipsa / Persona
		Registro Público de Concesiones de el Instituto / Organización
		Registro Público de Concesiones de el Instituto Federal de Telecomunicaciones / Organización
		Registro Público de Telecomunicaciones / Organización

		Reglamentaria / Organización
		República / Organización
		Resolución / Otro
		RESOLUCIÓN MEDIANTE LA CUAL EL PLENO DE EL INSTITUTO FEDERAL DE TELECOMUNICACIONES DETERMINA LAS CONDICIONES DE INTERCONEXIÓN NO CONVENIDAS ENTRE CONVERGIA DE MÉXICO / Organización
		RESOLUCIÓN MEDIANTE LA CUAL EL PLENO DE EL INSTITUTO FEDERAL DE TELECOMUNICACIONES MODIFICA LA / Organización
		Resolutivo SEGUNDO / Organización
		Resolutivo TERCERO / Otro
		Resolutivo Tercero / Persona
		RESOLUTIVOS PRIMERO / Persona
		Resolutivos Segundo / Organización
		S. A. de C. V. / Persona
		S.A. de C.V. / Persona

		S.A. DE C.V. / Persona
		S.A. DE C.V. Y RADIOMÓVIL DIPSA / Persona
		SCJN / Organización
		Segunda Sala / Otro
		Segunda Sala / Persona
		Segunda Sala de la SCJ / Otro
		Segunda Sala de la SCJN / Otro
		Segunda Sala de la Suprema / Organización
		Segunda Sala de la Suprema Corte de Justicia de la Nación / Organización
		Segunda Sala de la Suprema Corte de Justicia de la Nación / Otro
		SEGUNDO / Otro
		Segundo Tribunal Colegiado / Organización
		Segundo Tribunal Colegiado de Circuito / Organización
		Servicio Local / Organización
		Servicios de Interconexión / Organización
		SESI / Organización

		SEXTO / Organización
		Sistema Electrónico de Solicitudes de Interconexión / Otro
		Solicitud de Resolución / Organización
		Solicitud de Resolución / Otro
		Suprema Corte / Organización
		Suprema Corte de Justicia de la Nación / Otro
		Telcel / Organización
		Telecomunicaciones / Organización
		Telecomunicaciones / Otro
		TERCERO / Organización
		Tribunal Colegiado / Organización
		Tribunal Colegiado de Circuito / Organización
		Tribunal Pleno / Organización
		Tribunales Colegiados / Organización
		Tribunales Especializados / Organización
		Vigésimo Transitorio de el Decreto / Organización

		VII / Organización
		VII / Otro
		VII de la Ley Federal de Telecomunicaciones / Otro
		VII de la LFTR / Organización
		VIII / Organización
		VIII de la Ley Federal de Telecomunicaciones / Otro
		X / Otro
		XV / Organización
		XVI Sesión Ordinaria / Organización
		XXXVII / Organización

Glosario de términos

Entidad: Se considera entidad a personas, lugares y organizaciones (Galicia-Haro, S. N., Gelbukh, A., & Bolshakov, I. A 2004), aunque la definición de entidad, como hemos dicho anteriormente, puede adaptarse dependiendo del documento y de la información específica que buscamos extraer.

Patrón: Elementos que influyen en una decisión, y que definen el acto siguiente a realizar, o la conclusión a obtener. Por ejemplo, dado un contexto donde tenemos imágenes con personas, podemos reconocer los patrones (elementos) que definen una cara, y así poder identificarla. Así mismo, existen patrones en un documento legal que nos permiten diferenciarlo de un texto común, un cuento, una novela, o un poema.

Lenguaje Natural: Capacidad que tenemos los humanos para comunicarnos por medio de lenguas. Es decir, es el lenguaje que practicamos día a día al intercambiar ideas con otros humanos.

Lenguaje Formal: Modelo que describe un conjunto de cadenas finito o infinito, y sirve como acercamiento para entender o generar algunos aspectos del lenguaje natural.

Lenguaje Legal: Manifestación o variante del lenguaje natural que se utiliza específicamente en el contexto legal, en donde se manejan particulares normas lingüísticas como la sintaxis, vocabulario, términos y significados (Mattila, H. E. 2016).

Evento: Tupla compuesta por dos elementos en una misma oración: Fecha y verbo, del que se puede obtener más información implementando patrones.

Reconocimiento de Patrones: Acto de reconocer los elementos que nos ayudan a decidir dado un contexto.

Procesamiento de Lenguaje Natural: Conjunto de técnicas y procesos computacionales que trabajan sobre textos escritos en lenguaje natural para, con base en análisis lingüísticos, resolver tareas útiles que requieren de dicho conocimiento lingüístico (Liddy 2001).

Corpus lingüístico: “Un corpus lingüístico consiste en un conjunto de textos de materiales escritos y/o hablados, debidamente recopilados para realizar ciertos análisis lingüísticos” (Sierra 2010).

Reconocimiento de Entidades Nombradas: Extracción de un conjunto de cadenas específico de un texto escrito en lenguaje natural (no estructurado), donde lo obtenido corresponde con entidades de interés, y las cuales se hacen accesibles en una estructura computable.