

# Centro Nacional de Investigación y Desarrollo Tecnológico

Subdirección Académica  
Departamento de Ciencias Computacionales

## TESIS DE MAESTRÍA EN CIENCIAS

Un Método para la Identificación Automática de Argumentos en  
Discursos Políticos Escritos en Español

presentada por  
**Ing. Felipe de Jesús Ojeda Cruz**

como requisito para la obtención del grado de  
**Maestro en Ciencias de la Computación**

Director de tesis  
**Dr. Noé Alejandro Castro Sánchez**

Codirector de tesis  
**Dr. Héctor Jiménez Salazar**



Cuernavaca, Morelos a 11 de noviembre del 2016  
OFICIO No. DCC/225/2016  
Asunto: Aceptación de documento de tesis

C. DR. GERARDO V. GUERRERO RAMÍREZ  
SUBDIRECTOR ACADÉMICO  
PRESENTE

Por este conducto, los integrantes de Comité Tutorial del Ing. Felipe de Jesús Ojeda Cruz, con número de control M14CE064, de la Maestría en Ciencias de la Computación, le informamos que hemos revisado el trabajo de tesis profesional titulado "Un método para la identificación automática de argumentos en discursos políticos escritos en español" y hemos encontrado que se han realizado todas las correcciones y observaciones que se le indicaron, por lo que hemos acordado aceptar el documento de tesis y le solicitamos la autorización de impresión definitiva.

DIRECTOR DE TESIS

Dr. Noé Alejandro Castro  
Sánchez  
Doctor en Ciencias de la  
Computación  
08701806

CO-DIRECTOR DE TESIS

Dr. Héctor Jiménez Salazar  
Doctor en Ciencias en  
Ingeniería Eléctrica  
DIR/1603/2016

REVISOR 1

Dr. Manuel Mejía Lavalle  
Doctor en Ciencias  
Computacionales.  
8342472

REVISOR 2

Dra. Alicia Martínez  
Rebollar  
Doctora en Informática  
7399055

C.p. M.T.I. María Elena Gómez Torres - Jefa del Departamento de Servicios Escolares.  
Estudiante  
Expediente

NACS/lmz



SEP

SECRETARÍA DE  
EDUCACIÓN PÚBLICA



TECNOLÓGICO NACIONAL DE MÉXICO  
Centro Nacional de Investigación y Desarrollo Tecnológico

Cuernavaca, Mor., 17 de noviembre de 2016  
OFICIO No. SAC/329/2016

**Asunto:** Autorización de impresión de tesis

**ING. FELIPE DE JESÚS OJEDA CRUZ**  
**CANDIDATO AL GRADO DE MAESTRO EN CIENCIAS**  
**DE LA COMPUTACIÓN**  
**PRESENTE**

Por este conducto, tengo el agrado de comunicarle que el Comité Tutorial asignado a su trabajo de tesis titulado **"Un método para la identificación automática de argumentos en discursos políticos escritos en español"**, ha informado a esta Subdirección Académica, que están de acuerdo con el trabajo presentado. Por lo anterior, se le autoriza a que proceda con la impresión definitiva de su trabajo de tesis.

Esperando que el logro del mismo sea acorde con sus aspiraciones profesionales, reciba un cordial saludo.

**ATENTAMENTE**

"CONOCIMIENTO Y TECNOLOGÍA AL SERVICIO DE MÉXICO"

**DR. GERARDO VICENTE GUERRERO RAMÍREZ**  
**SUBDIRECTOR ACADÉMICO**



S. E. P. - TecNM  
CENTRO NACIONAL  
DE INVESTIGACIÓN  
Y DESARROLLO  
TECNOLÓGICO  
DIRECCIÓN

C.p. M.T.I. María Elena Gómez Torres - Jefa del Departamento de Servicios Escolares.  
Expediente

GVGR/mcr



# Dedicatoria

*“Raro y celestial don, el que sepa sentir y razonar al mismo tiempo”*

VITTORIO ALFIERI (1749-1803) DRAMATURGO Y POETA ITALIANO.

Con mucho cariño dedico este trabajo totalmente:

A mis **padres** por el apoyo incondicional, absoluto, total, ilimitado y leal que me han brindado en toda mi vida, gracias a ellos por permitirme superar y lograr esta hazaña.

# Agradecimientos

Son muchas personas a las que quiero agradecer, porque de una u otra forma estuvieron completamente involucradas en el desarrollo de mi maestría.

Quiero empezar por las personas más importantes de mi vida: “mis padres”, ustedes me apoyaron sin titubear y sin pensar, me ayudaron económicamente, me apoyaron incondicionalmente en todas las decisiones que tomaba. Sin el apoyo de ustedes todo esto que hice fue posible gracias a ustedes. Se los agradezco profundamente.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por su Programa Nacional de Posgrados de Calidad (PNPC) por medio del cual me apoyó económicamente como estudiante de tiempo completo.

A la red Temática en Tecnologías del Lenguaje por medio de la cual conocí a otros investigadores y estudiantes del área.

Al Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET), por darme la oportunidad de cursar la maestría.

De igual forma, agradezco a mi Mami, gracias por echarme la mano en momentos difíciles.

A mis hermanos Antonio y Daniel, sus porras fueron lo más importante para mí 😊.

Muchas gracias a mis revisores la Dra. Alicia, el Dr. Lavalle y el Mtro. Alcántara; por su dedicación y tolerancia al revisar mis trabajos. Así como proporcióname buenas ideas para tener un mejor trabajo de tesis.

Gracias a mis asesores, el Dr. Noé Castro y el Dr. Héctor Jiménez, por su paciencia, por estar atentos a mí, y hacer este trabajo posible. Les agradezco sus conocimientos que me proporcionaron, así también, muy lindas experiencias que tuve en diferentes eventos.

Quiero también mencionar a una personita, muy importante en este gran camino que recorrí, de hecho, no lo recorrí solo. Raquel, tu estuviste a lado de mí motivándome, apoyándome, ayudándome a tomar decisiones. Ahora te has vuelto un ser muy importante para mí. Le agradezco a la vida por toparme en tu camino.

Por último, pero no por menos, gracias a todos mis compañeros (Obed, Misael, Isai, Vicky, Nandí, Irving, Roberto, entre otros más), por sus consejos, y estar juntos para salir adelante. Gracias a ustedes la maestría fue mucho más sencilla llevarla a cabo y concluirla.

## Resumen

Con gran frecuencia usamos y escuchamos argumentos, ya que intentamos convencer y persuadir a nuestro receptor, por este motivo, recientemente ha habido un creciente interés en el tratamiento de los argumentos. Una de las principales aplicaciones en la identificación de argumentos es utilizar esta identificación para recuperar información, por ejemplo, en generación de preguntas y respuestas.

En este contexto, los argumentos se encuentran en todos los textos que se escriben con un determinado propósito comunicativo, por ejemplo, los discursos. Existen una gran cantidad de tipos de discursos con diferentes enfoques, sin embargo, los discursos políticos son los que normalmente escuchamos o leemos.

Por esta razón, en este trabajo de investigación, nos centramos en el desarrollo de un método para identificar argumentos que hacen uso de ciertas palabras claves que permiten detectar argumentos en los textos escritos, mediante la implementación de un algoritmo desarrollado con expresiones regulares, que cumplen ciertas pautas y reglas para identificar argumentos.

Los trabajos recientes en el tema de la argumentación utilizan muchas características lingüísticas para identificar argumentos. En este trabajo las características que se utilizan para identificar la estructura de los argumentos son los signos de puntuación, el contexto, las conjunciones, la posición de las proposiciones, la característica sintáctica y los marcadores argumentativos; éstos últimos son de gran importancia ya que son palabras claves que permiten identificar argumentos.

En esta tesis los marcadores se utilizan como característica principal, por ello, se desarrolló un método que permite encontrar nuevos marcadores tomando como base los marcadores ya identificados previamente.

Para la ejecución de las pruebas se tuvo que realizar un corpus manual de argumentos en español extraídos de discursos políticos. Se realizaron 3 experimentos teniendo como resultado de precisión 56%, 90% y 90% para los experimentos 1, 2 y 3 respectivamente. La diferencia de los últimos dos experimentos estriba en la cobertura, con un 79% para el experimento 2 y un 98% para el experimento 3.

## **Abstract**

Very often we use and hear arguments as they try to convince and persuade our receiver, for this reason, there has recently been a growing interest in the treatment of arguments. One of the main applications in the identification of arguments is to use this identification to retrieve information, for example, generation of questions and answers.

In this context, the arguments are found in all the texts that are written with a specific communicative purpose, for example, the speeches. There are a lot of types of discourse with different approaches, however, political speeches are normally hear or read.

For this reason, in this research, we focus on the development of a method to identify arguments that make use of certain keywords that detect arguments in written texts, by implementing an algorithm developed with regular expressions, which meet certain guidelines and rules to identify arguments.

Re-cent work on the subject of the argument used many linguistic features to identify arguments. In this work the features that are used to identify the structure of the arguments are the signs of punctuation, context, conjunctions, position of the propositions, syntactic feature and argumentative markers; These are very important since they are keywords that identify arguments.

In this thesis, markers are used as main characteristic, therefore, developed a method that allows to find new markers taking as bases markers already identified previously.

For the execution of the tests had to make a small manual Spanish corpus of arguments extracted political speeches. 3 experiments having as a result of precision 56%, 90% and 90% for experiments 1, 2, and 3 respectively were conducted. The difference between the last two experiments is in coverage, with a 79% for experiment 2 and 98% for experiment 3.

# Contenido

Lista de Figuras.....	IV
Lista de Tablas.....	V
Lista de Ecuaciones.....	VI
Capítulo 1 .....	1
1.1. Introducción.....	2
1.2. Planteamiento del problema .....	3
1.3. Objetivo .....	4
1.4. Estructura del documento .....	4
Capítulo 2 .....	6
2.1. Argumentación .....	7
2.2. Lógica.....	8
2.2.1. Lógica formal.....	8
2.2.2. Lógica informal .....	9
2.3. Discurso .....	9
2.4. Marcadores del discurso.....	10
2.5. Procesamiento de Lenguaje Natural.....	11
2.5.1. Los niveles de estudio del PLN.....	11
2.5.2. Etiquetado POS.....	12
2.6. <i>Freeling</i> .....	12
2.7. Expresiones regulares.....	14
2.8. Medidas de evaluación .....	14
2.9. Marco de trabajo.....	16
2.10. Discusión .....	17



Capítulo 3 .....	18
3.1. Identificación de argumentos .....	19
3.1.1. La identificación de estructuras en el discurso argumentativo de ensayos persuasivos.....	19
3.1.2. Extracción de argumentos y palabras de dominio para la identificación de los componentes argumentativos en textos .....	20
3.1.3. Extracción de argumentos de noticias .....	21
3.1.4. Hacia el reconocimiento basado en segmentos de la estructura de la argumentación en textos cortos .....	23
3.2. Tabla comparativa de los trabajos relacionados .....	27
3.3. Discusión de los trabajos relacionados.....	29
3.4. Discusión del capítulo del estado del arte.....	29
Capítulo 4 .....	30
4.1. Presentación.....	31
4.2. Metodología tradicional con respecto a la metodología propuesta en este trabajo..	33
4.3. Inducción de patrones .....	34
4.3.1. Incremento de marcadores .....	34
4.3.2. Método de inducción de patrones.....	36
4.3.3. Método de extracción automática de patrones contextuales mediante el análisis morfosintáctico.....	37
4.3.4. Proceso para la búsqueda extendida con expresiones regulares de patrones morfosintácticos .....	46
4.4. Proceso para la identificación de argumentos.....	47
4.4.1. Desarrollo de un modelo con expresión regular.....	49
4.4.2. Análisis del modelo de la expresión regular .....	53
4.4.3. Recursividad de la expresión regular.....	54
4.5. Discusión.....	56

Capítulo 5 .....	58
5.1. Rendimiento de las expresiones regulares .....	59
5.2. Corpus etiquetado manualmente .....	59
5.3. Ejecución de experimentos.....	62
5.3.1. Ejecución de experimentos del método inducción de patrones.....	62
5.3.2. Ejecución de experimentos del método identificación de argumentos .....	64
5.4. Discusión.....	66
Capítulo 6 .....	68
6.1. Presentación.....	69
6.2. Lecciones aprendidas.....	70
6.3. Contribuciones.....	72
6.3.1. Productos .....	73
6.3.2. Aportaciones.....	73
6.4. Trabajos futuros.....	74
Referencias .....	75
Anexo A. Etiquetas morfológicas empleadas por Freeling .....	79
Anexo B. Etiquetas de fragmentos de texto empleadas por Freeling .....	89
Anexo C. Expresión regular .....	93
Anexo D. Manual de instalación .....	94
Anexo E. Sistema identificador de argumentos en páginas web .....	97

## Lista de Figuras

Figura 2.1. Esquema argumentativo en un discurso, propuesto por Toulmin (1958) .....	10
Figura 2.2. Precision y Recall .....	15
Figura 2.3. Análisis sintáctico de Freeling representado por ETE, tomando un fragmento textual de un discurso político. ....	17
Figura 3.1. Diagrama de anotación de argumentos. ....	19
Figura 3.2. Modelo propuesto para organizar los componentes de los argumentos basado en roles de discusión. ....	24
Figura 4.1. Método de solución propuesto en esta tesis .....	32
Figura 4.2. Procedimiento del método de inducción de patrones. ....	37
Figura 4.3. Representación parcial de un árbol sintáctico del párrafo citado .....	39
Figura 4.4. Parte de un análisis sintáctico de Freeling 4.0 .....	39
Figura 4.5. Análisis de nivel contextual del marcador "por consiguiente" .....	40
Figura 4.6. Salida de estadísticas de las etiquetas de Freeling .....	45
Figura 4.7. Proceso del método extiende las funcionalidades de una expresión regular con análisis morfosintáctico .....	47
Figura 4.8. Diagrama de las reglas de Capaldi y como se implementa la inducción de patrones. ....	51
Figura 4.9. Un modelo de expresión regular para identificar argumentos.....	52
Figura 4.10. Diagrama de flujo del proceso de identificación de argumentos. ....	54
Figura 4.11. Argumentos identificados en formato JSON.....	55
Figura 4.12. Ejemplo de estructura de argumentos interrelacionados o sorites.....	56
Figura 5.1. Ejemplo de etiquetado de argumentos por colores. ....	61
Figura 5.2. Desglose de la lista de argumentos. Los colores indican un argumento de un subargumento.....	62

## Lista de Tablas

Tabla 2.1. Partes de una etiqueta Eagle por categoría .....	13
Tabla 2.2. Representación parcial de verbos de etiquetas POS por Freeling.....	13
Tabla 3.1. Tabla de resultados de varios métodos implementados por (Nguyen & Litman, 2015) .....	21
Tabla 3.2. Evaluación de caso base CRF: palabras + etiquetas POS + contexto 2/5.....	23
Tabla 3.3. Resultados obtenidos en distintos niveles de varios clasificadores. ....	25
Tabla 3.4. Tabla comparativa de los trabajos relacionados.....	27
Tabla 4.1. Comparación de la metodología tradicional con la metodología propuesta. ....	33
Tabla 4.2. Marcadores argumentativos encontrados por algunos recursos.....	35
Tabla 4.3. Continuación .....	36
Tabla 4.4. Análisis morfológico de oraciones con el marcador "por lo tanto" .....	41
Tabla 4.5. Representación morfológica de los marcadores de conclusión .....	43
Tabla 4.6. Continuación .....	44
Tabla 4.7. Total de marcadores identificados que usa el sistema. ....	48
Tabla 4.8. Continuación .....	49
Tabla 5.1. Conteo de marcadores en los argumentos identificados manualmente.....	60
Tabla 5.2. Marcadores argumentativos identificados en los discursos políticos consultados	63
Tabla 5.3. Resultados de los marcadores argumentativos encontrados con el método de inducción. ....	64
Tabla 5.4. Tabla de resultados de 3 experimentos.....	66
Tabla 6.1. Resumen de objetivos alcanzados.....	69

## **Lista de Ecuaciones**

Ecuación 2.8-1 Fórmula del precision .....	15
Ecuación 2.8-2 Formula del recall.....	15
Ecuación 2.8-3 Formula del F-measure .....	16

# Capítulo 1

## Introducción

---

En este capítulo se presenta un breve contexto de la temática de esta tesis, así como el panorama de este tipo de investigación. Las secciones que se presentan en este capítulo son los objetivos específicos y generales, el planteamiento del problema, la justificación y la estructura del documento de la tesis.

## 1.1. Introducción

En nuestra vida diaria lo que escuchamos y decimos frecuentemente contiene argumentos. Éstos están presentes en la información que se transmite por televisión, la radio, el periódico, las personas, internet, etc. También los construimos para plantear nuestras propias ideas sobre algún tema o para refutar la de otros (Rodríguez Jiménez, 2013).

Por lo demás, existe otro concepto muy importante, la minería de argumentos, relativamente nuevo; implica identificar automáticamente la estructura argumentativa dentro de los documentos, por ejemplo, premisas y conclusión, así como el esquema de la argumentación. Existen muchas aplicaciones como la mejora y recuperación de la información (*Linguistics*, 2014).

El objetivo principal de la argumentación es convencer, pero para abordar el estudio de la argumentación en todas sus formas es necesario considerar la lógica informal, para ser más preciso, la teoría de la argumentación (Díaz, 2002).

El argumento puede tener muchos significados, por ejemplo, puede ser una sucesión de hechos, episodios, etc.; o puede ser un resumen del asunto de una obra literaria. En este trabajo nos referimos al que lo define como el objeto de estudio de la lógica informal. Los argumentos deben propiciar inferencias o razonamientos pero además deben ser capaces de persuadir, convencer, permitir acuerdos o resolver diferencias de opinión (Harada Olivares, 2009). Es por eso que llamaremos a estos argumentos como argumentos lógicos, para distinguirlos de los argumentos informativos, también llamados descriptivos o narrativos.

Un argumento lógico se compone de premisas que pueden ser suposiciones, leyes, reglas, ideas ampliamente aceptadas u observaciones; y una conclusión. Las premisas y conclusiones juntas componen un argumento lógico (*Matemática: Razonamiento Y Aplicaciones 10/e*, 2006). Se dividirán los argumentos lógicos en argumentos con marcadores y sin marcadores, esta investigación se enfocará en los argumentos con marcadores.

La principal razón de identificar automáticamente argumentos en los textos, es que permite que los sistemas informáticos adquieran la capacidad de analizar razonamientos y luego generar conocimiento a través de la información obtenida de los argumentos.

Por otra parte, a partir de los años 60 del siglo XX se dio inicio (particularmente en Francia) a una serie de investigaciones sobre el discurso político; estas investigaciones dieron origen a una nueva disciplina llamada *análisis del discurso* (Murillo Serna, 2001). En los discursos políticos existen dos áreas de investigación, por un lado, el estudio del lenguaje y por el otro, el estudio de los fenómenos políticos. Por lo tanto, dado que los discursos políticos tienen un área para el estudio del lenguaje, y la lógica informal se usa en el lenguaje, este es el motivo por el cual se tomaron los discursos políticos para identificar y extraer los argumentos.

A pesar del creciente interés que se está dando al estudio de la argumentación, en la actualidad sigue considerándose una investigación pobre. Los recientes avances consisten en la minería de argumentos, que incluye la identificación de argumentos, así como sus componentes, sin embargo, aún no se cuenta con los avances necesarios para identificar con una gran efectividad los argumentos y sus componentes.

En la presente investigación se desarrolla un método para identificar de manera automática argumentos explícitos, principalmente en los discursos políticos, pero de igual forma puede utilizarse para otros tipos de textos que se encuentren en la red. El resultado final de este trabajo es una extensión del navegador de Chrome que permite identificar los argumentos en los sitios web, haciendo uso del algoritmo desarrollado que se encuentra en el servidor web.

## 1.2. Planteamiento del problema

El tema de la identificación de argumentos es parte de la minería de argumentos, la cual conlleva nuevos retos en el análisis del discurso. Es un desafío relativamente nuevo en el análisis de discursos basado en corpus (3rd Workshop on Argumen Mining, 2016). Además, los trabajos son relativamente nuevos y utilizan técnicas de Procesamiento del Lenguaje para identificar los argumentos, la estructura de éstos y las relaciones que pueden encontrarse con otros argumentos.

Un problema claro es que las investigaciones propuestas hasta ahora no son muy precisas o requieren que el lenguaje sea en un idioma específico y con temas particulares. Otro problema es que la argumentación es un tema intrínsecamente multidisciplinar que implica la filosofía, la psicología, los estudios de comunicación, la lingüística y la informática (Cabrio, Hirst, Villata, & Wyner, 2016) y eso ocasiona muchas dificultades en abordar su tratamiento tanto computacionalmente como manualmente (identificación por parte de lector).

Así mismo los argumentos no son muy comunes en los textos, porque normalmente el propósito del emisor es únicamente informar (Calzado, 2014), sin embargo en el tema político, al discutir temas recurrentes como la pobreza, la garantía de los derechos de los ciudadanos, salud pública, infraestructura, educación, economía, entre otros; los políticos comúnmente plantean conclusiones para guiar el razonamiento de las personas y así ser persuadidos. En consecuencia, es necesario ayudar al lector a identificar argumentos, permitiendo entender mejor el propósito principal de los discursos presentados por los políticos y/o dependencias de gobierno.

Existen numerosas justificaciones que llevan a cabo esta investigación. En primera, La identificación de argumentos permite ayudar en el debate, así como en el pensamiento crítico. Por esa razón, es crucial identificar argumentos, para defenderse apropiadamente (Harada Olivares, 2009). Las razones por las que se argumenta son dos: para convencer o para persuadir a los receptores. Estas razones conllevan un gran impacto a las personas porque nos tratan de influir, hacer cambiar de opinión y conducta (Arcentin, 1993).

La identificación de argumentos es una tarea muy importante porque ayuda a entender mejor el tema que se está tratando, permite identificar fallos en el razonamiento (falacias o sofismas), y facilita la creación de contraejemplos.

Desarrollar un sistema que permite identificar argumentos de manera automática trae varias ventajas a la Inteligencia Artificial, específicamente al Procesamiento del Lenguaje Natural (PLN) y por consiguiente a los seres humanos. Las ventajas obtenidas al identificar argumentos con los sistemas computacionales permite la generación automática de resúmenes de documentos; capacidad de recuperación, extracción y visualización de información, minería de



opiniones, generación de preguntas y respuestas, y análisis de sentimientos (Peldszus, 2014)(Padró, Collado, Reese, Lloberes, & Castellón, 2010) (Wellner & Pustejovsky, 2007).

Por estas razones se propone desarrollar un método basado en expresiones regulares que sigan ciertas reglas que identifican argumentos, estas reglas representan patrones para identificar diferentes características de los argumentos.

## 1.3. Objetivo

### Objetivo general

Desarrollar un método automático capaz de identificar argumentos explícitos a través de la identificación de marcadores lingüísticos argumentativos de conclusión y premisas en discursos políticos escritos en español.

### Objetivos específicos

- Crear un corpus de argumentos sobre discursos políticos etiquetados manualmente.
- Desarrollar el método de inducción de patrones para identificar una mayor cantidad de marcadores de discurso.
- Desarrollar un algoritmo recursivo para identificar argumentos explícitos.
- Desarrollar una herramienta que utilice de los métodos propuestos.

## 1.4. Estructura del documento

El contenido de la tesis se encuentra organizado en los siguientes capítulos:

- Capítulo 2. Marco teórico: En este capítulo se definen los principales conceptos para el tema de tesis desarrollado.
- Capítulo 3. Estado del arte: En este capítulo se presentan los trabajos relacionados con la minería de argumentos.
- Capítulo 4. Método de solución. En este capítulo se describen las dos fases del método de solución que describen el procedimiento para incrementar el repertorio de marcadores lingüísticos argumentativos y el procedimiento para identificar argumentos recursivamente.
- Capítulo 5. Resultados. En este capítulo se presentan las pruebas realizadas para la fase uno y fase dos del método de solución.

- Capítulo 6. Conclusiones y trabajos futuros. En este capítulo se describe un sistema identificador de argumentos en páginas web que implementa un algoritmo recursivo. Así como conclusiones y contribuciones que se derivan de esta investigación.

# Capítulo 2

## Marco Conceptual

---

En el presente capítulo se presentan conceptos que son utilizados como objeto de investigación de este trabajo, así como aquellos conceptos útiles que dan soporte al desarrollo.

Por tal motivo, en el marco teórico se incluye conceptos tales como, lógica, argumentación, marcadores argumentativos, entre otras.

## 2.1. Argumentación

La argumentación o argumento es una razón que se pretende defender por una idea u opinión mediante la persuasión (Alberto, 2008). Argumentamos cuando queremos defender nuestro punto de vista o apoyar o refutar un hecho, con el fin de convencer a los participantes involucrados y de esta manera que se pongan de nuestro lado. Para poder convencer (persuadir) a nuestros oyentes o lectores necesitamos dar credibilidad al discurso que realizamos de una manera ordenada que cambien las convicciones de los participantes (Marimón Llorca, 2008).

### Argumento

Un argumento lógico son grupos de proposiciones que se diferencian de premisas y conclusiones, las premisas y conclusiones son una relación inferencial lógica; para que un grupo de proposiciones se consideren argumentos, requiere que la conclusión implicada sea un razonamiento de las premisas. En un discurso no todo es un argumento, existen otras porciones del discurso que son en su mayor parte utilizadas en los actos del habla o escritos; por ejemplo, advertencias, informes, ilustraciones, entre otros. Existe el término demostración, que se puede utilizar como sinónimo de argumento ya que están estrechamente relacionados, pero existe una diferencia sustancial.

El experto en ciencias políticas Giandoménico Majone ve así la diferencia entre la demostración y la argumentación (Díaz, 2002):

“Primero, la demostración es posible sólo dentro de un sistema formalizado de axiomas y reglas de inferencia. La argumentación no parte de axiomas, sino de opiniones, valores o puntos de vista refutables; utiliza inferencias lógicas, pero no se acota en sistemas deductivos de enunciados formales. En segundo lugar, una demostración trata de convencer a todos aquellos que cuentan con los conocimientos técnicos requeridos, mientras que la argumentación se dirige siempre a un auditorio particular y trata de provocar o incrementar su adhesión a las tesis que se presentan para su consentimiento”

Un argumento se compone principalmente de premisas y conclusiones, pero existen argumentos más elaborados, lo que significa que llegan a contener otros elementos denominados soportes. Asimismo, un argumento puede contener a otros argumentos o estar relacionado con otros argumentos externos.

Las premisas son enunciados que ofrecen información y las conclusiones son enunciados del resultado del razonamiento deductivo o inductivo de las premisas (Rodríguez Arias, 2010).

Los soportes son oraciones que apoyan o refutan a la conclusión, estos se pueden considerar como premisas de la conclusión.

## 2.2. Lógica

La lógica es una ciencia, muchos pueden pensar que es una ciencia del razonamiento, esto puede interpretarse erróneamente. La lógica no estudia la observación de los hechos del pensamiento o razonamiento, ni los procesos que se llevan a cabo como muchas otras ciencias, por ejemplo, la física que investiga cómo funcionan las cosas del mundo físico y luego trata de interpretarlas por medio de reglas o leyes. En cambio, la lógica es una disciplina normativa, es decir, nos dice *cómo debe ser* y no *cómo lo hacemos nosotros*; por ejemplo, si le proponemos a una persona realizar una suma, y esta persona realiza dicha suma incorrectamente, nosotros debemos de convencerle que la suma da otro número el cual es el correcto. Cómo podemos apreciar, el lógico (persona) no está analizando el proceso de la suma mal realizada de la otra persona, sino que está dando él su respuesta correcta y con ello el razonamiento correcto; “lo que debe ser”. Por lo tanto, si alguien razona equivocadamente podemos hacerle algo de lógica para mostrarle que su razonamiento es incorrecto e invitarle a cambiar de conducta. Esto es lo que significa como ciencia normativa (Goldstein, 2008).

Conceptualizándolo de otra forma por palabras del filósofo (Kant, 2001) dice:

“La lógica es una ciencia racional, no según la mera forma, sino según la materia; una ciencia a priori de las leyes necesarias del pensamiento, pero no con respecto a objetos particulares, sino con respecto a todos los objetos en general; una ciencia, por lo tanto, del uso correcto del entendimiento y de la razón en general, pero no en sentido subjetivo, es decir, no según principios empíricos (psicológicos): cómo piensa el entendimiento, sino en sentido objetivo, es decir, según principios a priori: cómo debe pensar.”

Existen dos formas de representar la lógica; lógica formal y lógica informal. Esta última será de nuestro interés en esta tesis.

### 2.2.1. Lógica formal

La lógica formal es una parte de la lógica, su principal distinción se basa en la forma o estructura de los razonamientos, por lo que se puede tratar el contenido con mayor facilidad. La lógica formal puede representarse mediante lenguajes formales, sistemas deductivos y semánticas formales.

En la lógica formal, se realizan razonamientos deductivos, esto significa que obtienes la conclusión de la información contenida en las premisas, sin importar, que tan ciertas sean las premisas con base a la información general del mundo.

Un ejemplo de lógica formal es, “Todos los hombres son mortales, Sócrates es un hombre, por lo tanto, Sócrates es mortal” (“*Logical Fallacies*,” 2009).

En el ejemplo anterior, en la segunda premisa: “Sócrates es hombre”, nos está afirmando que es una persona, por lo tanto, podemos deducir bajo la primera premisa que Sócrates por ser un hombre es mortal, por ende, estamos realizando un razonamiento deductivo.

### 2.2.2. Lógica informal

La lógica informal es la forma en que se puede evaluar y analizar los argumentos en un lenguaje natural; al contrario, a la lógica formal en que los argumentos siguen un patrón en la redacción en el caso escrito. Los argumentos informales son los más difíciles de analizar debido a su naturaleza. Los argumentos informales que evalúa la lógica informal aborda todo tipo de razonamientos, como científicos, legales, entre otros; pero el objetivo principal es tener una relación de los argumentos que se dan en la vida diaria, por ejemplo social y político, como es el caso de los medios de comunicación (periódicos, revistas, televisión, internet, etc) (Groarke, 1996).

## 2.3. Discurso

Para poder definir que es un discurso político se necesita definir la noción de discurso. Con base a lo que dice van Dijk, un discurso tiene dos aspectos: es un suceso de comunicación y es interacción verbal; en el primer aspecto se refiere que es una secuencia coherente de actos de habla donde están entrelazados unos con otros; en el segundo aspecto, se refiere, a que los participantes se conceden turnos en la conversación, se defienden, tratan de persuadir, responden a las preguntas de los demás participantes, etc. Pero en general, siempre están relacionados con el ambiente social (Koutný, 2006).

Existen varios tipos de discursos, éstos se pueden clasificar en la forma de comunicarse, en las funcionalidades del lenguaje (Discurso emotivo, apelativo, informativo, entre otras), según la variación del lenguaje (Discursos socioculturales), según el modo de construcción (Discurso narrativo, descriptivo, argumentativo, entre otras) y en función de la temática (Discursos político, religioso, empresarial, entre otros) (Lucía, 2013). Ésta tesis se centra en el discurso político, el cual tiene otras características del discurso. Por ejemplo, un discurso político es apelativo, argumentativo, en algunos casos emotivos, etc.

### Discurso político

El discurso político es producido dentro de la escena política, estas escenas políticas pueden ser en sentido estricto, el discurso presidencial, partidos políticos, prensa política especializada, discurso emitido por los medios de comunicación en ciertos momentos y en algunas ocasiones, el magisterial, el del ejército, y la policía. Este tipo de discurso es estratégico, en la medida que se definen propósitos, medios y antagonistas; es performativo, lo que significa que quien lo emite no se sustenta con transmitir una convicción o informar, sino que también produce un acto, expresa un compromiso y asume una posición (Gutiérrez, 2005).

El discurso político se centra en la política. La política en un sentido amplio es toda actividad relacionada con el poder. El discurso político comunica y además hace reconocer un discurso de autoridad, tratando de imponer ideas al oyente y además, conseguir que actúe de acuerdo con las ideas expuestas en el discurso (Koutný, 2006).

## 2.4. Marcadores del discurso

Recientes investigaciones de la lingüística han llevado a interesarse sobre unas unidades llamadas marcadores del discurso, también denominados marcadores lingüísticos, introductores, textuales, entre otros. Unidades como, *por tanto*, *en suma*, *en fin*, *entre otras*. Estas unidades son los que dan estructura a los textos, son aquellos que guían la realización de inferencias cuando se procesa un texto. Los marcadores residen en su carácter de marcas de sentido de procesamiento, no conceptual (Sánchez Avendaño, 2005); en otras palabras:

“Los marcadores del discurso son unidades lingüísticas invariables, no ejercen una función sintáctica en el marco de la predicación oracional y poseen un cometido coincidente en el discurso: el de guiar, de acuerdo con sus distintas propiedades morfosintácticas, semánticas y pragmáticas, las inferencias que se realizan en la comunicación (Portolés, 2011)”

En los marcadores desde el punto de vista de su forma, pueden ser unidades simples o compuestas, como conjunciones, adverbios y locuciones, sintagmas nominales, verbales o preposicionales. Su función principal es relacionar dos enunciados.

Existen una gran clasificación de marcadores, éstos se pueden clasificar por su organización global del texto (iniciadores, distribuidores, de transición, entre otras), porque introducen operaciones discursivas (de expresión de punto de vista, de manifestación de certeza, de confirmación, de reformulación, entre otras) y, por último, conectores que sirven para poner en relación lógico-semántica segmentos textuales (aditivos o sumativos, contra argumentativos, de base causal, entre otras).

### Marcadores argumentativos

Los marcadores argumentativos identifican la estructura de los argumentos en un discurso textual (Calsamiglia & Tusón, 1999). En Figura 2.1, Toulmin (1958) propone un esquema de los argumentos en un discurso.

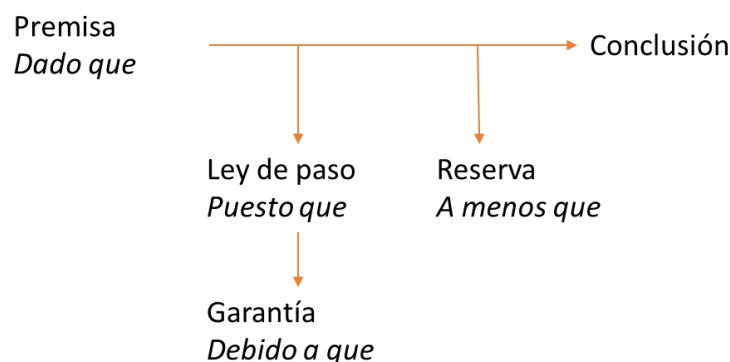


Figura 2.1. Esquema argumentativo en un discurso, propuesto por Toulmin (1958)

(Capaldi, 2011), propone unas reglas que hacen uso de marcadores lingüísticos para identificar las premisas y conclusiones de los argumentos y por lo consiguiente siguiendo estos pasos identifican todos los argumentos que puede haber en un discurso.

## 2.5. Procesamiento de Lenguaje Natural

El Procesamiento de Lenguaje Natural (PLN) es una disciplina reciente, que nace en la década de los 60 del siglo XX, pertenece como subárea de la Inteligencia Artificial y de la Lingüística, con el objetivo de estudiar los problemas de la generación y comprensión automática del Lenguaje Natural (Fabra, 2012).

El PLN realiza las principales tareas (Gelbukh & Sidorov, 2010) (Alberich, 2007):

- Síntesis del discurso
- Análisis del lenguaje
- Comprensión del lenguaje
- Reconocimiento del habla
- Síntesis de voz
- Generación de lenguajes naturales
- Traducción automática
- Respuestas a preguntas
- Recuperación de la información
- Extracción de la información

### 2.5.1. Los niveles de estudio del PLN

Existen distintos niveles o análisis de lingüística que se utilizan para el procesamiento del lenguaje natural, dependiendo del análisis que se quiera realizar es el nivel a emplear.

A continuación, se colocan los distintos niveles del PLN, no obstante, en este trabajo no se tomará el primer nivel, por las razones en que se analiza el texto escrito, así como tampoco el nivel pragmático.

El primer nivel se llama **nivel fonético y fonológico**, este nivel se emplean cuando queremos analizar los sonidos del lenguaje hablado y sus representaciones abstractas (fonemas); en este nivel se aplica reconocimiento de la voz y de la síntesis de voz.



El segundo nivel es el **nivel morfológico**; estudia la forma en que se adoptan los lemas al contexto de uso, así como también, las unidades mínimas de modificación de forma, llamadas morfemas, este análisis extrae lexemas (significado con referencia al mundo externo) y morfemas (significado correspondiente al mundo de las categorizaciones o estructura).

El tercer nivel denominado **nivel sintáctico o sintaxis**, se refiere a las relaciones en los subconjuntos de una frase (denominadas sintagmas) o en la frase general. Básicamente establece la relación entre los términos dentro de una oración, es decir, identifica el rol del término a partir de una frase y las dependencias con los otros términos.

El quinto **nivel semántico**, se centra en tratar el significado de los términos, en otras palabras, establecer relaciones entre significados y significantes; tratando de este modo la ambigüedad, tanto en casos de sinonimia, antonimia así como de polisemia (Alberich, 2007).

El último nivel lingüístico, es la **pragmática**, en él se estudia las interpretaciones que le dan los hablantes a los enunciados en contexto. Es el nivel más complejo de todos por ser una disciplina que estudia el lenguaje en función a la comunicación, ya que cada persona produce significados muy diversos, pero todos llegan a modificar una situación (Reyes, 1990).

### 2.5.2. Etiquetado POS

En la lingüística computacional, también conocido como etiquetado gramatical porque lo que se refiere el etiquetado POS (*Parts Of Speech*) es asignar a cada palabra del texto su categoría gramatical, es decir, dado el siguiente ejemplo: “El gato come carne”. La palabra come es de categoría gramatical verbo-masculino-indicativo-presente-tercera persona-singular, así sucesivamente con las demás palabras.

En este trabajo se hará uso del etiquetado POS, proporcionado con la herramienta Freeling. Así mismo en este trabajo se permitirá usar el etiquetado *chunk* (fragmentos) que no son más que categorizar conjuntos de palabras, por ejemplo, cuales son sintagmas nominales.

## 2.6. Freeling

Freeling es una librería de procesamiento de multilinguaje de código abierto proporcionando una gama amplia de funcionalidades de análisis para varios idiomas.

Freeling puede manejar los idiomas de inglés, español y catalán, y en un menor grado gallego, italiano, galés, portugués, asturiano y ruso.

Además, maneja diferentes tipos de análisis como análisis morfológico, etiquetado POS, análisis sintáctico superficial, análisis sintáctico completo, análisis de dependencias, correferencias y gráfico semántico. Cada uno de ellos es configurable, los tipos de opciones en cada análisis son reconocimiento de los números, reconocimiento de las horas y fechas, cuantificadores, radios y porcentajes. Detección de entidades nombradas, clasificación de entidades nombradas, detección de multipalabra, entre otras (Padró & Stanilovsky, 2012).

## Etiquetas Freeling

Freeling maneja dos tipos de etiquetas POS, las etiquetas *Eagle* y *chunk*. Las etiquetas *Eagle* representan la información morfológica de las palabras. En la siguiente tabla se presenta la construcción de la etiqueta para cada palabra.

Tabla 2.1. Partes de una etiqueta Eagle por categoría

ETIQUETAS			
Posición	Atributo	Valor	Código
<i>Columna 1</i>	<i>Columna 2</i>	<i>Columna 3</i>	<i>Columna 4</i>

Existen varias categorías, algunas son los adjetivos, adverbios, determinantes, entre otras. Para cada categoría se presenta los atributos, valores y códigos que puede tomar (Tabla 2.1). En la columna 1, indica la posición en la que aparecen los atributos (columna 2). La columna 2 indica los atributos que puede tener cada categoría. Por ejemplo, categoría, tipo, grado, persona, entre otros. La columna 3 indica los valores que puede tener cada atributo. Por último, la columna 4 representa los códigos que *Freeling* estableció para su representación.

Un ejemplo se observa en la Tabla 2.2 con la categoría “verbos”.

Tabla 2.2. Representación parcial de verbos de etiquetas POS por Freeling

VERBOS			
Pos.	Atributo	Valor	Código
1	Categoría	Verbo	V
2	Tipo	Principal	M
		Auxiliar	A
		Semiauxiliar	S
3	Modo	Indicativo	I
		Subjuntivo	S
		Imperativo	M
		Infinitivo	N
		Gerundio	G
		Participio	P
...	...	...	...

Si se quiere saber su representación morfológica del verbo “cantada”, su lema es “cantar” y su etiqueta es “VMP00SF”, que significa, que su categoría es verbo de tipo principal de modo participio, y así sucesivamente. En el Anexo A se pueden observar todas las tablas por categorías.

Por otro lado, están las etiquetas *chunk*, estas etiquetas agrupan varias palabras que representan una unidad. La unidad se construye con base a una palabra que representa el segmento agrupado, por ejemplo, la frase “Juan quiere comer plátanos” *Freeling* identifica “comer plátanos” como *grup-verb-inf*, que significan “infinitivo cuya cabeza es preposicional”. Así mismo los agrupamientos se hacen cada vez mayor hasta cubrir toda la frase. Por ejemplo, toda la frase presentada es *grup-verb* que significa que la frase es un segmento verbal. En el Anexo B aparecen todas las etiquetas *chunk* que maneja *Freeling*.

## 2.7. Expresiones regulares

Las expresiones regulares también conocidas como *regex*, *regexpr* entre otras; definen una serie de patrones que luego son utilizados para trabajar con cadenas de texto, en otras palabras, podemos considerar a las mismas como un lenguaje específico diseñado para localizar texto (Deitel, 2004). Esta serie de patrones son secuencias de caracteres y símbolos que definen un conjunto de cadenas. La necesidad de localizar texto es para validar la entrada y asegurar que los datos están en un formato específico (Montoro, 2013).

Por ejemplo, un código postal debe consistir de cinco dígitos, y validar un campo que sea únicamente texto. Así como estos ejemplos existen muchas aplicaciones que tienen las expresiones regulares, casi todo lo que se usa en un sistema computacional contiene expresiones regulares, ya han sido parte de los desarrolladores.

Las expresiones regulares de los ejemplos mencionados son:

- Para validar que sea código postal: `[0-9]{5}`
- Campo de texto: `([A-Z])\w+`

## 2.8. Medidas de evaluación

Las métricas habituales para el ámbito de la recuperación de la información (RI) y el reconocimiento de patrones son la cobertura (*recall*), precisión (*precision*) y medida F (*F-measure*). Estas métricas tienen como funcionalidad evaluar la efectividad de los algoritmos desarrollados.

El *recall* mide la proporción de términos correctamente conocidos respecto al total de términos reales. Mientras que *precision* mide el número de términos correctamente reconocidos respecto al total de términos predichos, ya sean estos términos verdaderos o falsos, es decir el *precision* está midiendo la pureza. Por otro lado, también el *F-measure* es una medida del RI que caracteriza con un único valor la bondad de un clasificador (Alcina & Valero, 2009).

$$precision = \frac{|documentos\ relevantes \cap documentos\ recuperados|}{|documentos\ recuperados|}$$

Ecuación 2.8-1 Fórmula del precision

Observe la Ecuación 2.8-1 del *precision*, que indica el valor absoluto de la intersección de los documentos relevantes y de los documentos recuperados entre el valor absoluto de los documentos recuperados. Entretanto la fórmula del *recall* es el valor absoluto de la intersección de los documentos relevantes y documentos recuperados, todo eso entre el valor absoluto de los documentos relevantes.

$$recall = \frac{|documentos\ relevantes \cap documentos\ recuperados|}{|documentos\ relevantes|}$$

Ecuación 2.8-2 Formula del recall

Como se puede observar en la Figura 2.2, los documentos relevantes son todos aquellos puntos que están en el área sombreada de color gris fuerte (derecho), mientras que los documentos recuperados son los que están dentro del círculo, por lo tanto la intersección de éstos dos es el numerador del *precision* y del *recall* (Ecuación 2.8-2), es decir son los puntos grises que se encuentran en área verde. Mientras que el denominador difiere en estas dos métricas, como se aprecia en las respectivas fórmulas de cada métrica.

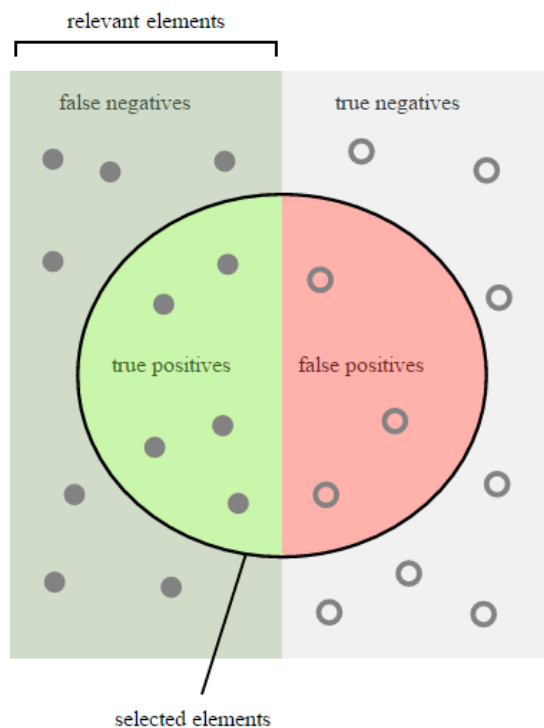


Figura 2.2. Precision y Recall

Por otro lado, el *F-measure* no es más que una combinación de *precision* y *recall*, como se muestra en la Ecuación 2.8-3.

$$F = \frac{2 \times recall \times precision}{cobertura + precision}$$

Ecuación 2.8-3 Formula del F-measure

## 2.9. Marco de trabajo

Los marcos de trabajo o comúnmente llamados *framework* son un conjunto de librerías o bibliotecas que están relacionados que permiten realizar algún tipo de aplicación de software.

Normalmente un lenguaje de programación puede tener muchos *framework*, un desarrollador independiente o el desarrollador del propio lenguaje de programación crean un framework para agilizar la labor de otros desarrolladores, cubriendo los puntos más importantes de un diseño, Por ejemplo, logueo, ataques CSRF (*Cross-site request forgery*) entre otras, conexiones y accesos a bases de datos, diseño de interfaces, entre muchas más.

En este trabajo se hace uso de tres marcos de trabajo en dos lenguajes de programación,

YiiFramework desarrollado sobre y para PHP; éste se utilizó para pre-procesamiento de los datos, y Django desarrollado sobre y para Python, utilizado para las pruebas y de la implementación del sistema final.

Del mismo modo se tiene ETE Toolkit, un marco en Python especializado en trabajar con árboles su representación se deriva del formato Newick. Newick es un estándar para la representación de árboles en forma legible por ordenador (Wikipedia, 2016). Permite la creación árboles, visualización y anotación, etc; este framework se utilizó para trabajar con los resultados de Freeling (Huerta-Cepas, Serra, & Bork, 2016). ETE Toolkit se desarrolló en un principio para la filogenómica, por lo que utilizarlo en árboles de Freeling tendría mucho potencial, por consiguiente, podría utilizarse para muchos análisis más. La manipulación de esta herramienta con los árboles hace más fácil interactuar con los árboles de Freeling.

En la Figura 2.3 se aprecia la utilidad del marco ETE Toolkit. En este trabajo el marco representa exactamente el análisis sintáctico de Freeling.

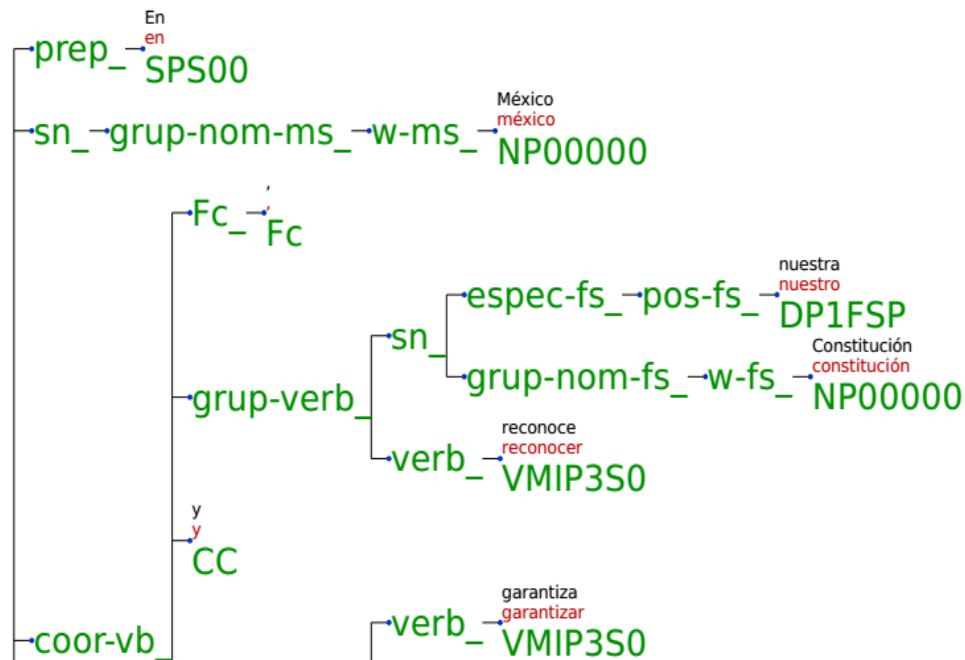


Figura 2.3. Análisis sintáctico de Freeling representado por ETE, tomando un fragmento textual de un discurso político.

## 2.10. Discusión

El tema de la argumentación es un tema meramente de la filosofía y es por eso que en el ámbito de la computación resulta complicado abarcar esa área, ya que este estudio es muy abstracto. En el lenguaje cotidiano normalmente se utiliza el concepto argumento como algo más general, es decir, cualquier oración dada sobre un tema es considerado argumento, lo cual es erróneo en muchos casos. Aunque argumentar se trata de convencer y persuadir, y las personas intentan imponer sus ideas, la realidad es que pocas veces se argumenta, ya que se tiene que cumplir con ciertas características; por ejemplo, un argumento se compone esencialmente de premisas y conclusiones, y estas premisas pueden contener más argumentos. En este capítulo se introdujo otros conceptos sobre la temática computacional, la razón de esta selección es que nos ayudan a cumplir nuestro objetivo de identificar argumentos de forma diferente a los demás métodos. Por otro lado, se investigaron medidas de desempeño que se apegue a nuestra problemática.

# Capítulo 3

## Estado del Arte

---

El estudio de la argumentación en el área computacional es reciente porque apenas iniciaron los trabajos que tratan el tema. Los trabajos más antiguos encontrados en el transcurso de esta investigación datan a partir del año 2005. Los trabajos mencionados a continuación son los más recientes encontrados y por ello fueron seleccionados como estado del arte. El mayor auge de la argumentación en la computación se da a partir del año 2014 con el primer y el segundo *workshop* de la minería de la argumentación.

Estos trabajos se enfocan en temas de dominio específico, es decir, textos bajo ciertas temáticas como la salud o el deporte, y con algún idioma en particular, por ejemplo, inglés o griego.

### 3.1. Identificación de argumentos

Este grupo de trabajos tratan sobre la identificación de argumentos, las relaciones argumentativas o sus componentes, utilizando diferentes métodos de Procesamiento de Lenguaje Natural e Inteligencia Artificial.

Se tomaron en cuenta los trabajos que están más relacionados a la tarea de identificación de argumentos y los más actualizados, ya que muchos de estos trabajos son la mejora de otros trabajos previos.

#### 3.1.1. La identificación de estructuras en el discurso argumentativo de ensayos persuasivos

El objetivo del trabajo de (Stab & Gurevych, 2014b), es identificar los componentes y relaciones de los argumentos.

Para realizar dichas tareas hicieron un corpus con un novedoso esquema de anotación para no perder la estructura de los argumentos (Stab & Gurevych, 2014a), Figura 3.1. La anotación contiene partes de los argumentos. Estas anotaciones son las conclusiones mayores, conclusiones y premisas que está conectados con el apoyo y el ataque de relaciones argumentativas. El corpus comprende 90 ensayos incluyendo 1,673 frases.

De este corpus se analizó y se obtuvieron 90 conclusiones mayores, 429 conclusiones, 1033 premisas y 327 casos no argumentativos.

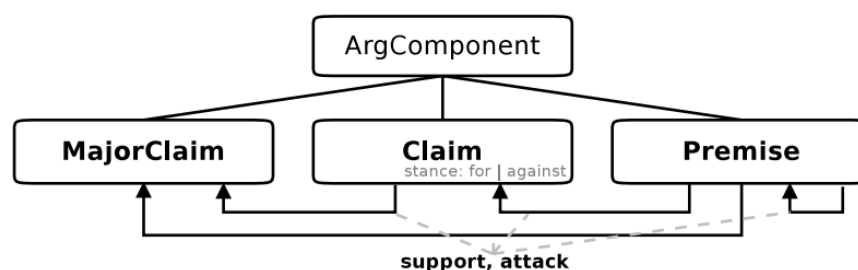


Figura 3.1. Diagrama de anotación de argumentos.

La primera parte del trabajo, que es identificar los componentes, consiste en identificar premisas y conclusiones, para ello, se obtuvieron cinco características esenciales, estas son características estructurales que están basados en *tokens*, localización y signos de puntuación. Características léxicas que son n-gramas, verbos, adverbios y modelos; los verbos y adverbios juegan un papel muy importante para identificar las partes de un argumento, verbos como 'creer' o 'pensar' proporcionan señales para indicar la existencia de la conclusión mayor y adverbios como 'debería' o 'podría' son frecuentemente usados en discursos argumentativos para expresar conclusiones. Las características sintácticas se capturan de un árbol sintáctico en él se obtienen dos funciones, el número de sub-clausulas y la profundidad del árbol; además



se crearon reglas de producción del árbol sintáctico, es decir, son funciones de etiquetas, por ejemplo, una regla es *VP* -> *VBG*, *NP* o *PP* -> *IN*, *NP*. Los marcadores de discurso son palabras que indican que se introduce una premisa o una conclusión, por ejemplo, cuando se habla de conclusión estos pueden ser 'por lo tanto', 'así' o 'consecuentemente'; y en los casos de premisas son 'porque', 'razón' o 'además'. Asimismo, palabras que denotan primera persona. Por último, tenemos la característica contextual que consiste en la relación de las premisas y su conclusión, por ejemplo una premisa puede ser identificada como tal, si está relacionada con una conclusión.

Cada una de estas características está evaluada como características booleanas, es decir, si es verdadero en el caso de la característica sintáctica significa que aparece en el sub-árbol de un componente argumentativo.

La evaluación se llevó con matrices de confusión para los puntajes de evaluación. Los clasificadores que se usaron son McNemar Test con una precisión de 55%, Máquinas de vectores de soporte (MVS) con una precisión de 77.3%; y un análisis humano de 87.7%.

Por otro lado, la identificación de las relaciones argumentativas (soporte y ataque) también se basó de un conjunto de características, que son características estructurales, características léxica, características sintácticas, indicadores y tipo predicho. Los resultados se compararon con los mismos clasificadores con una precisión de McNemar 84.3%, humano de 95.4% y MVS de 86.3%.

### **3.1.2. Extracción de argumentos y palabras de dominio para la identificación de los componentes argumentativos en textos**

En el trabajo de (Nguyen & Litman, 2015) usa ciertas características para encontrar argumentos de (Stab & Gurevych, 2014b) pero con la diferencia de que hace uso de temas específicos; es decir, temas de política, ciencia, deporte, etc.

La mayoría de trabajos realizados en la identificación de argumentos se hacen con n-gramas y reglas sintácticas.

El objetivo de este trabajo es identificar argumentos implícitos y explícitos con base a modelos de tema para extraer argumentos y palabras de dominio, debido a que las características lingüísticas genéricas no son efectivas para cualquier tema, como son los artículos científicos.

El pre-procesamiento consiste en extraer palabras argumentativas (señales léxicas) y palabras de dominio (terminologías en temas de argumentos). Por ejemplos, en las palabras argumentativas normalmente en un argumento genérico (cualquier tema) se utilizan expresiones como '*my view is that*', '*I think*', '*to conclude*' seguido del contenido del argumento, entonces aquellas palabras argumentativas son '*view*', '*conclude*', '*think*'. Asimismo, las palabras de dominio son '*art*' y '*life*' si es un caso de tema poético. En cuanto al aspecto estructural se utilizó arboles de dependencias para extraer pares de verbos y objeto principal, por ejemplo '*I.think*' o '*view.be*' en vez de utilizar reglas de producción.

Para aprender las palabras argumentativas y palabras de dominio se utiliza el algoritmo LDA. Los datos de desarrollo para construir el modelo por temas salieron de 6794 ensayos sacados de [www.essayforum.com](http://www.essayforum.com). Posteriormente examinaron las palabras argumentativas frecuentes. Las semillas utilizadas como palabras de dominio son aquellas que no son *stop-words* y palabras argumentativas. Cada palabra de semilla de dominio está asociada con una frecuencia de ocurrencia  $f$  como una relación de las ocurrencias de semillas más de las ocurrencias totales de todas las semillas de dominio de indicaciones de ensayo.

Los datos utilizados para los estudios fueron obtenidos del corpus de (Stab & Gurevych, 2014a).

Los modelos predictivos para realizar la comparación son *baseline* de (Stab & Gurevych, 2014b) la re-implementación del mismo por los autores (Nguyen & Litman, 2015) y su propio modelo; llamadas BaseR, Basel y AD respectivamente. El modelo *baseline* trata sobre las características lingüísticas para identificar argumentos presentados en el trabajo anterior.

Los resultados se pueden observar en la siguiente tabla.

Tabla 3.1. Tabla de resultados de varios métodos implementados por (Nguyen & Litman, 2015)

<b>Models</b>	BaseR	Basel	AD	Basel	AD
<b>#features</b>	100	100	100	130	70
<b>Accuracy</b>	0.77	0.78	0.79+	0.80	0.83*
<b>Kappa</b>	NA	0.63	0.65*	0.64	0.69*
<b>F1</b>	0.73	0.71	0.72	0.71	0.76+
<b>Precision</b>	0.77	0.76	0.76	0.76	0.79
<b>Recall</b>	0.68	0.69	0.70	0.68	0.74+
<b>F1:MajorClaim</b>	0.62	0.54	0.51	0.48	0.59
<b>F1:Claim</b>	0.54	0.47	0.53*	0.49	0.56*
<b>F1:Premise</b>	0.83	0.84	0.84	0.86	0.88*
<b>F1:None</b>	0.88	1.00	1.00	1.00	1.00

### 3.1.3. Extracción de argumentos de noticias

En el trabajo de (Sardianos, Katakis, Petasis, & Karkaletsis, 2015) extraen argumentos de blogs y noticias. Identifican los argumentos, así como sus componentes. Los argumentos pueden ser normalmente descompuestos en una conclusión y o más premisas que lo justifiquen. El enfoque propuesto trata de identificar los segmentos que representan elementos de argumentos (premisas y conclusiones) en los textos de la web social. Este trabajo se ocupa de la lengua griega e incluye un pequeño conjunto de dominios temáticos, incluyendo artículos sobre política, economía, cultura, diversos sectores sociales temas y deportes.

El enfoque propuesto explota las representaciones distribuidas de palabras, extraídas de un gran corpus no anotada. Entre los aspectos novedosos de este trabajo es el dominio temático

en sí, que se refieren a la web social, en contraste con otras investigaciones, que se concentra principalmente en los documentos legales y las publicaciones científicas.

La razón por la que se centraron en la web social es que hay un aumento enorme de comunidades de la web social, hace que la identificación de los argumentos en estos textos sea una necesidad.

Un resultado de este trabajo fue la construcción de un nuevo corpus anotado manualmente.

El motivo de desarrollo de este trabajo es principalmente por necesidades en el área de la administración electrónica y la formulación de políticas, con el objetivo de llevar a cabo la extracción de argumentos con un amplio corpus recogida de la web social, destinado principalmente a los periódicos y blogs en línea.

El objetivo de esta investigación es desarrollar un enfoque para la tarea de extracción de argumentos, basado en el aprendizaje de máquina, que cumplirá con estos requisitos y será aplicable a la lengua griega. El enfoque se basa en campos aleatorios condicionales (*Conditional Random Field* o CRFs en inglés), un marco probabilístico para el etiquetado y la segmentación de los datos estructurados tales como secuencias, que se ha aplicado a una amplia gama de tareas de segmentación de reconocimiento de entidades nombradas, y análisis superficial para el análisis de opiniones basadas en los aspectos. Más allá de las características como palabras y etiquetas POS, el enfoque se aprovecha de un pequeño léxico de palabras claves, que generalmente indican la presencia de un segmento de premisa, y representaciones distribuidas de palabras. Éste mapeo de palabras a vectores tiene generalmente más de 100 dimensiones, que se crea sin la intervención humana a partir de la observación de la utilización de palabras con un amplio corpus no anotado. Específicamente hacen uso de la segunda función de Word2vec, lo que puede hacer conjeturas muy precisas sobre el significado de una palabra en función de su uso. Proporcionando suficientes datos, el uso y el contexto de cada palabra. El enfoque Word2vec intenta organizar palabras con significado similar cerca uno del otro, y es la característica que se explota en esta investigación.

El funcionamiento del enfoque de este trabajo se basa en el enfoque propuesto por (Goudas, Louizos, Petasis, & Karkaletsis, 2014), que sugiere dos fases en su método. La primera fase identifica las oraciones argumentativas, empleando clasificadores como la regresión logística (Strano & Colosimo, 2006), Naive Bayes, etc. Las características utilizadas en la clasificación se dividieron en características seleccionadas como la posición de la oración dentro del documento, así como el número de comas y conjunciones dentro de la frase. Por otra parte, examinaron el número de verbos en la sentencia, la existencia y el número de palabras claves y entidades, el número de palabras y adverbios en el contexto de una oración, y finalmente, la longitud media de las palabras en la proposición. Además de las características anteriores también examinaron la relación entre las distribuciones a través de unigramas, bigramas, trigramas de palabras y etiquetas POS. Después de la extracción de las oraciones argumentativas, se procedió a la identificación de los componentes de los argumentos. En esta etapa, se aplica un clasificador CRF en un corpus manual. Las características necesarias para esta tarea fueron las palabras de las frases, las listas de diccionario geográfico de palabras claves y léxico de los verbos y los adjetivos que aparecen con mayor frecuencia en las oraciones argumentativas de los datos de entrenamiento. Este trabajo de investigación se quiere reducir la

dependencia en los diccionarios geográficos utilizados por (Goudas, 2014), mediante la explotación de representación distribuida para las palabras, método propuesto con el Word2vec. Esto ayuda a ampliar el aspecto de las palabras que pueden ser manejados por el clasificador y, por tanto, conseguir crear un modelo CRF de grano más fino.

Los resultados de la evaluación para la extracción de argumentos con CRFs y representación distribuida de palabras, con la ayuda de un corpus notado manualmente, conteniendo segmentos anotados que corresponden a los elementos de los argumentos. Se coleccionaron 300 artículos de noticias escritas en griego, el dominio temático de los artículos varía en política, económico cultural, entre otros. El corpus fue anotado por dos estudiantes con una experiencia moderada en la anotación.

Se realizaron tres tipos de pruebas con 3 experimentos cada uno. El cual el último fue el más convincente. Las características que presenta la mejor prueba son: La incrustación de palabras para expandir el diccionario geográfico con Word2vec. Los resultados se pueden observar en la Tabla 3.2.

Tabla 3.2. Evaluación de caso base CRF: palabras + etiquetas POS + contexto 2/5

<i>Context</i>	<i>Precision</i>		<i>Recall</i>		<i>F1</i>	
0	20.74%	+2.63	11.29%	+1.88	14.60%	±2.20
+2	39.70%	+4.55	27.59%	+3.54	32.53%	±3.90
±5	38.72%	+5.29	27.60%	±3.36	32.21%	±4.06

### 3.1.4. Hacia el reconocimiento basado en segmentos de la estructura de la argumentación en textos cortos

En el trabajo de (Peldszus, 2014) desarrollan un sistema que permite identificar argumentos en micro textos. La anotación de la estructura de los argumentos está basada en la teoría de la estructura de la argumentación de (Freeman, 1991). Su idea central en la teoría de anotación que es modelar la argumentación como un intercambio hipotético dialéctica entre el proponente, quien presenta y defiende sus afirmaciones, y el oponente, quien les cuestiona críticamente de una manera reglamentada.

Para abordar el problema de la argumentación, se recogieron un pequeño corpus de microtextos alemanes que contienen una complejidad lingüística y retórica.

Realizaron varios experimentos con diferentes máquinas de aprendizaje, para demostrar que los anotadores entrenados pueden determinar la estructura de los argumentos en microtextos. La razón por la que se centraron en los microtextos es que hay un aumento rápido en la web,

por lo tanto, sugieren que hay un recurso faltante que podría facilitar el desarrollo de los sistemas de reconocimiento automáticos. Por esa razón, realizan un experimento de generación de texto diseñada para controlar la complejidad lingüística y retórica de los microtextos. Estos textos serán anotados con la estructura argumentativa propuesta, y se presenta la clasificación automática en varios niveles de granularidad del esquema.

La fiabilidad del esquema de la anotación ha sido evaluada en dos experimentos. El primero por anotadores sencillos. De ahí se construyeron 23 microtextos, que cubren varias combinaciones de construcciones básicas de la argumentación. El segundo experimento fue por tres anotadores expertos. Se anotaron los mismos microtextos y se consiguió un alto nivel de coincidencia de  $k=0.83$ .

Al final el corpus se divide en dos partes, por un lado, los 23 microtextos utilizados en los experimentos de anotación y por el otro lado, 92 microtextos que se han recogido en un experimento de generación de texto controlada.

Posterior a todo ello, se realizó una limpieza y anotación de los argumentos, eso incluye la eliminación de segmentos redundantes e insignificantes de algunos argumentos, así como la reestructuración de algunos argumentos para mantener la coherencia al eliminar los argumentos.

Por otra parte, este trabajo presenta un modelo para organizar los componentes de los argumentos. Como se presenta en la Figura 3.2. Consiste en 4 niveles (role, typegen, type, combined). En el caso de *role*, se refiere al tipo de persona que está involucrada en la discusión. *Proponent* si es que tiene la idea central de la discusión (tesis), y oponente quien puede aportar o atacar al *proponent*. El *typegen* especifica si propone la idea central, y los soportes, si atacan o definen la idea central. El *type* especifica el tipo de apoyo, si es normal, o apoya con un ejemplo. Por último, se tiene el último nivel *combined*, que especifica si la función de un segmento es válida sólo en combinación con el de otro segmento.

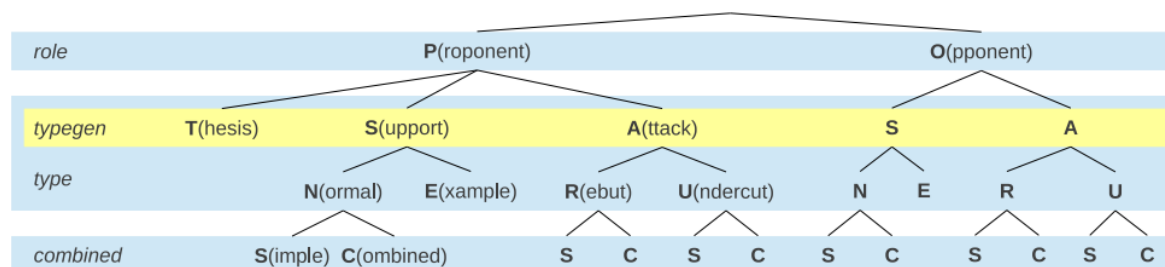


Figura 3.2. Modelo propuesto para organizar los componentes de los argumentos basado en roles de discusión.

Por otro lado, a los textos se le aplicaron algunas características automáticamente con las herramientas OpenNLP-tools, como la tokenización, lematización, etiquetado POS, morfología de las palabras y análisis de dependencias. Usando la salida lingüística y otros recursos, se extrajeron, lemas unigramas, bigramas, etiquetado POS, morfología del verbo principal, triples

dependencias, el valor sentimental de los fragmentos (SentiWS), marcadores de discurso (léxico de marcadores de discurso en Germán), los primeros tres lemas, presencia de marcador de negación.

Por último, se realizó la clasificación con algunos clasificadores normalmente utilizados en estos trabajos relacionados. Estos clasificadores son, Naïve Bayes, Máquinas de Vectores de Soporte, máxima entropía y campos aleatorios condicionales. Algunas herramientas fueron utilizadas en Weka, a excepción de máxima entropía que fue con MaxEnt toolkit y los campos aleatorios condicionales con Mallet. Se realizaron 10 experimentos para encontrar el mejor clasificador con las características mencionadas, sin embargo, los resultados no fueron significativos, es decir, los clasificadores fueron buenos en algunos niveles, pero en otros no.

Tabla 3.3. Resultados obtenidos en distintos niveles de varios clasificadores.

level	Majority			OneR			CRF		
	A	F	K	A	F	n	A	F	n
role	78±1	69±1	0±0	83±3	79±4	33±13	86±5	84±6	49±16
typegen	49±1	33±1	0±0	58±3	47±3	23±7	68±7	67±8	46±12
type	48±1	31±1	0±0	56±3	45±3	22±6	62±7	58±8	38±10
comb	74±1	62±1	0±0	81±4	77±4	44±12	<b>84±5</b>	<b>81±7</b>	55±13
target	24±1	9±1	0±0	37±5	29±4	24±6	47±11	<b>45±11</b>	<b>38±12</b>
role+typegen	47±1	30±1	0±0	56±3	45±3	22±6	67±7	65±8	49±11
role+type	46±1	29±1	0±0	54±3	43±3	21±6	61±7	56±8	38±11
role+type+comb	41±1	24±1	0±0	50±4	38±3	19±6	56±7	51±8	36±9
role+type+comb+target	20±1	7±1	0±0	28±4	19±3	18±5	36±10	30±9	28±10
level	Naïve Bayes			MaxEnt			LibLinear		
	A	F	n	A	F	K	A	F	it
role	84±5	84±5	52±14	<b>86±4</b>	<b>85±5</b>	<b>52±</b>	<b>86±4</b>	84±4	50±14
typegen	<b>74±5</b>	74±5	57±8	70±6	70±6	51±10	71±5	71±5	53±9
type	<b>68±5</b>	67±5	52±8	63±6	62±6	43±9	65±6	62±6	44±9
comb	74±6	75±5	42±11	<b>84±5</b>	<b>81±7</b>	<b>56±12</b>	<b>84±3</b>	<b>81±4</b>	54±10
target	38±6	38±6	29±6	47±8	44±8	37±9	<b>48±5</b>	44±5	<b>38±6</b>
role+typegen	<b>69±6</b>	69±6	55±9	68±7	67±7	51±10	<b>69±5</b>	67±6	52±9
role+type	61±5	61±5	45±7	63±6	<b>61±6</b>	<b>45±9</b>	<b>64±5</b>	60±5	<b>45±8</b>
role+type+comb	53±6	51±6	36±8	58±6	54±7	41±8	61±5	56±5	44±8
role+type+comb+target	22±4	19±4	16±4	36±6	<b>33±6</b>	29±6	<b>39±5</b>	32±4	<b>31±5</b>

En la tabla de arriba se muestra los resultados de los distintos tipos de clasificadores. Los resultados marcadores en negrita indican los mejores resultados para cada clasificador en cada nivel.

### 3.2. Tabla comparativa de los trabajos relacionados

Tabla 3.4. Tabla comparativa de los trabajos relacionados.

Trabajos	Característica	Tipo de identificación	Corpus	Características	Herramientas y/o recursos	Lugar de aplicación	Idioma de aplicación
<b>Identificación de estructuras en el discurso argumentativo de ensayos persuasivos</b> (Stab & Gurevych, 2014b)	-Componentes de argumentos -Relaciones de argumentos	Corpus propio Etiquetado manualmente	-Léxicas -Sintácticas -Marcadores del discurso -Contexto	-Clasificador McNemar -Máquinas de vectores de soporte	-Ensayos persuasivos	Inglés	
<b>Extracción de argumentos y palabras de dominio para la identificación de los componentes argumentativos en textos</b> (Nguyen & Litman, 2015)	-Argumentos completos -Palabras de dominio	Corpus de Stab- Gurevych	-N-gramas -Reglas sintácticas -Señales léxicas -Palabras de dominio léxico	-LDA (Aprender palabras de dominio) -Baseline y derivados	-Temas específicos (Deporte, política, ciencia, etc.)	Inglés	
<b>Extracción de argumentos de noticias</b> (Sardianos et al., 2015)	-Argumentos completos -Componentes de argumentos	Corpus manual anotado	-Palabras de dominio léxico -Posición de la oración -Número de comas -Conjunciones -N-gramas -Etiqueta POS -Diccionario geográfico	-CRF (campos aleatorios condicionales) -word2vec	-Blogs (política, economía, cultura, deportes) -Noticias	Griego	



Tabla 3.4. Tabla comparativa de los trabajos relacionados.

Trabajos	Característica	Tipo de identificación	Corpus	Características	Herramientas y/o recursos	Lugar de aplicación	Idioma de aplicación
<p><b>Hacia el reconocimiento basado en segmentos de la estructura de la argumentación en textos cortos</b> (Peldszus, 2014)</p>	<p>-Argumentos completos -Componentes de argumentos</p>	<p>Corpus anotado manualmente basado en la anotación de Freeman</p>		<p>-OpenNLP tolos -SentiWS -Clasificadores: Naïve Bayes MVS Máxima entropía CRF</p>	<p>-Microtextos de cualquier tema</p>	<p>Alemán</p>	
<p><b>Método propuesto: Identificación automática de argumentos lógicos en discursos políticos escritos en español</b></p>	<p>-Argumentos Completos -Componentes de argumentos</p>	<p>Corpus propio Anotado manualmente</p>	<p>-Signos de puntuación -Contexto -Conjunciones -Marcadores argumentativos -Posición de las proposiciones -Sintácticas</p>	<p>-Freeling -Regex</p>	<p>-Discursos políticos</p>	<p>Español</p>	

### 3.3. Discusión de los trabajos relacionados

En la Tabla 3.4, se consideraron las columnas: tipo de identificación en la minería de argumentos, corpus, características aplicadas para identificar, herramientas o recursos utilizados, en que tipo de texto fue aplicado el trabajo, así como el idioma utilizado tanto de trabajo desarrollado como los corpus. Como se aprecia en la tabla anterior, algunos trabajos se esmeran en identificar las partes de los argumentos y ello implica identificar el argumento completo.

En la columna 2 (corpus), se identifica que trabajos elaboraron su propio corpus y que trabajos se apoyaron de otras fuentes.

En la columna 3 (características) se refieren a las cualidades que se requieren para identificar las partes de los argumentos. Algunos hacen uso de más características que otras, pero observando los resultados obtenidos, se puede concluir que el mecanismo utilizado que aplica estas características influye en la precisión de los resultados. Cada trabajo trata los datos de las características de distintas maneras, para servir como entrada a los clasificadores. Esto refleja de igual forma en los resultados.

En la columna 4 (Herramientas y/o recursos), son todos aquellos métodos o aplicaciones utilizadas para llevar a cabo la identificación de argumentos. Algunos trabajos hacen uso de más clasificadores únicamente para identificar qué clasificador arroja los mejores resultados; como sucede con el trabajo de (Peldszus, 2014).

Por otro lado, se tienen las columnas de “lugar de aplicación” y “lengua”, columna 5 y 6 respectivamente, que señalan el ámbito aplicado en los trabajos.

De los trabajos mencionados, no se tomó ninguna herramienta ni algoritmo para implementarse en esta tesis porque en un principio el objetivo de la tesis era otro

### 3.4. Discusión del capítulo del estado del arte

Los trabajos en el área de la identificación de argumentos automáticos no tienen mucho que se empezó, así que hay pocos trabajos relacionados con ese tema, es por eso que solo se tomaron cuatro trabajos de investigación. Además, estos trabajos tienen características importantes y que son los más relevantes para la identificación de argumentos. Existen más trabajos que tratan sobre la argumentación (por ejemplo, en los *Workshop on Argument Mining*), sin embargo, tienen otros enfoques o que son trabajos previos, por lo tanto, estas investigaciones son la continuación.

Otro dato importante es que los trabajos presentados utilizan modelos matemáticos y estadísticos de la Inteligencia Artificial y el modelo de la tesis es prácticamente sencilla en términos programables, es decir mientras que los otros trabajos intentan encontrar otras características lingüísticas en los diferentes niveles del PLN, la tesis se enfoca prácticamente aún nivel sintáctico y morfológico.

# Capítulo 4

## Método de solución

---

En este capítulo se propone desarrollar un método para la identificación de argumentos, los procesos de este método son: Un proceso que induce patrones para extraer marcadores argumentativos, y un algoritmo recursivo que implementa una expresión regular para identificar argumentos. Previo a estos procesos se requiere realizar el pre-procesamiento de los textos.

## 4.1. Presentación

El método de solución propuesto en este trabajo se muestra en la Figura 4.1. Se describen las fases y módulos de este método:

- Se requiere en primera instancia un módulo para extraer y pre-procesar páginas Web (Actividad 1). Este módulo realiza la descarga de discursos políticos del sitio [www.gob.mx](http://www.gob.mx), además realiza un procesamiento de limpieza, que consiste en la eliminación de caracteres especiales, así como lenguajes de marcación y programación. Como resultado de este módulo se tiene únicamente el texto.
- Fase 1: Inducción de patrones. Esta fase permite encontrar nuevos marcadores lingüísticos de argumentos, esto significa un incremento en el repertorio de palabras claves o introductores para identificar argumentos, cuanto mayor sea el número de marcadores, mayor será la efectividad de la identificación de argumentos (fase 2).
  - Identificar argumentos manualmente (Actividad 2). En esta actividad se buscaron argumentos que fueron utilizados como semillas en la extracción de patrones contextuales. Los argumentos fueron extraídos de los discursos descargados y procesados previamente.
  - Extraer patrones contextuales (Actividad 3). Esta actividad de la fase, identifica patrones contextuales de los marcadores que previamente se identificaron manualmente. Para ello se utilizaron dos herramientas: Freeling una herramienta lingüística, y ETE un marco de trabajo en Python.
  - Aplicar patrones para la búsqueda de marcadores (Actividad 4). Esta actividad requiere como entrada etiquetas Freeling o una expresión regular que puede contener etiquetas Freeling. El procedimiento es la sustitución de las etiquetas por el texto proporcionado para el análisis. Como resultado se tienen marcadores identificados, estos pueden ser nuevos o los mismos.
  - Pruebas (Actividad 5). Esta actividad consiste en verificar si las palabras encontradas representan marcadores de argumentos, es decir, si las palabras encontradas no tienen otra función que no sean introductores de premisas o conclusiones.

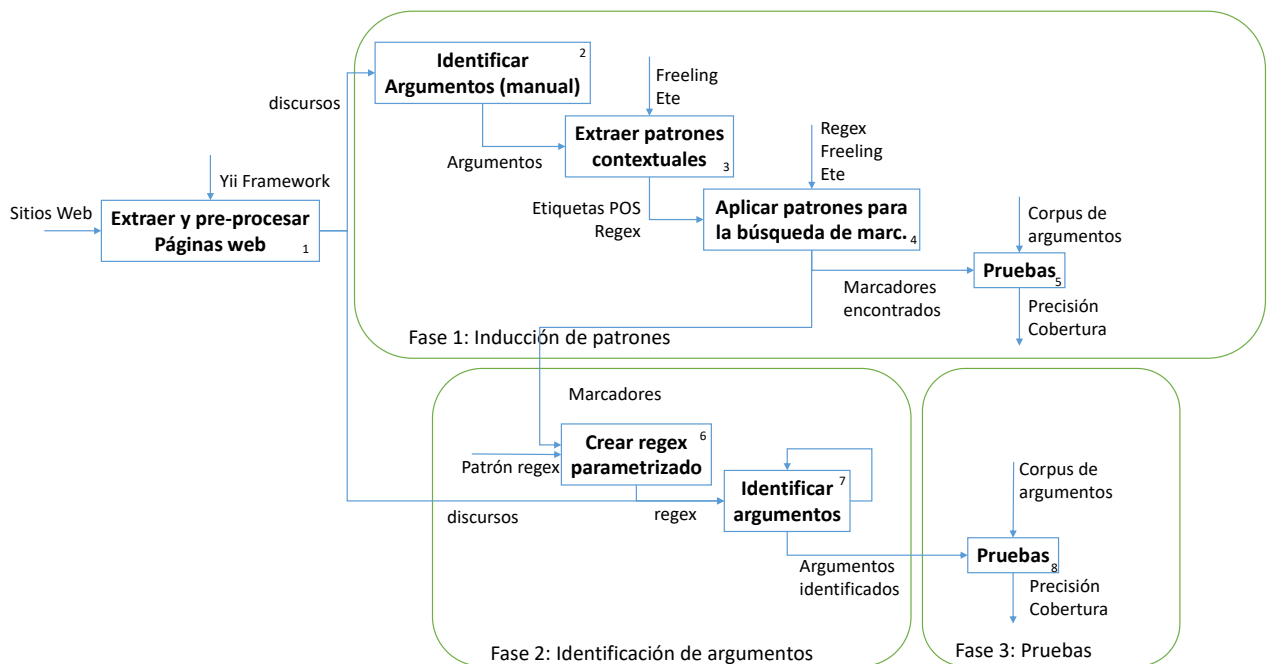


Figura 4.1. Método de solución propuesto en esta tesis

- Fase 2: Identificación de argumentos. Esta fase permite identificar argumentos explícitos, se requiere el repertorio de marcadores argumentativos y una expresión regular prediseñada que permite identificar argumentos, tomando como base los marcadores
  - o Crear regex parametrizada (Actividad 6). Es un modelo de expresión regular que cumple con ciertas reglas proporcionados por (Capaldi, 2011) para identificar argumentos. La expresión regular se considera dinámica porque carga los archivos de marcadores.
  - o Identificar argumentos (Actividad 7). Actividad que consiste en identificar las partes de un argumento (premisas, conclusiones y soportes) así como los sub-argumentos. La utilización de sub-argumentos requiere la utilización de recursividad. Los parámetros requeridos son: la expresión regular y un texto (discurso). Como resultado se tiene argumentos identificados en el texto proporcionado como parámetro.
- Fase 3: Pruebas. Esta fase permite llevar a cabo las pruebas de la fase 2: identificación de argumentos. El corpus utilizado en la actividad 5 es el mismo que se utiliza en la fase 3. El corpus se realizó manualmente, tomado de los discursos descargados en el primer módulo del método de solución.
  - o Pruebas (Actividad 8). La actividad consiste en verificar la efectividad de los argumentos identificados por el método desarrollado. De igual forma verifica el grado de precisión y cobertura de las partes de los argumentos.

## 4.2. Metodología tradicional con respecto a la metodología propuesta en este trabajo

La mayoría de los trabajos relacionados con la identificación de argumentos siguen en general una metodología de procesamiento, en la que es posible distinguir los siguientes puntos:

1. Selección y preparación del corpus: Indica la elección del corpus en el que se va a realizar la identificación de los argumentos, del mismo modo indica la generación del propio corpus. Las consideraciones para trabajar con un corpus son el idioma, el tema de aplicación y el tipo de identificación del argumento, por ejemplo, relaciones de argumentos o componentes de argumentos.
2. Implementación de características lingüísticas: Se establecen un conjunto de características que tienen los argumentos o patrones comunes de los argumentos para utilizarlos en la identificación automática.
3. Selección y prueba de clasificadores: Se toman varios métodos computacionales que utilizan las características lingüísticas como datos de entrada, posteriormente se toman los clasificadores con mejores resultados.

En la Tabla 4.1 se comparan la metodología tradicional con el propuesto.

Tabla 4.1. Comparación de la metodología tradicional con la metodología propuesta.

Procedimiento	Metodología tradicional	Metodología propuesta
<b>Selección y preparación del corpus</b>	Sí	Sí
<b>Implementación de características lingüísticas</b>	Sí	Sí
<b>Selección y prueba de clasificadores</b>	Sí	No
<b>Implementación de un modelo para identificar argumentos.</b>	No	Sí

Como se puede apreciar en la figura anterior existe una diferencia en la forma en que se identifican argumentos, la diferencia se encuentra en las filas “selección y prueba de clasificadores” e “implementación de un modelo” ya que los demás trabajos utilizan algoritmos de aprendizaje

automático mientras que en el método propuesto se desarrolló un modelo de expresión regular que identifica las características lingüísticas más comunes.

### 4.3. Inducción de patrones

En esta sección se presenta un proceso para extraer patrones contextuales de marcadores argumentativos. Los marcadores argumentativos son piezas importantes para identificar argumentos explícitos, es decir, los argumentos hacen uso de palabras claves introducen premisas y conclusiones. Ante esto se desarrolló un proceso llamado “inducción de patrones” que nos permite identificar nuevos marcadores, con lo que se incrementa la efectividad para identificar argumentos.

#### 4.3.1. Incremento de marcadores

(Capaldi,2011) propone marcadores para identificar argumentos. El especialista Capaldi menciona algunos marcadores de premisas y de conclusión. Los marcadores que se mencionan son:

- Conclusiones
  - Por lo tanto
  - Por ende
  - Así
  - De ahí que
  - En consecuencia
  - Por consiguiente
  - Se desprende que
  - Como resultado
  - Llegamos a la conclusión
- Premisas
  - Puesto que
  - Ya que
  - Como

- En tanto que
- Dado que
- Por cuanto
- Viendo que

Además de los marcadores que menciona (Capaldi,2011), existen muchos sinónimos de estas palabras que hacen el papel de marcadores en distintos contextos del texto. Sobre esta proposición se buscaron más marcadores. El resultado de la búsqueda que se hizo en esta tesis es un conjunto de 28 marcadores de premisas y 30 marcadores de conclusión adicionales a los marcadores de (Capaldi,2011).

En la siguiente tabla se aprecian todos los marcadores argumentativos encontrados en los recursos *sensagent*, *interglot*, *thefreedictionary* y *RAE*, así como también los marcadores de (Capaldi,2011); dando una totalidad de 37 marcadores de premisas y 37 marcadores de conclusiones.

Tabla 4.2. Marcadores argumentativos encontrados por algunos recursos.

<b>Marcadores de conclusión</b>	<b>Marcadores de premisas</b>
así es que	dado que
conque	mientras que
de esta manera	mientras
de esta forma	porque
por ello	al
por esa razón	cuando
por eso	cuándo
por tanto	si
de acuerdo con esto	supuesto que
en conformidad	visto que
a cambio de	ya que
en vista de ello	cada vez que
de manera que	siempre que
por lo tanto	como
por ende	puesto que
así	teniendo en cuenta que
de ahí que	ahora que
en consecuencia	tanto
por consiguiente	al igual que
se desprende que	cómo



Marcadores de conclusión	Marcadores de premisas
como resultado	igual que
luego	tan
llegamos a la conclusión	en tanto que
entonces	en el supuesto de que
así pues	en la medida en que
se deriva	en lo que
por lo que	que
así que	pues
de modo que	por cuanto
por ese motivo	viendo que
por tal motivo	como quiera que
por dicho motivo	a partir de
por tal razón	siendo así
por dicha razón	y
por esa causa	sea como fuere
por tal causa	pero
por dicha causa	en efecto

Es posible que, dependiendo del dominio del tema o el tipo del discurso, se utilicen otros marcadores argumentativos. Ante este hecho, se desarrolló el método para extraer marcadores de manera automática, como se verá en la siguiente sección.

### 4.3.2. Método de inducción de patrones

El método de inducción se basa en la siguiente consideración: sea  $T$  un texto representado con etiquetas POS (*Part Of Speech*) o partes de un fragmento (*chunk*) que contiene un marcador de (Capaldi, 2011) ( $M$ ). Considérese ahora un contexto izquierdo  $C_i$  y derecho  $C_d$  de  $M$  en  $T$ ; es decir, la cadena  $C_iMC_d$  ocurre en  $T$ . Si para otro término  $M'$  se cumple que  $C_iM'C_d$  ocurre en  $T$ , diremos que  $M'$  satisface el contexto de  $M$  y, por lo tanto, puede ser un marcador equivalente a  $M$ . En otras palabras, un marcador conocido en el texto, tiene texto en la izquierda y derecha, si otro marcador se acompaña con un texto similar en la izquierda o derecha, entonces es un marcador equivalente al primer marcador. Observe la Figura 4.2, en ella se aprecia el flujo de los pasos para encontrar marcadores diferentes. Este método tomó como base las ideas presentadas por Marti Hearst en la identificación de hipónimos (Hearst, 1992).

El método desarrollado se compone de dos fases, el método para la extracción automática de patrones contextuales mediante el análisis morfosintáctico, y el método para la búsqueda extendida de expresiones regulares con patrones morfosintácticos.

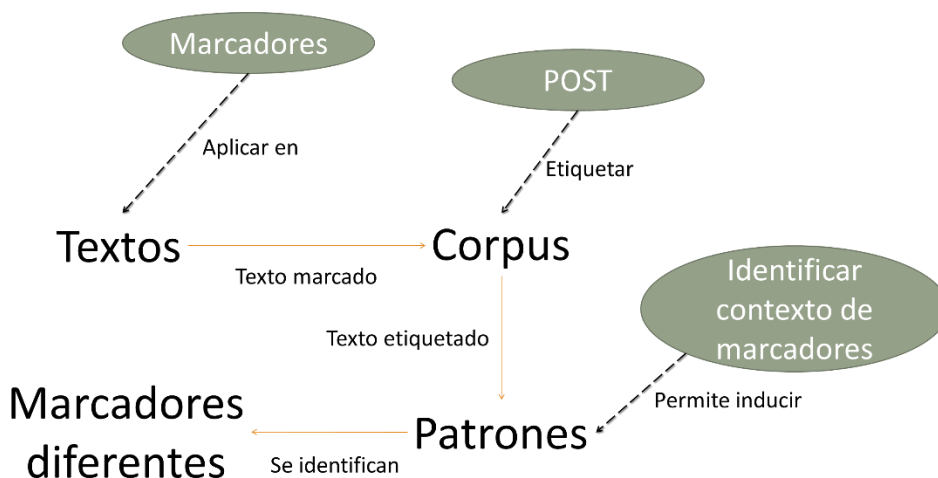


Figura 4.2. Procedimiento del método de inducción de patrones.

El método de extracción de patrones requiere como parámetros un conjunto de textos, y el marcador o marcadores que se quieren analizar. Después del análisis automático se genera una tabla con estadísticas (Figura 4.6) que muestra un conjunto de etiquetas *chunking* y *Eagle*, cada tipo de etiqueta se acompaña de la cantidad de veces que aparece en los contextos de la palabra analizada y también se acompaña del porcentaje con respecto al total de textos analizados.

El método siguiente requiere como entrada una expresión regular, que a su vez implica que en ella se incluyan las palabras claves y la sintaxis de estos. Asimismo, la expresión regular puede estar formada de etiquetas *Eagle* o *chunk*, que se identifican del texto o caracteres especiales, pues cada etiqueta finaliza con guión bajo “\_”. El resultado es una expresión regular con las etiquetas sustituidas por texto.

En las siguientes subsecciones se presenta el funcionamiento detallado de cada proceso.

### 4.3.3. Método de extracción automática de patrones contextuales mediante el análisis morfosintáctico

*Freeling* realiza varios análisis: análisis morfológico, etiquetado POS, análisis de dependencias, análisis sintáctico de poca profundidad, y análisis sintáctico completo. Este último fue el utilizado por ser el análisis más completo y por lo tanto proporciona mayor cantidad de información de análisis. Este análisis codifica la forma de las palabras, así como sus relaciones entre ellas en un conjunto de caracteres con símbolos que representan un significado específico. Este análisis fue el utilizado para identificar los patrones que flanquean a los marcadores.

El análisis sintáctico además de contener análisis morfológico contiene las relaciones entre palabras y oraciones, lo que incrementa la posibilidad de extraer patrones más precisos, hasta llegar a un grado de patrones a nivel de sintagmas u oraciones. Estas relaciones entre palabras, *Freeling* lo llama *chunking tags* (etiquetas de fragmentos). Algunas de las etiquetas que existen son: *a-fp* (Adjetivo, Adjetivo femenino plural), *coord.* (Conjunción, Conjunción de coordinación), *grup-sp* (Preposición, fragmento preposición), entre otras, véase el Anexo B

Un análisis sintáctico completo que realiza *Freeling*, se puede observar en la Figura 4.3.

Los últimos nodos serán llamados hojas, cada hoja contiene una palabra del texto, que no tienen modificaciones y además internamente contienen su lematización (proceso lingüístico que halla el lema de las palabras) y su etiquetado POS. Por ejemplo, si se toma de la Figura 4.3 el primer nodo hoja el valor es “El”, sus datos adicionales son el lema “el” y la etiqueta POS “*DA0MS0*”. Así sucesivamente con cada nodo hoja. Por otro lado, las etiquetas *chunk* son todos los nodos padres de los nodos hojas. Obsérvese el primer nodo hijo “El” de la figura correspondiente, el cual tiene como nodo padre “*j-ms*” y este a su vez tiene “*espec-ms*”, un nodo más arriba es el “*sn*” que es el nodo padre de dos nodos hijos. Esto significa que las palabras “El Estado Mayor Presidencial” están representados por la etiqueta “*sn*” (sintagma nominal).

Se tomó un fragmento textual del discurso “Estado mayor presidencial” del sitio [www.gob.mx](http://www.gob.mx) en el que se menciona lo siguiente (República, 2015a):

“El Estado Mayor Presidencial (EMP) es un órgano técnico militar que tiene como misión fundamental proteger al Presidente de la República. Esta responsabilidad es de la más alta importancia para el País, pues implica el resguardo de la representación del Estado Mexicano y de la República, por consiguiente, de la estabilidad, de la gobernabilidad y de la seguridad nacional”.

En el párrafo citado se observa dos marcadores subrayados “pues” y “por consiguiente”. Obsérvese en la Figura 4.4 el nodo hoja “por consiguiente” representa el introductor de conclusión “por consiguiente”, el contexto izquierdo de esta marcador es la coma (“,”) y el contexto derecho es la preposición “de”, cada uno de estas palabras o signos representan una etiqueta POS (Etiquetas *Eagle*), las etiquetas son “*Fc*” para la coma y “*SPS00*” para la preposición pero las etiquetas POS no se visualizan en la figura 4.4. A nivel sintáctico (Etiquetas *chunk*), el contexto izquierdo es “*Fc*” y derecho “*sp-de*”. Nótese que la etiqueta “*sp-de*” representa el fragmento textual “de la estabilidad,” (nodos hijos).

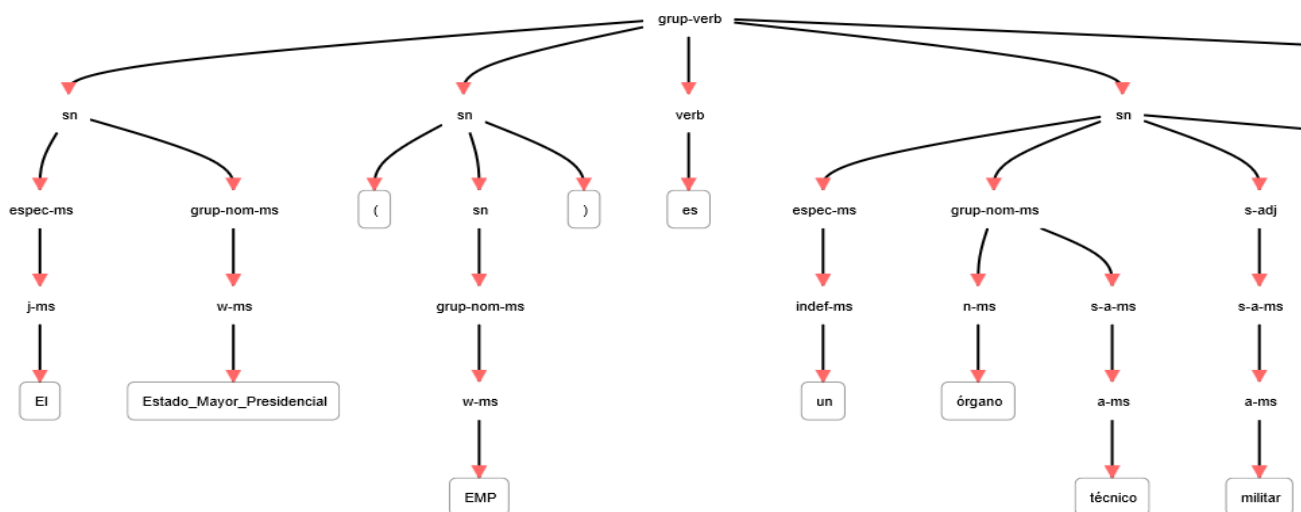


Figura 4.3. Representación parcial de un árbol sintáctico del párrafo citado

No se debe intentar entender los contextos izquierdos y derechos tomando como base los niveles horizontales del árbol. Simplemente saber que las ramas (flechas) que están a la izquierda y derecha del marcador analizado representan los contextos textuales. Para facilitar el entendimiento se empieza analizando desde los nodos hijos y subiendo por los nodos padres.

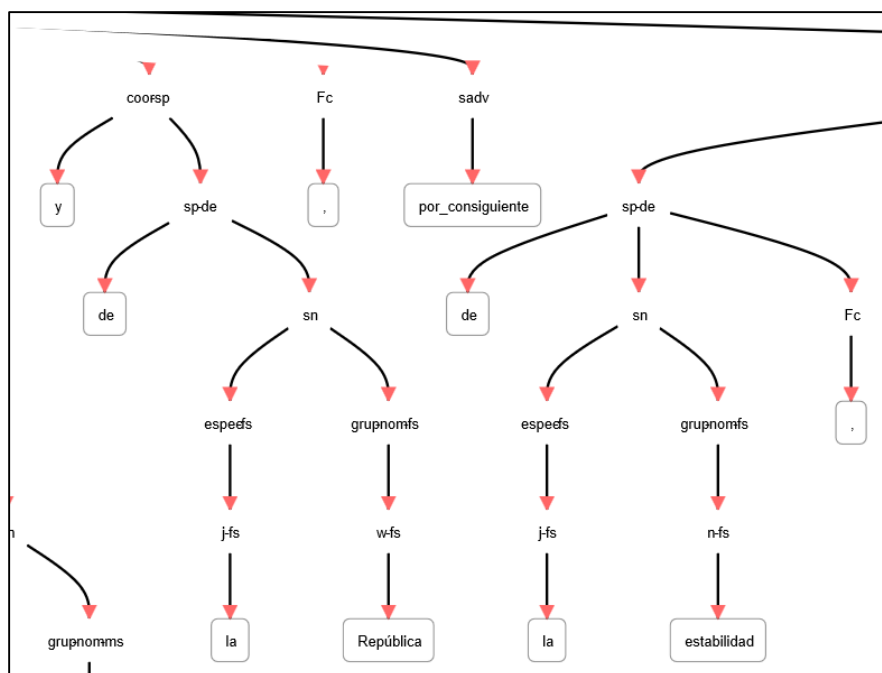


Figura 4.4. Parte de un análisis sintáctico de Freeling 4.0

## Niveles contextuales

Los niveles contextuales son las palabras que flanquean al marcador que se quiere analizar. En la Figura 4.5 se pueden observar los distintos niveles contextuales que tiene un marcador. El método de inducción de patrones tiene la capacidad de extraer distintos niveles contextuales. La extracción de estas palabras es su representación morfológica o sintáctica.

Sin embargo, en la mayoría de las situaciones los contextos de un marcador pueden tener más de una etiqueta *chunk*. Por ejemplo, el contexto izquierdo de segundo nivel del marcador “por consiguiente” de la Figura 4.4 es el nodo hijo “República”. Las etiquetas *chunk* (nodos padres) de este nodo son “*w-fs*”, “*grup-nom-fs*”, “*sn*”, “*sp-de*”, “*coor-sp*”; esto se sabe, porque si traducimos las etiquetas mencionadas a palabras tenemos las siguientes combinaciones:

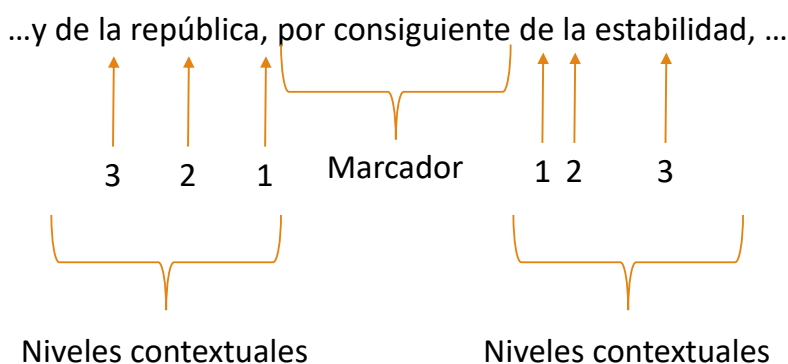


Figura 4.5. Análisis de nivel contextual del marcador "por consiguiente"

*Coor-sp* representa “y de la República” con base en sus nodos hijos, seguido de los nodos “,” y “por consiguiente” ...

Así mismo con las siguientes etiquetas *chunk*:

*Sp-de* representa “de la República”, seguido de “,” y “por consiguiente” ...

*Sn*: representa “la República”, seguido de “,” y “por consiguiente” ...

*Grup-nom-fs* y *w-fs*: representa “República”, seguido de “,” y “por consiguiente”...

Si se analiza el tercer nivel izquierdo, se empieza por el nodo “la”, es decir, la etiqueta “*DA0FS0*”; hasta recorrer el árbol en dirección hacia arriba, llegando al nodo padre.

Sobre este enfoque se desarrolló un método automático que, dada la forma de la palabra, el lema o la etiqueta *Eagle*, y el árbol sintáctico de *Freeling*; proporciona como resultado una tabla de estadísticas de ocurrencias contextuales (Figura 4.6) o una expresión regular, donde se observan los patrones contextuales de las palabras analizadas.

## Patrones contextuales

Los patrones contextuales son un conjunto de etiquetas *Eagle* o *chunk* que se repiten en cada contexto del marcador analizado. Por ejemplo en la Tabla 4.4 se aprecia la extracción de un conjunto de oraciones que son partes de argumentos. Las etiquetas *Eagle* de primer y segundo nivel del marcador “por lo tanto” de la Tabla 4.4 son:

1. (AQ0CSS00 | CC) por lo tanto (VMIP3P0 | NCMS000)
2. (CC | Fc) por lo tanto (VMN0000 | DA0FS0)
3. (CC | Fc) por lo tanto (Fc | RG)

Los patrones contextuales de los tres ejemplos son las etiquetas que más se repiten contextualmente. Por ejemplo, de los tres fragmentos textuales presentados, el *Fc* se repite en dos de los tres fragmentos, en el contexto izquierdo. Entonces por ser muy repetitivo se toma la etiqueta *Fc* como un patrón contextual. Si se analiza el segundo nivel del contexto izquierdo, la etiqueta *CC* se repite en el fragmento dos y tres.

Sin embargo, en el contexto derecho no sucede así, no existe ninguna etiqueta que se repite, pero se puede notar que la etiqueta *VMIP3P0* y *VMN0000* tienen una similitud de izquierda a derecha en los primeros dos caracteres “*VM*”, esto significa que en el segundo nivel contextual en 2 de los 3 fragmentos empieza con un verbo principal (VM). Este último indicio es una característica implementada en este método. Los detalles se verán en la siguiente sección.

Tabla 4.4. Análisis morfológico de oraciones con el marcador "por lo tanto"

	Nivel 4	Nivel 3	Nivel 2	Nivel 1	Marcador	Nivel 1	Nivel 2	Nivel 3	Nivel 4
<b>Fragmento 1</b>	la	incapacidad	temporal	y	por lo tanto	tienen	derecho	a	recibir
<b>Lematizado</b>	el	incapacidad	temporal	y	por_lo_tanto	tener	derecho	a	recibir
<b>Etiquetas</b>	DA0FS0	NCFS000	AQ0CS00	CC	RG	VMIP3P0	NCMS000	SP	VMN0000
<b>Fragmento 2</b>	de	carbón	y	,	por lo tanto	reducir	la	eficiencia	de
<b>Lematizado</b>	de	carbón	y	,	por_lo_tanto	reducir	el	eficiencia	de
<b>Etiquetas</b>	SP	NCMS000	CC	Fc	RG	VMN0000	DA0FS0	NCFS000	DA0FS0
<b>Fragmento 3</b>	el	aire	y	,	por lo tanto	,	también	se	incrementa
<b>Lematizado</b>	el	aire	y	,	por_lo_tanto	,	también	se	incrementar
<b>Etiquetas</b>	DA0MS0	NCMS000	CC	Fc	RG	Fc	RG	PO0CN00	VMIP3S0

## Truncamiento de patrones contextuales

El truncamiento de patrones contextuales consiste en considerar como patrón una parte de las etiquetas. Retomando los fragmentos de la Tabla 4.4, el primer nivel contextual derecho tiene

una similitud en dos fragmentos con las etiquetas *VMIP3P0* y *VMN0000*, la similitud se encuentra en los primeros dos caracteres de las etiquetas “*VM*”, analizando de izquierda a derecha de manera secuencial, es decir no saltarse los caracteres para encontrar coincidencias como sucede con el “0” en el último carácter. Esto significa que el primer nivel contextual de dos fragmentos inicia con un verbo principal (*VM*), por consiguiente, podemos considerar que existe un patrón morfológico con las etiquetas truncadas a esto se le conoce como truncamiento de patrones contextuales.

Esta característica es muy importante porque incrementa la cantidad de patrones en los textos, al principio no se tenía en cuenta y las posibilidades de los contextos se elevaba demasiado. Con esta forma se permitió otra forma de detectar patrones.

El truncamiento de patrones está implementado en este método, en otras palabras, el método propuesto permite al usuario introducir como parámetro el nivel de truncamiento, así como el nivel contextual para extraer patrones contextuales.

### Posición de los patrones

La posición de los patrones es el orden en que están las etiquetas. Éstos permiten identificar solo aquellos textos que poseen un orden secuencial en los patrones identificados, en otros términos, se captura otro tipo de patrón además del tipo de palabra (etiquetas). Este otro tipo de patrón se le conoce como “posición de los patrones”.

Esta característica está relacionada con el orden de aparición de las etiquetas. Tomando como ejemplo los 3 fragmentos de la Tabla 4.4, el patrón contextual de “por lo tanto” en el primer nivel queda de la siguiente forma:

- (*Fc*) por lo tanto (*VM*)

El segundo nivel izquierdo queda de la siguiente forma:

- (*CC* | *Fc*) por lo tanto...

Entonces se aprecia que *CC* está antes que *Fc* y en este caso por ningún motivo *Fc* está antes que *CC*, a esto se le conoce como “posición de los patrones”.

### Tipo de marcador

El tipo de marcador se refiere a cómo están compuestos morfológicamente los marcadores argumentativos. En la Tabla 4.5 se tiene un conjunto de marcadores de conclusión, a estos se les extrajo su representación morfológica representadas en las columnas 2, 3 y 4.. Observe la Tabla 4.5, muchos marcadores tienen una similitud parecida en las etiquetas. Si se trunca las etiquetas la similitud será mayor. De igual forma, la posición de las etiquetas tiene un parecido. Por ejemplo, en la columna “contexto izquierdo” una gran cantidad de marcadores empieza con la etiqueta “*SPS00*”.

Por eso, se consideró una característica más en este método “el tipo morfológico del marcador”.

Tabla 4.5. Representación morfológica de los marcadores de conclusión

<b>Marcadores</b>	<b>Contexto izquierdo</b>	<b>Contexto central</b>	<b>Contexto derecho</b>
<b>Así es que</b>	RG	VSIP3S0	CS
<b>A cambio de</b>	SPS00		
<b>Así pues</b>	RG		
<b>Así que</b>	RG	CS	
<b>Conque</b>	CS		
<b>Como resultado</b>	CS	NCMS000	
<b>De esta manera</b>	SPS00	DDOFS0	NCFS000
<b>De esta forma</b>	SPS00	DDOFS0	NCFS000
<b>De acuerdo con</b>	SPS00		
<b>De manera que</b>	SPS00	NCFS000	PROCNO00
<b>De ahí que</b>	SPS00	RG	CS
<b>De modo que</b>	CS		
<b>En consecuencia</b>	RG		
<b>Entonces</b>	RG		
<b>En conformidad</b>	SPS00	NCFS000	
<b>En vista de ello</b>	SPS00	PDONS000	
<b>Se desprende que</b>	P00CN000	VMIP3S0	CS
<b>Se deriva</b>	P00CN000	VMIP3S0	
<b>Luego</b>	RG		
<b>Llegamos a la conclusión</b>	VMIP1P0	SPS00	DAOFS0
<b>Por consiguiente</b>	RG		
<b>Por lo tanto</b>	RG		
<b>Por ende</b>	RG		
<b>Por lo que</b>	SPS00	DAONS0	PROCNO00
<b>Por ello</b>	SPS00	PDONS000	
<b>Por esa razón</b>	SPS00	DDOFS0	NCFS000
<b>Por eso</b>	SPS00	PDONS000	
<b>Por tanto</b>	SPS00	RG	
<b>Por ese motivo</b>	SPS00	DDOMS0	NCMS000
<b>Por tal motivo</b>	SPS00	DDOCS0	NCMS000
<b>Por dicho motivo</b>	SPS00	VMP00SM	NCMS000
<b>Por tal razón</b>	SPS00	DDOCS0	NCFS000



Marcadores	Contexto izquierdo	Contexto central	Contexto derecho
Por dicha razón	SPS00	VMP00SF	NCFS000
Por esa causa	SPS00	DDOFS0	NCFS000
Por tal causa	SPS00	DDOCS0	NCFS000
Por dicha causa	SPS00	VMP00SF	NCFS000
A consecuencia de	SPS00		
Por consecuencia	RG		

El tipo de marcado permite reducir el margen de error al identificar nuevos marcadores, ya que en las pruebas (se presentan más adelante) con esta característica se eliminan falsos negativos, identificando de este modo solo palabras que cumplen con alguno de los patrones de los marcadores.

## Resultados del método

El método de extracción de patrones permite dos formatos de salida. Por un lado, la generación de una tabla de estadísticas de ocurrencias de etiquetas (Figura 4.6) y por el otro, la generación automática de una expresión regular. Si observa la Figura 4.6, la primera fila indica el total de lemas analizados y encontrados, en este caso se analizó el lema “por lo tanto” de 11 argumentos. En la primera columna se indica cuáles son las etiquetas encontradas. Las etiquetas con guión bajo al final son las etiquetas *chunk* para diferenciarlas de las etiquetas *Eagle*. La columna 2 indica la cantidad de encuentros que hubo en el contexto izquierdo o derecho según sea el caso. Por último, tenemos la columna de porcentaje, que indica en qué grado aparece la etiqueta de un total de 11 argumentos analizados.

Todas las características mencionadas en los puntos anteriores solo funcionan para la salida de la generación automática de expresiones regulares.

Para construir una expresión regular se tomó como base una expresión regular predefinida desarrollada manualmente con base en la observación. La expresión regular predefinida contiene ciertas características, por ejemplo, se sabe que todo signo de puntuación siempre aparece unido a la palabra que lo antecede, así mismo se sabe que todas las palabras van separadas por un espacio.

Lema(s) encontrados 11			
-----Context: right-----			
Tag	Matches	Percent	
coord_	1	9.09091	+
n-fs_	1	9.09091	
dem-fs_	1	9.09091	
NCFS000	1	9.09091	
grup-nom-fs_	1	9.09091	
espec-fs_	1	9.09091	
Fd	1	9.09091	
Fc_	7	63.6364	+
sn_	2	18.1818	+
Fc	8	72.7273	+
F-no-c_	1	9.09091	
DDOFS0	1	9.09091	
-----Context: left-----			
Tag	Matches	Percent	
verb_	3	27.2727	
coord_	1	9.09091	
VMIP3S0	2	18.1818	
Fc_	6	54.5455	+
CC	1	9.09091	+
Fc	7	63.6364	+
grup-verb_	1	9.09091	

Figura 4.6. Salida de estadísticas de las etiquetas de Freeling

El siguiente fragmento de expresión regular es una salida del proceso de extracción automática de patrones contextuales, y sólo se está considerando el contexto derecho de los marcadores analizados. El fragmento fue generado automáticamente, tomando como datos de entrada 20 argumentos con el marcador “por lo tanto” y 5 de “en consecuencia” obtenidos de distintos argumentos:

```
\s((?:RG_) | (?:SPS00_) \s[\wáéíóú]+(?:\s(?:[\wáéíóú]+))?) | (?:CS ) (?:\s(?:[\wáéíóú]+))?) | (?:P00CN000 ) \s[\wáéíóú]+(?:\s(?:[\wáéíóú]+))?) | ( (?:Fc_ ) (?:P00C ) (?:VMIP_ ) (?:DA0F_ ) | (?:Fc_ ) (?:P00C ) (?:VMIF_ ) (?:VMN0 ) | (?:DA0F ) (?:NCFS ) (?:SPS0 ) (?:CS ) ...) [^\w]
```

Se puede notar que el texto sombreado con color gris, es el patrón del tipo de marcador. También se pueden observar las etiquetas truncadas, así como el orden de los patrones. Esta expresión regular tiene 4 niveles contextuales analizados (color verde) encerrados en paréntesis y separados por espacio en blanco, y a menos que sea un signo gramatical no habrá espacios. Cada grupo de patrón está separado por el operador lógico “OR” que en una expresión regular se indica con una barra vertical (“|”). En este ejemplo se aprecian en color verde

los cuatro grupos de patrones, que implican el orden de las etiquetas, el truncamiento, y el tipo de etiqueta.

#### 4.3.4. Proceso para la búsqueda extendida con expresiones regulares de patrones morfosintácticos

Las expresiones regulares son patrones textuales, que permiten realizar una búsqueda a nivel de sintaxis porque son equivalentes a un autómata finito, son también conocidas como regex, regexp o regexen, (López & Romero, 2014).

Sin embargo, en este trabajo se extendió su funcionalidad, a tal grado de poder realizar búsquedas a nivel morfosintáctico. Para llegar a esa funcionalidad se realiza un pre-procesamiento de análisis morfosintáctico (pasos 1 al 5) y posteriormente se aplica los pasos 6 y 7.

A continuación, los pasos que realiza el proceso para la búsqueda extendida con expresiones regulares de patrones morfosintácticos, el cual procede de la siguiente manera:

1. Como dato de entrada se recibe un texto, posteriormente se analiza automáticamente extrayendo el contexto de las palabras analizadas.
2. De manera automática, se obtiene un árbol con la API de *Freeling* y se codifica al formato Newick para su posterior utilización con el framework ETE.
3. Se eliminan etiquetas redundantes automáticamente.
4. Se crea automáticamente una expresión regular con la posibilidad de respetar la posición de las etiquetas *Eagle* o *chunk*.
5. Como otro dato de entrada se proporciona de manera manual otro texto a analizar para encontrar marcadores argumentativos.
6. Se busca la representación textual de las etiquetas en el texto proporcionado de manera automática.
7. Se sustituye automáticamente la representación textual con cada una de las etiquetas que están en la expresión regular.

Al finalizar como dato de salida se obtiene una expresión regular con nuevos valores sustituidos previamente, para ello, se tuvo que apoyar del análisis de *Freeling* y del *framework ETE*.

En el paso 6 se aplica la expresión regular al segundo texto proporcionado, para encontrar los marcadores que respeten el patrón extraído del primer texto.

El paso 4 puede omitirse, entonces se debe colocar una expresión regular propia o la que genera el sistema, por ejemplo, considerando las etiquetas con el mayor porcentaje, esto lo proporciona el método anterior.

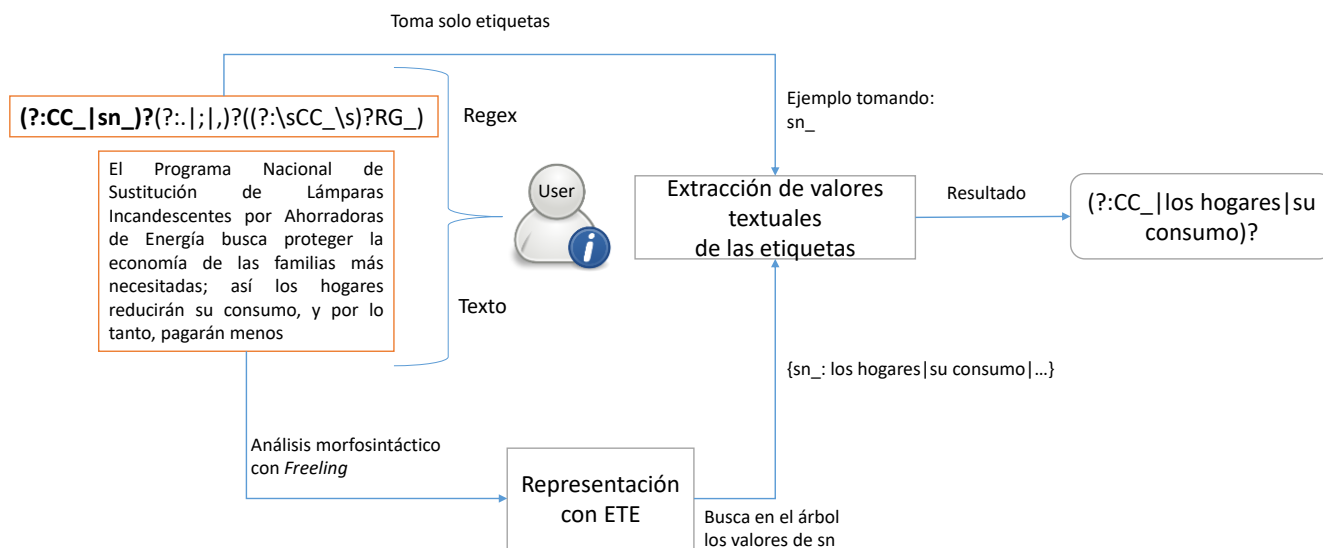


Figura 4.7. Proceso del método que extiende las funcionalidades de una expresión regular con análisis morfosintáctico

En la Figura 4.7 se observa un ejemplo del funcionamiento de este proceso. El usuario requiere dos parámetros, el texto que en este caso es un argumento, y una expresión regular que contenga etiquetas *chunk* o *Eagle*. Las etiquetas tienen que estar con guión bajo al final para diferenciarlas del resto de caracteres. Posteriormente el método toma únicamente las etiquetas y las busca en el texto proporcionado. Para ello el método realiza al texto un análisis morfosintáctico con Freeling y representa el árbol con ETE. Como ejemplo se extrae la etiqueta “sn\_”, que significa “los hogares” y “su consumo”, el método crea una pequeña estructura de expresión regular para que posteriormente sustituya “sn\_” en la expresión regular original con las palabras encontradas en el texto proporcionado.

El motor de expresión regular, así como la estructura de estos no fue modificado, simplemente se incrementó el alcance de búsqueda con el análisis morfológico y sintáctico. Esto nos permite dar un mayor alcance para las expresiones regulares, por mencionar un caso simple: buscar todos los verbos o sustantivos acompañados de un adverbio.

## 4.4. Proceso para la identificación de argumentos

En este capítulo se presenta el proceso para identificar argumentos y un sistema que lo implementa. El resultado es un archivo en formato JSON que representa las partes de los argumentos.

La fase 2 (identificación de argumentos) del método de solución contiene las actividades 6 (crear regex parametrizada) y 7 (identificar argumentos). La actividad 6, necesita dos parámetros de entrada, dos listas de marcadores que contienen premisas y conclusiones como se

observa en la Tabla 4.7, así como un archivo con extensión regex, que contiene una expresión regular predefinida.

Tabla 4.7. Total de marcadores identificados que usa el sistema.

Conclusiones	Premisas
a cambio de	dado que
pues implica	mientras que
conque	mientras
como resultado	porque
de esta manera	cuando
de esta forma	cuándo
de acuerdo con	si
de manera que	supuesto que
de ahí	visto que
de modo que	ya que
en consecuencia	cada vez que
entonces	siempre que
en conformidad	como
en vista de ello	puesto que
se desprende que	teniendo en cuenta que
se deriva	ahora que
luego	tanto
a la conclusión	al igual que
por consiguiente	cómo
por lo tanto	igual que
por ende	tan
por lo que	en tanto que
por ello	en el supuesto de que
por esa razón	en la medida en que
por eso	en lo que
por tanto	que
por ese motivo	pues
por tal motivo	
por dicho motivo	
por tal razón	
por dicha razón	
por esa causa	
por tal causa	

Conclusiones	Premisas
por dicha causa	
quiere decir que	
significa que	
así (?!como)	

#### 4.4.1. Desarrollo de un modelo con expresión regular

En este trabajo se desarrolló un modelo regex que permite identificar argumentos explícitos en discursos políticos, tomando como base central los marcadores argumentativos.

La expresión regular desarrollada toma como principio las siguientes ideas de Capaldi, donde se mencionan ciertas pautas para identificar argumentos en discursos.

A continuación, se detallan las reglas para identificar argumentos:

- No identificar premisas y conclusiones por su contenido.
- No identificar premisas y conclusiones por su posición en el párrafo.
- Identificar argumentos basados en determinadas palabras (marcadores).
- El uso de signos de puntuación y conjunciones indican la existencia de proposiciones que pueden funcionar como premisas y conclusión.
- Una conclusión de un argumento puede ser la premisa de otro argumento.

Para llevar a cabo la implementación de las reglas se tomaron algunos argumentos del libro de (Capaldi,2011) así como argumentos de la página del gobierno ([www.gob.mx](http://www.gob.mx)).

En total fueron 6 argumentos, los cuales, aunque parezcan pocos, en realidad se usaron para realizar cambios en su estructura y así tratar de cubrir todas las posibilidades, por ejemplo, combinaciones de conjunciones con signos de puntuación o el orden de las proposiciones). La razón de esta estrategia es que permite una mayor gama de argumentos de diferentes formas. Otra razón es que buscar otros argumentos que cumplieran con el resto de las reglas de (Capaldi ,2011) llevaría mucho tiempo en encontrarlos. Los cambios que se realizaron en los seis argumentos se hicieron con mucho cuidado.

Básicamente lo que se hizo es mover los componentes de los argumentos, también se cambiaron los marcadores por sinónimos, así como conjunciones y signos de puntuación, todo ello sin perder la esencia del significado del argumento.

La lista de argumentos tomados como base para la generación de la expresión regular, aparecen a continuación:

1. Como el señor Rivera es extranjero, y puesto que los extranjeros no pueden votar en las elecciones norteamericanas, llegamos a la conclusión de que el señor Rivera no puede votar en las próximas elecciones.
2. Por ello, para la adecuada instrumentación de estas acciones ejecutivas, el Gobierno ha solicitado la asesoría y acompañamiento de la OCDE, que agrupa a 34 países de mayor desarrollo en el mundo y que en consecuencia está a la vanguardia en las mejoras prácticas para la buena gobernanza.
3. Todo Estado es una comunidad de algún bien; pues la humanidad siempre actúa con el fin de obtener aquello que considera bueno. Pero, si todas las comunidades aspiran a algún bien, el Estado o comunidad política, que es la más elevada de todas, y que abarca a todo el resto, aspira al bien en medida mayor que cualquier otra, y al bien supremo.
4. Los cupos permiten a las personas físicas usuarias, disminuir los costos de sus importaciones, haciéndolas más baratas y, en consecuencia, más atractivas y competitivas respecto de otras operaciones que no se realizan bajo este esquema.
5. Si mañana queremos tener investigadores y universitarios de excelencia, desde hoy debemos contar con alumnos de excelencia. De ahí que, en el Gobierno de la República, hemos fortalecido los apoyos directos a las instituciones de Educación Superior y a los jóvenes universitarios.
6. Canadá valora la riqueza de los productos del campo mexicano, como el jitomate y el aguacate, de ahí que la balanza comercial agroalimentaria que se reporta de 2015 es superavitaria para nuestro país en 300 millones de dólares.

Cada uno de los argumentos presentados, fue utilizado para probar la expresión regular (modelo), en caso de que no se identifique alguno de los argumentos, se decide modificar la expresión regular para poder identificarlo. Una vez que se consiguió identificarlos a todos, el siguiente paso fue modificarlos, pero conservando la estructura original, es decir conservando los argumentos originales. Las modificaciones fueron realizados con base en las reglas de (Capaldi, 2011), después se analizaba el nuevo resultado con la expresión regular, en caso de no identificar alguno, se ajustaba el modelo para identificarlo; sin embargo, el modelo se modificaba de tal forma que aún se podían identificar los argumentos originales, este procedimiento fue realizado varias veces hasta asegurarse de que sean realizado todas las combinaciones de las reglas, como consecuencia se tuvo una colección de argumentos modificados tomando como base los 6 primeros presentados anteriormente.

Con base en las reglas de (Capaldi,2011), y pautas proporcionados por él, se diseñó una gráfica (Figura 4.8) de las reglas para identificar argumentos en discursos.

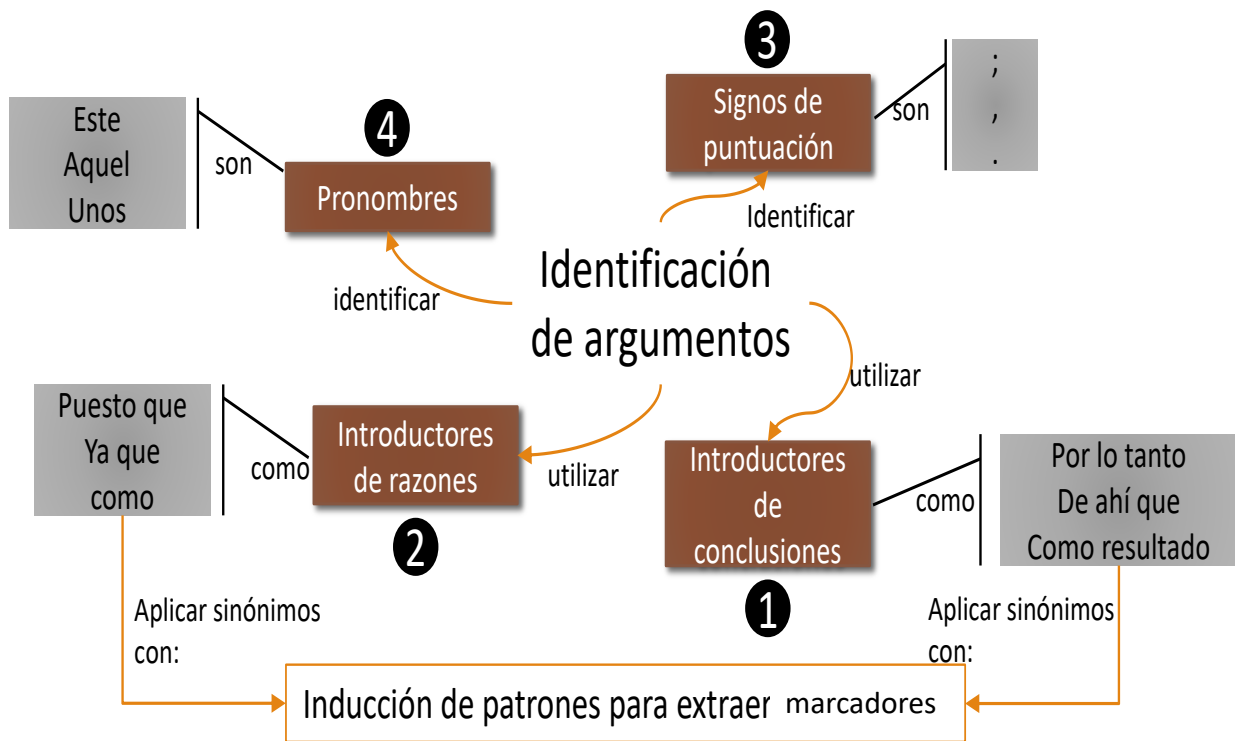


Figura 4.8. Diagrama de las reglas de Capaldi y como se implementa la inducción de patrones.

Como resultado de la implementación de las ideas de Capaldi se tiene la siguiente expresión regular en diagrama de ferrocarril (Avallone, 2011) mostrada en la Figura 4.9.

En el diagrama, se observan 10 grupos, estos grupos capturan la información identificado por ellos. Cada grupo captura información diferente, dependiendo de la estructura de los argumentos, pero no todos los grupos se activan al mismo tiempo, la mayoría de ellos permanecen sin información capturada.

A continuación, se presenta el papel que desempeña cada grupo.

- Grupo #1: Captura un marcador de premisa si comienza al principio del argumento.
- Grupo #2: Captura la proposición-premisa que cuenta con un marcador de premisa (Grupo #1).
- Grupo #3: Captura la proposición-premisa que comienza al principio de un argumento y sin contar con un marcador de premisa.
- Grupo #4: Captura el marcador de conclusión.
- Grupo #5: Captura la proposición-conclusión que cuenta con un marcador de conclusión (Grupo #4)



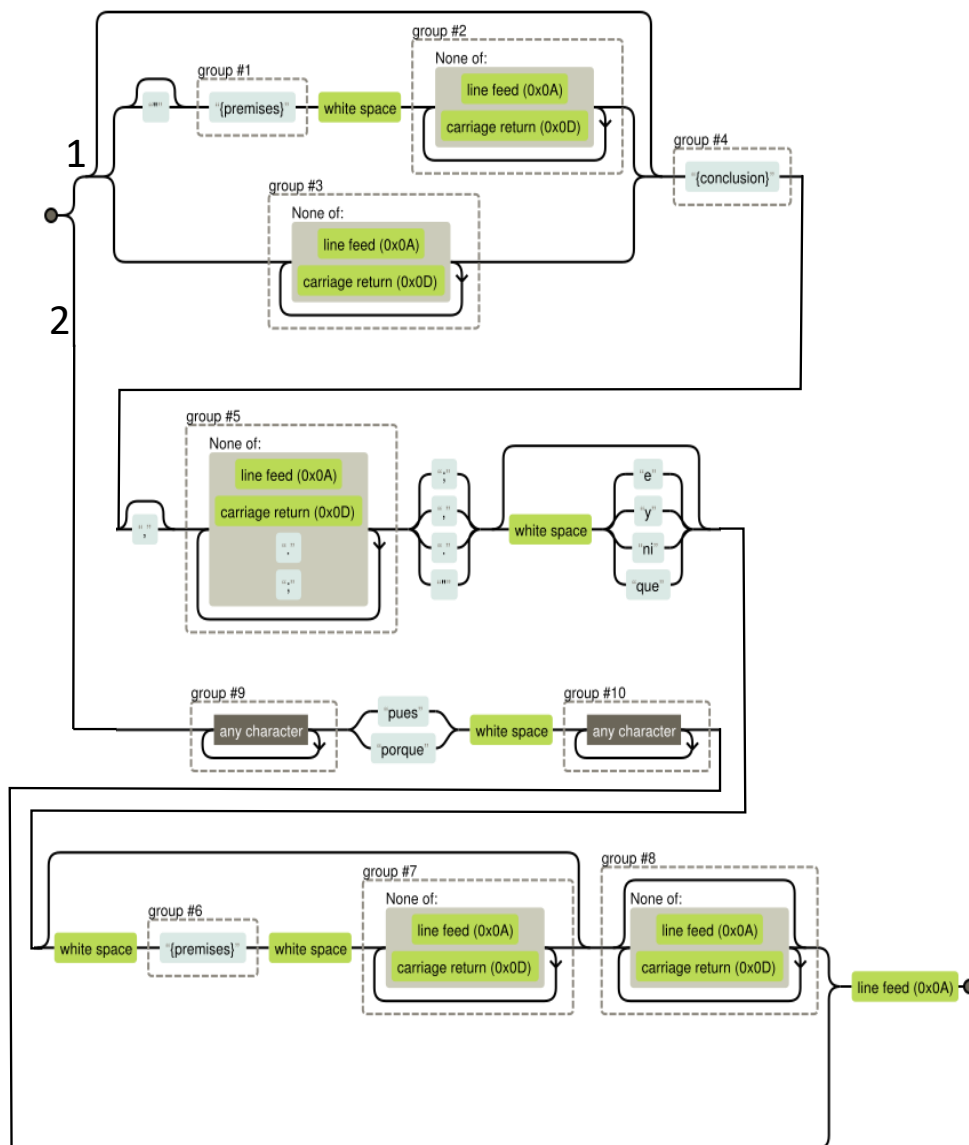


Figura 4.9. Un modelo de expresión regular para identificar argumentos

- Grupo #6: Captura un marcador de premisa que está después de una proposición-conclusión (Grupo #5).
- Grupo #7: Captura la proposición-premisa que está después de un marcador de pre-misa (Grupo #6)
- Grupo #8: Captura el resto del argumento, que pueden indicar un soporte positivo o negativo al argumento, por ejemplo, oraciones que contengan las palabras “sin embargo”, “pero”, “además”.

- Grupo #9: Captura la proposición-conclusión que no cuenta con un marcador de conclusión.
- Grupo #10: Captura la proposición-premisa que cuenta con un marcador especial (premisa y conclusión al mismo tiempo).

#### 4.4.2. Análisis del modelo de la expresión regular

En esta subsección se da un análisis de la expresión regular respecto a las relaciones entre los grupos de captura.

Los grupos 1 y 2, se activan al mismo tiempo. Lo que indica que un argumento puede empezar con una premisa o se activa el grupo 3 en caso de que no exista marcador de premisa. Sin embargo, los argumentos pueden no comenzar con premisas, así que el grupo 4 y 5 se ejecutan en primera instancia para capturar la conclusión al comienzo del argumento. Posteriormente, entra en funcionamiento el grupo 6 y 7 para capturar premisas cuando se encuentren después de una conclusión, aunque también se puede dar el caso en que no existan argumentos con premisas después de la conclusión, por lo tanto, es opcional la activación del grupo 6 y 7. Cabe mencionar, que los grupos 1-2-3 y los grupos 6-7 no pueden activarse juntos, esto significa que forzosamente se requiere una conclusión en medio del argumento (grupo 4 y 5). Además, el grupo 8 se activa siempre y cuando exista alguna parte adicional del argumento.

El diagrama presentado en la figura 4.9 contiene al inicio desde el punto negro dos caminos por recorrer, lo cual indica que se puede tomar cualquiera de los dos caminos, esto dependerá de la estructura del argumento. El párrafo anterior describe el primer camino de la expresión regular.

Asimismo, el otro camino activa los grupos 9 y 10 al mismo tiempo. Esto indica que el grupo 9 es una conclusión sin marcador y el grupo 10 es la premisa o un conjunto de premisas.

El fragmento de abajo está compuesto de varias líneas, la primera y la segunda línea representan el argumento analizado, las demás líneas representan las capturas de cada grupo de la expresión regular. En ese argumento se encuentra un marcador especial “**pues**” señalado en negritas, y este marcador indica la existencia de una conclusión antes del marcador y después una o más premisas. Como se puede observar el grupo 9 tiene capturado “Todo Estado es una comunidad de algún bien;” y el resto se encuentra en el grupo 10. Puede comprobar el funcionamiento de la expresión con este argumento desde el sitio <https://regex101.com/#python>. El cual permite la depuración paso a paso de la expresión regular.

**Encuentro #0 Tamaño: 344 Rango: 496-839**

Todo Estado es una comunidad de algún bien; **pues** la humanidad siempre actúa con el fin de obtener aquello que considera bueno. Pero, si todas las comunidades aspiran a algún bien, el Estado o comunidad política, que es la más elevada de todas, y que abarca a todo el resto, aspira al bien en medida mayor que cualquier otra, y al bien supremo.

**Grupo #1 Tamaño: 0**

**Grupo #2 Tamaño: 0**

Grupo #3	Tamaño: 0
Grupo #4	Tamaño: 0
Grupo #5	Tamaño: 0
Grupo #6	Tamaño: 0
Grupo #7	Tamaño: 0
Grupo #8	Tamaño: 0
Grupo #9	Tamaño: 44

Todo Estado es una comunidad de algún bien;

Grupo #10	Tamaño: 295
-----------	-------------

la humanidad siempre actúa con el fin de obtener aquello que considera bueno. Pero, si todas las comunidades aspiran a algún bien, el Estado o comunidad política, que es la más elevada de todas, y que abarca a todo el resto, aspira al bien en medida mayor que cualquier otra, y al bien supremo.

### 4.4.3. Recursividad de la expresión regular

La expresión regular propuesta, es la que tuvo el mejor resultado en los experimentos (Tabla 5.4). Se realizaron otros experimentos con pequeñas diferencias de expresiones regulares, pero no se tuvieron los resultados esperados. En la sección “ejecución de experimentos” se indica cuáles fueron esos cambios.

La expresión regular para ejecutarse requiere leerse desde un archivo de texto, luego ser parametrizada (actividad 6 del método de solución). Se prosigue con la compilación, que quiere decir, crear el objeto regex y cargarse en memoria para su posterior uso.

Posteriormente en la actividad 7, se toma como valor de entrada el objeto regex y se cargan los discursos pre-procesados. El pre-procesamiento consiste en eliminar caracteres que no tienen nada que ver con los discursos. En la Figura 4.10 se observa el proceso general. Identificar todo el argumento es detectar si todo el párrafo es un argumento; para ello, se comprueba si hay un marcador de conclusión o especial, en caso de haberlo se identifica el párrafo como argumento. Al instante se registran las partes del argumento; es decir, sus premisas, conclusiones y si existen soportes.

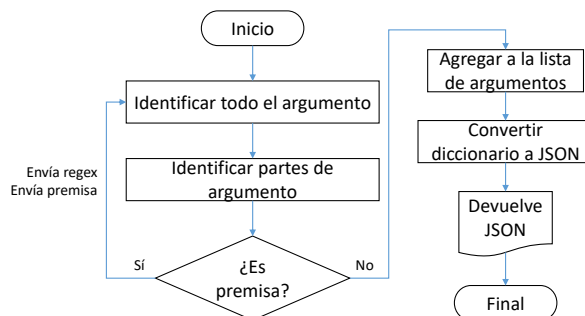


Figura 4.10. Diagrama de flujo del proceso de identificación de argumentos.

Si la parte identificada es una premisa entonces se vuelve a ejecutar el método “identificar argumentos”, enviando el regex compilado y las premisas; a esto se le conoce como recursividad. Se puede enviar al mismo tiempo un conjunto de premisas o premisas separadas, esto dependerá de la forma del argumento. Por ejemplo, si la conclusión se encuentra en medio de un párrafo, entonces lo que está a la derecha e izquierda de la conclusión son conjuntos de premisas, y cada conjunto de premisas pueden tener varias proposiciones, en este caso se envían los dos conjuntos de premisas por separado.

```
{
  "position": [
    1800,
    2152
  ],
  "argument": "México tiene en su juventud la mayor opor
próximas décadas. Por ello, en el Gobierno de la Repúb
propongan. De ahí la importancia del Diálogo sobre la
del país.",
  "subargument": [
    {
      "position": [
        0,
        234
      ],
      "argument": "México tiene en su juventud la ma
las próximas décadas. Por ello, en el Gobierno
como se lo propongan. ",
      "subargument": [],
      "parts": {
        "marcon": "Por ello",
        "position": [
          0,
          121
        ],
      },
      "conclusion": " en el Gobierno de la Repúb
propongan",
      "premises2": "México tiene en su juventud
durante las próximas décadas. ",
      "supportpre": " "
    }
  ]
}
```

Figura 4.11. Argumentos identificados en formato JSON.

De otra manera, si no es premisa o ya no existen argumentos qué identificar dentro del conjunto de premisas entonces el argumento se agrega a la lista de argumentos, luego se realiza la conversión a JSON y se retorna el valor. El valor de retorno es un archivo JSON con los siguientes atributos (Figura 4.11):

- Posición del argumento
- Argumento completo
- Lista de sub-argumentos que contiene los mismos atributos que se están mencionando.
- Las partes del argumento:
  - El marcador de conclusión

- Posición de la conclusión
- El cuerpo de la conclusión
- Los marcadores de premisas
- Los cuerpos de las premisas o soportes

Las expresiones regulares buscan las mejores coincidencias posibles que satisfacen al patrón, esto es, busca todos los caminos posibles que coincidan completamente con la expresión regular (Microsoft, 2016), por lo tanto, el motor de expresión regular siempre buscará el primer argumento padre, en otras palabras encontrará el marcador que se encuentra en el argumento principal. Esto es importante, ya que de este modo se puede identificar perfectamente el orden en que van apareciendo los argumentos.

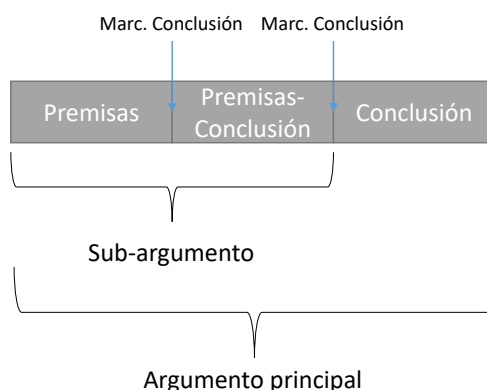


Figura 4.12. Ejemplo de estructura de argumentos interrelacionados o sorites.

Observe el diseño (Figura 4.12), en el cual existen dos marcadores de conclusión; divididos con una flecha azul. El segundo marcador, esto es, el marcador de la derecha es el primero que identifica el motor de expresión regular, por lo tanto, divide el argumento principal en premisas y conclusión. Se aplica la recursividad y se toma los dos bloques de la izquierda “premisas” y “premisas-conclusión”. Al encontrar en las premisas otro marcador de conclusión, el sistema lo cataloga como sub-argumento; así sucesivamente, el método tomará de nuevo las premisas del sub-argumento para encontrar otro sub-argumento del primero. Así es la funcionalidad de los motores de expresiones regulares, de modo que es una ventaja para identificar argumentos interrelacionados.

## 4.5. Discusión

La idea de desarrollar el método de inducción de patrones se debe a la importancia que se le dio a los marcadores argumentativos en este trabajo, porque el modelo de expresión regular desarrollado para identificar argumentos tiene como principal característica y como comienzo del procedimiento de búsqueda, identificar palabras claves que son introductores de argumentos de premisas y conclusiones.

El método de inducción de patrones se compone de varios sub-métodos que tienen funcionalidades diferentes pero que se complementan mutuamente. El método de extracción automática de patrones contextuales sufrió varios cambios en el transcurso de las pruebas ya que no se obtenían buenos resultados, así que se implementaron varias características más del análisis contextual, como la posición de las etiquetas.

Por otro lado, el método para la identificación de argumentos se empezó a desarrollar con expresiones regulares ya que resultaba fácil identificar marcadores argumentativos, además que se tenía un conjunto de marcadores de premisas y conclusiones, así fue como se decidió utilizar expresiones regulares para las siguientes etapas de este trabajo, así como analizar para utilizar otras características lingüísticas de los argumentos.

Gran parte del trabajo se desarrolló a partir de las reglas de Capaldi, por el simple motivo que está enfocado a discursos.

# Capítulo 5

## Resultados

---

En este capítulo se presenta la sección de rendimiento de las expresiones regulares ya que la solución propuesta está basada sobre éste modelo. Del mismo modo se presentan los experimentos de la fase 1 y 2 del método de solución de esta tesis. Para la ejecución de los experimentos se tuvo que realizar un corpus etiquetado manualmente de argumentos.

## 5.1. Rendimiento de las expresiones regulares

En esta sección se presenta el rendimiento de las expresiones regulares al momento de encontrar los patrones en el texto. El motivo de la existencia de esta sección es que esencialmente todo el trabajo se basa en las expresiones regulares, y por lo tanto es importante analizar el impacto de este tipo de modelo. Además si en las expresiones regulares, no se implementan bien pueden darse problemas en el motor de las expresiones regulares en el cual el análisis regresa muchas veces a un estado anterior para analizar los otros posibles casos en la búsqueda de texto; a este problema se le conoce como retroceso catastrófico (Goyvaerts, 2016). Este problema puede surgir porque se obtiene una expresión regular a partir de un texto ya analizado morfosintácticamente, y nos da una expresión regular extensa, por lo que se incrementan los pasos del motor.

Para las pruebas de testeo y rendimiento se usaron dos herramientas web, el <https://regex101.com/#python> y <http://pythex.org/>. La página regex101 consiste en una colección de herramientas, con depuración, pruebas de unidad, entre otras; la página pythex consiste en aplicar solo la expresión regular, pero se puede obtener el tiempo estimado que se tarda en ejecutar una expresión regular.

Un ejemplo del problema del retroceso catastrófico sucede en esta expresión, el cual fue generada por el proceso de inducción de patrones:

```
(grup-verb_|\.|CC_|coord_|Fc_)\s?(RG_)\s?(Fc_|espec-ms_|sn_|coord-vb_|pos-ms_|verb_|morfema-verbal_|indef-ms_|dem-ms_)
```

Esta prueba se realizó con todos los patrones de “por lo tanto” de 10 argumentos, en el primer nivel y considerando etiquetas *Eagle* y *chunking*. Se aplicó estos patrones extraídos por los mismos 10 argumentos. Como resultado se obtuvo el análisis con un total de 22382 pasos. Sin embargo, analizando más argumentos (10 argumentos con el marcador “en consecuencia”), las herramientas presentaron el problema de retroceso catastrófico. El análisis llevó aproximadamente 2 segundos con análisis de pasos y depuración. Sin realizar el testeo llevó aproximadamente .6085s ejecutándolo cuatro veces. En conclusión, si el texto a analizar incrementa en caracteres, los valores de las etiquetas también incrementarían. En todas nuestras pruebas solo se presentó este problema introduciendo etiquetas *chunking*.

A partir de entonces, todas las demás pruebas se hicieron sobre etiquetas *Eagle*. Ya que no presentaron problemas de este tipo, y las ejecuciones se realizaron en menos de un segundo con cantidad de textos variados, así como expresiones regulares diferentes.

## 5.2. Corpus etiquetado manualmente

Para llevar a cabo las pruebas del método para identificar argumentos se tuvo que crear un corpus de argumentos explícitos. Lo siguiente explica la estrategia tomada para realizar la colección de argumentos:



1. Se extrajeron 1200 discursos y comunicados de prensa del sitio [www.gob.mx](http://www.gob.mx) del gobierno de México.
2. Se realizó el pre-procesamiento de los datos, para eliminar cualquier carácter ajeno al texto de nuestro interés (discursos). Para ello se utilizó varias librerías en PHP para la manipulación de páginas web.
3. De los 1200 discursos se tomaron 200 discursos de manera aleatoria.
4. Posteriormente fueron analizados manualmente discurso por discurso, intentando localizar las palabras claves que funcionan como marcadores argumentativos, para ello se tomó la lista de marcadores presentados en la Tabla 4.2.
5. Una vez localizados se identificaron 78 fragmentos textuales que hacen uso de marcadores argumentativos. Los fragmentos textuales son párrafos en los discursos políticos.
6. Luego, se leyeron los fragmentos detenidamente para verificar si son argumentos y además comprobar los componentes de los argumentos, como resultado se obtuvo 61 argumentos identificados.
7. En seguida, se detectaron 62 marcadores de conclusión en los 61 argumentos identificados.

En la tabla siguiente se pueden apreciar los 62 marcadores identificados de los 200 discursos, así como también las palabras que no son marcadores. Cabe mencionar que un fragmento textual puede contener más de un argumento, asimismo un argumento puede contener más de un marcador de conclusión. Por ejemplo, un argumento puede tener la siguiente sintaxis: “Por lo tanto... porque...”. Observe la sintaxis y entre las palabras “por lo tanto” y “porque” se encontraría la conclusión, esto quiere decir que la conclusión hace uso de dos marcadores: uno a la izquierda y otro a la derecha.

Tabla 5.1. Conteo de marcadores en los argumentos identificados manualmente.

<b>Palabras</b>	<b># De Marc.</b>	<b># De No Marc.</b>
Por lo tanto	19	
Como resultado	5	1
así	11	36
Por ello	4	1
De ahí	6	
En consecuencia	9	
porque	3	

Palabras	# De Marc.	# De No Marc.
pues implica	1	
Por consiguiente	1	
se llegó a la conclusión	1	
significa que	1	
quiere decir que	1	
Total:	62	38

En la Figura 5.1 se puede observar un segmento de los argumentos identificados manualmente limitados por colores, es decir, cada color representa la conclusión de un argumento o todo el argumento en el mismo párrafo. Por ejemplo, en el segundo párrafo el color rojo indica un argumento completo simple; en cambio en el primer párrafo el color rojo indica la conclusión del argumento, pero que sus premisas son el color amarillo y naranja. Ahora el color naranja es la conclusión de un subargumento, y sus premisas son aquellas que están en color amarillo. Por último, tenemos el color amarillo que indica un subargumento del subargumento.

En la Figura 5.2 se desglosa los argumentos de la Figura 5.1 para visualizar el resultado final.

“La inflación incide directamente en el bienestar de las familias, ya que mide la variación de los precios de una canasta de bienes y servicios representativa del consumo de los hogares mexicanos. Es decir, una inflación alta significa que los precios se están elevando considerablemente y, por lo tanto, el ingreso de las familias alcanza para menos. Por el contrario, como ahora ocurre: una inflación baja quiere decir que los precios de los principales productos y servicios que adquieren los mexicanos, prácticamente, no están aumentando..

En el sector industrial estamos conscientes del trascendente papel que debe desempeñar la iniciativa privada, para transformar a México en una sociedad y economía del conocimiento. Y por lo tanto, tenemos claro que debemos redoblar esfuerzos para sumarnos a la dinámica que ha mostrado su Administración, para fortalecer el desarrollo científico y tecnológico de nuestro país..

Figura 5.1. Ejemplo de etiquetado de argumentos por colores.

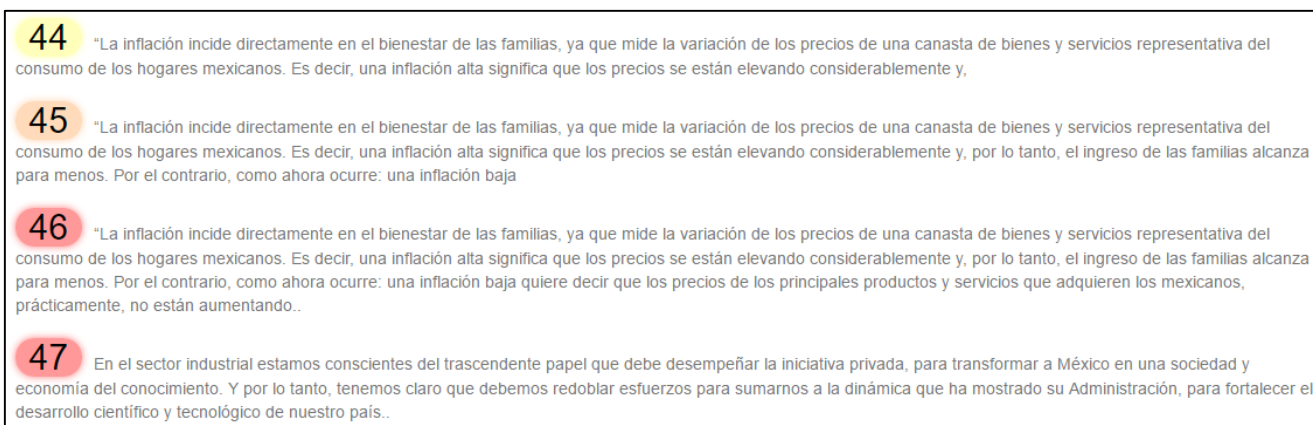


Figura 5.2. Desglose de la lista de argumentos. Los colores indican un argumento de un subargumento.

Es así como se construyó el etiquetado para representar los sorites y los argumentos simples.

### 5.3. Ejecución de experimentos

Se realizaron dos tipos de pruebas, el primero es el proceso de inducción de patrones, y el segundo el método de identificación de argumentos.

Se realizaron 15 experimentos con el proceso de inducción de patrones, cada una de ellas se aplicaron con características diferentes; por ejemplo, los niveles de análisis, el truncamiento, entre otras.

Se usaron 20 argumentos para las pruebas que contienen el marcador “en consecuencia” y 10 argumentos con el marcador “por lo tanto”. Después de realizar los experimentos algunos tuvieron buenos resultados, así que se prosiguió incrementar los argumentos para realizar un experimento final.

Por otro lado, se encuentran las pruebas del método para identificar argumentos. El número total de experimentos en estas pruebas fueron 3, aplicando diferentes características para mejorar la precisión.

#### 5.3.1. Ejecución de experimentos del método inducción de patrones

Dado que Freeling maneja una precisión aproximada del 97%(Padró et al., 2010), se eligió para realizar el etiquetado POS. De igual manera, cabe señalar que la precisión que arrojan los métodos desarrollados dependerá en parte de Freeling.

Para analizar la efectividad de los procesos, se realizó una consulta manual de 200 discursos de la presidencia de la República y secretarías federales, de los cuales 29 contienen marcadores conocidos y de ellos se extrajeron manualmente varios argumentos con diferentes

marcadores de conclusión sugeridos por (Capaldi,2011). En la siguiente tabla (tabla 2) se indican los marcadores extraídos:

Tabla 5.2. Marcadores argumentativos identificados en los discursos políticos consultados

	MARC. ARG.	NO MARC. ARG.
<b>COMO RESULTADO</b>	3	3
<b>DE AHÍ</b>	5	1
<b>EN CONSECUENCIA</b>	9	-
<b>IMPLICA</b>	1	-
<b>SE LLEGÓ</b>	1	-
<b>POR CONSIGUIENTE</b>	1	-
<b>POR LO TANTO</b>	17	2
<b>ASÍ</b>	9	37
<b>POR ELLO</b>	6	1
<b>TOTAL</b>	52	44

De la tabla anterior se observa que los marcadores más utilizados son “por lo tanto”, “en consecuencia” y “así”. Dado que este último presenta mucha ambigüedad porque puede funcionar como adverbio, conjunción, interjección, entre otras; se decidió tomar los dos primeros para hacer la generación automática de expresiones regulares. En total se logró generar automáticamente 16 expresiones regulares (16 pruebas), de las cuales fueron dos las que lograron identificar la mayor cantidad de palabras que sí cumplen la función de marcadores.

En el Anexo C se encuentra una expresión regular generada automáticamente, éste tuvo el mejor resultado de todos los experimentos. Se tomó como datos de entrada 20 argumentos con el marcador “por lo tanto” y 5 de “en consecuencia” obtenidos de distintos discursos.

La expresión regular mostrada en el Anexo C contiene etiquetas POS truncadas para considerar una mayor posibilidad de encuentros, ya que además de que se extrajeron únicamente los patrones POS también se consideró la posición, por ejemplo: el contexto de un argumento en particular con el marcador “por lo tanto” contiene las etiquetas *PP1, VMI, Fc*; y se está respetando la posición de estos. Es decir, primero tiene que estar la etiqueta *PP1*, seguida de la etiqueta *VMI* y posteriormente la etiqueta *Fc*, así también con los demás argumentos analizados.

Un ejemplo de la expresión regular mostrada es la identificación del marcador de conclusión “de ahí” del siguiente argumento: “El diagnóstico de la pobreza y el diálogo de la pobreza se vuelve cada vez más complicado y de ahí la importancia en que permanentemente midamos para ver cómo estamos evolucionando” (Social, 2015).

El contexto derecho del marcador “de ahí” es: “la importancia en que” que en términos morfosintácticos es: *DA0F* → la, *NCFS* → importancia, *SPS0* → en, *CS* → que. Así la

expresión regular identificó un nuevo marcador tomando como base los marcadores “por lo tanto” y “en consecuencia”

Tabla 5.3. Resultados de los marcadores argumentativos encontrados con el método de inducción.

<b>Inducción de patrones (Reglas-aplicadas)</b>	<b>Precisión</b>	<b>Cobertura</b>	<b>F1</b>
Experimento 1	56%	25%	34.5%
Experimento 2	92%	25%	39%

Se realizaron varios experimentos. Dos de ellos se encuentran en la Tabla 5.3 porque fueron los que tuvieron la mejor precisión. El experimento uno consiste en las siguientes características: posición de los patrones, que se refiere al orden de las etiquetas POS; contexto izquierdo y derecho (3 niveles), se refiere a que sólo se tomaron 3 palabras a la izquierda y a la derecha; patrones de “por lo tanto” y “en consecuencia”, son los marcadores analizados; y truncado a 2, se refiere a que no se consideró toda la anotación morfosintáctica. El experimento dos, sólo se consideró el contexto derecho (4 niveles); es decir cuatro palabras a la derecha; truncado a cuatro, es decir los cuatro primeros caracteres de las etiquetas; posición de los patrones, el orden de las etiquetas fueron tomados en cuenta; tipo de marcador, se refiere al tipo de información morfológica, por ejemplo, marcadores de tipo adverbio; y patrones de “por lo tanto” y “en consecuencia”, que son los marcadores analizados.

### Discusión de los experimentos

En los primeros experimentos no se obtuvieron los resultados esperados. Después se agregaron varias características de manipulación al método, como la posición y el truncamiento. Estas características permitieron reforzar la precisión de los patrones. Porque además del patrón de la etiqueta, se agregó el patrón de las posiciones, esto permitió un modelo de expresión regular más riguroso que llevó de nuevo a su reducción en la precisión, debido a la cantidad limitada de argumentos utilizados en la extracción de patrones. Sin embargo, al aplicar la característica de truncamiento de las etiquetas, permitió aumentar la precisión de los patrones.

Por otro lado, la cobertura se incrementó en los últimos experimentos, pero permaneció constante. La razón probable puede deberse a la cantidad de argumentos analizados para extraer sus patrones, así como el análisis de dos tipos de marcadores.

#### 5.3.2. Ejecución de experimentos del método identificación de argumentos

El objetivo de esta tesis es identificar los argumentos explícitos, sin embargo, se presenta además una característica adicional, el método desarrollado permite identificar a un buen grado los componentes de un argumento. Es por eso que en el último experimento se tienen los datos de efectividad considerando las partes identificadas de los argumentos. No obstante, si se consideran las partes de los argumentos, la precisión y la cobertura es inferior a la identificación de todo el argumento.

El modelo de expresión regular identifica el argumento completo y sus partes por lo cual se tienen dos resultados: Las medidas de rendimiento para identificar argumentos y los componentes de los argumentos.

En esta sección se realizaron tres experimentos para la identificación de argumentos, de los cuales solo el último alcanzó un grado alto de medida deseable.

Cada experimento tiene una expresión regular distinta, sin embargo, todas tienen una gran cantidad de paréntesis, la razón detrás de esto es que los paréntesis permiten agrupar mejor las partes de los argumentos o permiten introducir condicionales como “|” que simboliza el “o”, es decir, si es “X” o “Y”. Pero no todos los paréntesis que agrupan texto son capturados por el motor de la expresión regular. Los paréntesis que contienen un signo de interrogación seguido de dos puntos (:) indican que el contenido no será captura, sin embargo, si existe otro paréntesis dentro de ellos y no contiene el signo de interrogación y los dos puntos, entonces el texto será capturado.

La lista de marcadores de premisas y conclusiones de cada experimento fueron sustituidos en las variables {premisas} y {conclusiones} respectivamente. Para que la expresión regular localice esos marcadores.

Las características de cada experimento se enlistan a continuación:

- Experimento 1:
  - Se utilizaron 61 argumentos identificados manualmente.
  - Se agregan los marcadores a la expresión regular presentados en la Tabla 4.7, a excepción de la última fila de la columna uno. En este experimento solo estaba la palabra “así”.
  - La expresión regular tomada en este experimento es la siguiente:

```
(?: (?: (?: "?( {premisas} )\s ( [^\n\r]+ ) | ( [^\n\r]+ ) )? ( {conclusiones} ),? ( [^\n\r;]+ ) (?: ; | , | \. | " ) (?: \s (?: e | y | ni | que) )? (?: \s ( {premisas} )\s ( [^\n\r]+ ) )? | ( .+ )\s (?: pues | porque )\s ( .+ )
```

- Experimento 2:
  - Se utilizaron 61 argumentos identificados manualmente.
  - Se agregan los marcadores a la expresión regular presentados en la Tabla 4.7, a excepción de la última fila de la columna uno. Es decir, no se consideró el marcador “así” por motivos de ambigüedad.
  - Se siguió considerando el mismo patrón de la expresión regular del experimento 1, con la diferencia en la cantidad de marcadores utilizados.
- Experimento 3:
  - Se utilizaron 61 argumentos identificados manualmente.
  - Se agregan los marcadores a la expresión regular presentados en la Tabla 4.7, sin omitir ningún marcador.
  - Se hicieron cambios en la expresión regular, quedando del siguiente modo:

```
(?:(?:("?(premisas)\s)([\n\r]+)|([\n\r]+))?(conclusiones),?([\n\r\.;]+)(?:;|,|\.|")(?:\s(?:e|y|ni|que))?(?:\s(premitas)\s([\n\r]+))?(.*)|(.)?(?:pues|porque)\s(.+)
```

Las dos expresiones regulares indicadas en los experimentos 1 y 3, tienen pequeñas diferencias pero que realizan grandes cambios a la hora de identificar las partes de los argumentos. Estas diferencias están marcadas en color gris, en ellos se pueden apreciar las partes agregadas o eliminadas con respecto a una expresión regular de la otra. De esta manera pudimos aumentar la efectividad de los resultados.

En la tabla siguiente se aprecian los tres tipos de experimentos realizados con sus resultados.

Tabla 5.4. Tabla de resultados de 3 experimentos.

PRUEBAS	PRECISIÓN	COBERTURA	F1
EXPERIMENTO 1	56%	83%	66.87%
EXPERIMENTO 2	90%	79%	84.1%
EXPERIMENTO 3- SIN PARTES	90%	98%	93.8%
EXPERIMENTO 3- CON PARTES	87%	95%	90.8%

En el experimento 3, se hicieron dos cambios muy importantes. La primera reside en que la lista de marcadores es posible introducir expresiones regulares, como en el caso del marcador “así (¿!como)”, esto permite anular palabras contextuales de los marcadores, y así evitar un grado alto de ambigüedad. Además, como se aprecia en la expresión regular del experimento 3, se realizaron algunas modificaciones permitiendo de este modo, capturar si el argumento posee algún soporte negativo o positivo. El cambio aumentó la precisión para identificar las partes del argumento, así como el argumento completo.

### Discusión de los experimentos

En resumen, el cambio de la expresión regular y la capacidad de introducir expresión regular en cada marcador permitió una considerable mejora.

De todos los experimentos, el experimento 3 de la Tabla 5.4 tuvo el mejor resultado; por ello se analizó también los resultados métricos de los componentes de los argumentos. No obstante, los resultados fueron un poco más bajos.

Si observa los modelos propuestos de expresiones regulares notará pequeños cambios, pero que estos cambios mejoraron los resultados.

## 5.4. Discusión

Durante las pruebas se presentaron varios inconvenientes, como la velocidad. La lentitud o el retroceso catastrófico de las expresiones regulares causaban problemas por el gran tamaño

de los patrones. Por tal motivo se realizaron cambios a los métodos, esto permitió resolver los problemas mencionados y al mismo tiempo se mejoraron los resultados con las métricas utilizadas para medir el desempeño de los métodos propuestos.

Las métricas utilizadas son las medidas más utilizadas para el reconocimiento de patrones y recuperación de información, aunque en la tabla de resultados en el método de inducción se tuvo una cobertura baja, esto solo indica que hicieron falta más datos de análisis para extraer nuevos patrones y combinarlos con los ya existentes. Se cree que utilizando varios marcadores para extraer patrones se enriquece la expresión regular, y por consiguiente se tendrá una mayor cobertura.

Se encontraron nuevos marcadores que no estaban en la lista de marcadores. Así que se agregaron a la lista. De igual forma se tomaron en cuenta para realizar el corpus.

El corpus se realizó minuciosamente con todos los conocimientos investigados para identificar argumentos. Sin embargo, en el corpus solo están identificados argumentos lo suficientemente explícitos que usan marcadores. Porque incluso identificar argumentos implícitos es un tema que nadie ha cubierto, puesto que no se encontraron trabajos en los *workshop* y en otros lados.



# Capítulo 6

## Contribuciones y trabajos futuros

---

En este capítulo se presenta la implementación del método de solución en un sistema web para el uso común, así como también las aportaciones de éste trabajo de tesis, y posibles mejoras que se puedan realizar.

## 6.1. Presentación

Del método de solución propuesto para identificar argumentos, se utilizó la segunda fase para implementar en un sistema que permitirá identificar argumentos en los textos, sobre todo en los discursos políticos escritos.

Este sistema desarrollado tiene una estructura relativamente diferente al método de solución. El usuario interactúa con cualquier sitio Web y puede interactuar con la extensión de Chrome por si quiere identificar argumentos en dicha página.

Una extensión Web es un pequeño programa que incorporan nuevas funciones al navegador para personalizar así la experiencia del usuario.

Algunas razones por las cuales se implementó una extensión en el navegador se debe a que es más fácil para el usuario hacer uso de una extensión que de una aplicación, sobre todo porque la interacción se hace directamente desde una página web, asimismo la instalación es más fácil que una aplicación de escritorio como el uso de un servicio web.

Tabla 6.1. Resumen de objetivos alcanzados.

<b>Objetivos</b>	<b>Logros</b>
<i>Crear un corpus de argumentos sobre discursos políticos etiquetados manualmente</i>	Se extrajeron 1200 discursos políticos automáticamente desde la página <a href="http://www.gob.mx">www.gob.mx</a> , de los cuales 200 se tomaron aleatoriamente. Después siguiendo las reglas para identificar argumentos con marcadores, se leyeron los discursos para identificar los argumentos que hacen uso de estos marcadores. Posteriormente, se etiquetaron todos los argumentos detectados.
<i>Desarrollar el método de inducción de patrones para identificar más marcadores de argumentos</i>	Se pensó que el uso de marcadores sigue un patrón o conjunto de patrones, así que se tomó como base la idea de identificar el contexto izquierdo y derecho de cada marcador, para ello se analizó morfosintácticamente con la herramienta Freeling las palabras que flanquean a los marcadores, como resultado se tuvieron etiquetas que representan ese análisis. Posteriormente se juntaron las etiquetas para buscar coincidencias, la tarea se hizo al principio de forma manual y luego automático. El resultado son los patrones contextuales. Además de ello se tomó en consideración algunas características, presentadas en la sección 4.3.3.

<b>Objetivos</b>	<b>Logros</b>
<i>Desarrollar un algoritmo recursivo para identificar argumentos explícitos</i>	Se desarrolló un modelo de expresiones regulares que cumple con las reglas de Capaldi para identificar argumentos en discursos políticos. La realización de esta tarea fue hecho paso a paso de forma manual, es decir la primera instancia de expresión regular fue considerando identificar los marcadores e identificar todo el párrafo; realizado este paso se continuó con identificar proposiciones con signos de puntuación y conjunciones, luego identificar qué tipo de proposición es, y por último reutilizar el modelo para identificar otros argumentos dentro de otros argumentos.
<i>Desarrollar una herramienta final para hacer uso del método para identificar argumentos</i>	La implementación del algoritmo de identificación de argumentos, se hizo en dos partes, la primera parte fue realizar un sistema del lado del cliente y la segunda del lado del servidor, es decir, requiere de un servidor para su funcionamiento. Así mismo el sistema del cliente es una extensión del navegador Chrome, que trae como consecuencia facilitar el uso y la instalación.

Por otro lado, se presentan las conclusiones de toda la investigación, así como las posibles mejoras que se pueden implementar.

## 6.2. Lecciones aprendidas

A lo largo de esta investigación de tesis hubo varios retos que enfrentar, así como tomar decisiones en el alcance de la tesis.

Al principio de la investigación el nombre de la tesis era: “Método para la identificación automática de falacias por apelación de las emociones en noticias escritas en español”. Pero se cambió de noticias a discursos políticos, debido a la poca cantidad de argumentos falaces que se encontraban en las noticias, esta conclusión se llegó porque los argumentos intentan convencer. Las falacias son argumentos que intentan convencer apoyándose de premisas, que no apoyan lo suficiente a la conclusión (Rodríguez Jiménez, 2013), en otras palabras el razonamiento es débil porque carece de un buen sustento entre las premisas y conclusiones. Normalmente en las noticias por su naturaleza no convencen sino informan. Existen decenas de tipos de falacias, cada uno con una particularidad diferente que radica en el modo o la forma de justificarse, pero se pueden clasificar en dos grandes grupos, falacias formales e informales (Vega Reñón, 2012).

De este modo, se continuó con la investigación con discursos políticos. Más tarde se investigaba la forma de identificar falacias de manera manual que apelen a las emociones. Los resultados que se obtuvieron fue la propuesta de un método para identificar si un argumento es falaz, para ello en la propuesta inicial de la tesis se tenía que identificar primero si existe un argumento y luego verificar si el argumento es falaz. Así que el método que se propuso consistía en dos fases, por un lado, la fase para identificar argumentos y por el otro lado la fase para identificar falacias.

En las presentaciones de avances que se realizaron, los revisores como los asesores llegaron a un acuerdo y propusieron omitir la segunda fase y concentrarse en la primera, debido a que cada fase eran trabajos grandes que requerirían de tiempo para ser desarrollados, además que son áreas relativamente nuevas en el campo de la computación. Por tal motivo en esta tesis solo se encuentra el método para la identificación de argumentos. El método propuesto en esta tesis solo puede identificar argumentos explícitos, es decir, que hacen uso de marcadores argumentativos; esta restricción se debe a que el objetivo principal era identificar falacias y por consiguiente la primera fase no se llegó a realizar un estudio profundo de los trabajos relacionados para realizar la identificación de los argumentos implícitos.

La fase dos hace uso de la semántica que es el cuarto nivel del Procesamiento de Lenguaje Natural, esto implica tener un cierto conocimiento del mundo para poder identificar argumentos. Las técnicas que se investigaron para desarrollar la fase dos se enlistan a continuación:

- Uso de un léxico afectivo creado por (BACA GOMEZ et al., 2014). Esto con el fin de poner una polaridad emocional a las proposiciones de los argumentos (premisas y conclusiones). El resultado sería un grado emocional que varía desde 0 a 1.
- Identificar la existencia de implicación textual de las premisas a la conclusión, usando un diccionario Tesauro y el analizador *Freeling*. El resultado obtenido serían cadenas léxicas. Las cadenas léxicas son relaciones de cohesión a través del texto, lo cual indica una continuidad discursiva de la información que viaja en el texto (Martínez, 2002).
- Un método de control difuso que requiera como entrada el grado emocional de las proposiciones y las cadenas léxicas. Usando la herramienta *scikit-fuzzy* nos daría como resultado el grado de justificación del argumento; es decir, si las premisas apoyan suficientemente a la conclusión se consideraría un buen argumento, de lo contrario sería considerado un argumento falaz o falacia.

Como se aprecia estos puntos requieren desarrollarse y ejecutarse para realizar las pruebas, esto conlleva un tiempo adicional para su implementación. Además, no se encontró ningún trabajo sobre falacias informales, por lo tanto, los buenos resultados que se esperarían obtener de la propuesta de la fase dos sería una hipótesis. Por esta razón se decidió concentrarse en la identificación de argumentos para finalizar a tiempo la tesis.

El método para identificar argumentos es también un trabajo relativamente nuevo (Mining, 2016) y por ende no existen numerosas investigaciones sobre ello. El primer reto que nos enfrentamos fue descubrir cuál es la primera característica más importante en un argumento

y con base en eso, identificar la ubicación del argumento en el discurso político. El segundo mayor reto es identificar la dimensión del argumento en un texto y el tercer reto como identificar sus partes. Después de varias lecturas se encontró el libro “Como ganar un discurso” de (Capaldi, 2011), el cual mencionaba características importantes para identificar argumentos en discursos.

Uno de los mayores retos que hubo en la realización de la tesis, fue entender correctamente los trabajos relacionados debido a que es un tema nuevo para nosotros y entender en que parte de la lógica se encuentra la argumentación. Lo mejor era centrarse en el marco conceptual para entender perfectamente los términos que se usan en la tesis y en los trabajos relacionados para que así se pueda realizar una mejor investigación. Al no realizar un buen entendimiento de los conceptos, algunos trabajos que más se relacionan fueron encontrados posteriormente a la propuesta de solución.

Por otra parte, el estado del arte tuvo que cambiarse ya que se tenían trabajos investigados sobre falacias o trabajos sobre discursos políticos. El cambio que hubo provocó una demora grande en tiempo, y se tuvo que empezar de nuevo en la búsqueda. Sin embargo, gracias a los *workshops* que salieron a finales de 2014 sobre la argumentación, estos ayudaron a encontrar trabajos relacionados con la tesis. Es importante mencionar, para que el problema del cambio del estado del arte no se presente, se debe realizar un buen estudio sobre la magnitud de complejidad de la tesis.

Por último, las pruebas fueron un gran reto porque se requiere de un buen conocimiento para identificar argumentos en los textos. Así que el proceso de identificar argumentos fue lento, porque no era seguro si lo que se encontraba era un argumento. Cada uno de los posibles argumentos que se identificaron se analizaban a detalle para corroborar si en verdad era un argumento, de serlo así se identificaba sus partes como sus dimensiones. Este proceso llevó más de un mes. El corpus de los trabajos del estado del arte de esta tesis, se apoyaron de expertos o conocedores del tema de la argumentación para tener resultados concisos y seguros. Sin el apoyo adecuado de expertos en la argumentación que ayuden a detectar argumentos, la tesis se hizo más tediosa.

### 6.3. Contribuciones

Al inicio de la investigación no se tenían los conocimientos necesarios para identificar argumentos, debido a que identificarlos manualmente es una tarea complicada. Fue así como se dio la tarea de encontrar ciertas pautas que ayudaran a su identificación. Como fuentes principales para tratar los argumentos se consideraron los libros “Como ganar un discurso” y las “Claves de la argumentación” de los autores (Capaldi, 2011) y (Weston, 2011) respectivamente. En estos libros se mencionan la posibilidad de identificar argumentos y la construcción de los mismos. Estos dos puntos son importantes para el método de solución que se propuso.

Una problemática que se presentó es la ausencia de herramientas y recursos que tratan sobre este tema. No se encontraron muchos trabajos que se enfoquen en la argumentación y de los pocos que se encontraron se aplicaban para otros idiomas (griego, alemán e inglés).

En las subsecciones siguientes se comentará de los productos generados y las aportaciones de la tesis.

### 6.3.1. Productos

El método de solución está hecho en su mayor parte de expresiones regulares, estos modelos permiten hacer un análisis sintáctico y morfológico a los textos. Para llegar a un nivel de análisis morfosintáctico *se desarrollaron varios procedimientos que permiten unir las capacidades de búsqueda de sintaxis de las expresiones regulares con el poder de análisis de Freeling.*

Por lo demás, en esta tesis se proporciona *una herramienta al usuario que le ayude a la identificación de los argumentos* para un mejor análisis del tema en cuestión, esta herramienta implementa una extensión del navegador que consulta a un servicio web. Con la obtención de los argumentos que es el primer paso en la minería de argumentos de la Inteligencia Artificial, se pueden procesar su información y realizar diferentes análisis como la recuperación de información, comprensión del lenguaje, respuestas a preguntas, detectar contradicciones, identificar argumentos falaces de cualquier tipo, entre muchas más.

### 6.3.2. Aportaciones

La construcción de argumentos, así como la identificación son tareas muy complejas de desarrollar automáticamente, incluso para el ser humano. La única forma de poder localizar a estos argumentos es con la práctica. Por lo tanto, en este trabajo se desarrolló *un modelo generado a base de reglas para identificar argumentos en discursos*, que estén bien estructurados y que usen marcadores argumentativos. El cual es un nuevo recurso basado en expresión regular.

Por otro lado, el propósito principal de este trabajo es identificar el argumento completo en un párrafo. No obstante, *las expresiones regulares y las reglas de (Capaldi, 2011), permitieron identificar sus componentes argumentativos*; es decir, las partes del argumento como las premisas, la conclusión, los soportes y los argumentos interrelacionados.

Como resultado de este trabajo de tesis se tiene la *generación manual de un corpus listo para su procesamiento electrónico de argumentos de la temática de discursos políticos* (únicamente para el idioma español), con un total de 61 argumentos más 25 argumentos utilizados para la extracción de patrones en las pruebas.

Además se construyó *una lista de marcadores argumentativos de premisas y conclusiones* extraído de los recursos: sensagent, interglot, thefreedictionary y (Calsamiglia & Tusón, 1999), por supuesto el método de inducción de patrones permitió encontrar otros que no se encontraron en estos recursos.

El método de “inducción de patrones” se puede utilizar para cualquier palabra(s) en la que se quiera obtener sus patrones contextuales. Ya sea que esta palabra se proporcione en forma de lema, etiqueta POS o la(s) palabra(s) como tal.

## 6.4. Trabajos futuros

Los modelos propuestos cumplen con algunos requisitos para identificar argumentos o marcadores argumentativos:

- Signos de puntuación.
- Conjunción.
- Posición de las proposiciones (premisas y conclusiones).
- Contexto
- Características morfosintácticas

Sin embargo, aún existen otros requisitos que no se lograron cubrir por falta de tiempo. Como *identificar la dimensión de un argumento más allá de un párrafo*. La regla principal para cubrir este requisito es identificar pronombres o verbos en segunda y tercera persona que se encuentren en la primera oración de un párrafo. Este requisito se pretende considerar como un trabajo futuro.

El método que identifica argumentos también identifica sus componentes. Pero a pesar de ello, *no identifica el tipo de soporte*, en otras palabras, si cierta oración apoya o refuta al argumento. Para resolver el problema de los soportes se pueden considerar otros tipos de marcadores argumentativos que indican el tipo de soporte. Por ejemplo, palabras como “sin embargo”, “no obstante”, “pero” indican soportes negativos; palabras como “además”, “así como”, “asimismo” indican soportes positivos o que apoyan al argumento. Esta característica peculiar en los argumentos también se puede implementar más adelante.

Una de las principales dificultades que se presentó en esta tesis fue evitar la ambigüedad de ciertos marcadores, como sucede con el “así”. El desarrollo del *método de inducción* permitió identificar hasta cierto grado esta ambigüedad. Sin embargo, este método trabaja de forma independiente al método para identificar argumentos, de hecho, el método para identificar argumentos puede trabajar sin depender del método de inducción. Otro trabajo que se puede hacer, es *la combinación de estos dos métodos* para reducir la ambigüedad.

# Referencias

---

- Alberich, M. (2007). *Procesamiento del Lenguaje Natural - Guía Introductoria*. Retrieved from [http://www.sopadebits.com/content/file\\_download/cms\\_file/file/58/pln-1.0-20070630.pdf](http://www.sopadebits.com/content/file_download/cms_file/file/58/pln-1.0-20070630.pdf)
- Alberto, A. (2008). argumentación. Retrieved February 23, 2015, from <http://mimosa.pntic.mec.es/ajuan3/lengua/argument.htm>
- Alcina, A., & Valero, E. (2009). *Terminología y sociedad del conocimiento* (Peter Lang).
- Arcentin, G. (1993). Impacto persuasivo del gesto en el discurso político: una experiencia de consejo (asesoría) y de laboratorio. *Revista Latinoamericana de Psicología*, 25, 61. Retrieved from [http://panamericano.edu.gt/colegio2012/archivos/SC-COMU021/el\\_impacto\\_de\\_discurso.pdf](http://panamericano.edu.gt/colegio2012/archivos/SC-COMU021/el_impacto_de_discurso.pdf)
- Avallone, J. (n.d.). Regexper. Retrieved April 11, 2016, from [www.regexper.com](http://www.regexper.com)
- BACA GOMEZ, Y. R., Castro Sánchez, N. A., Martínez, A., Farias, H., Irazu, D., & Rosso, P. (2014). Impacto de la ironía en la minería de opiniones basada en un léxico afectivo. In *CEUR Workshop Proceedings* (Vol. 1199, pp. 41–44).
- Cabrio, E., Hirst, G., Villata, S., & Wyner, A. (2016). Natural Language Argumentation: Mining, Processing, and Reasoning over Textual Arguments. Retrieved June 7, 2016, from <http://www.dagstuhl.de/16161>
- Calsamiglia, H., & Tusón, A. (1999). *Las cosas del decir: Manual del análisis del discurso*. Las cosas del decir. Manual de análisis del discurso.pdf
- Capaldi, N. (2011). *Como ganar una discusión* (Tercera). Retrieved from <http://es.scribd.com/doc/102500113/Como-Ganar-Una-Discusion#scribd>
- Deitel, P. J. (2004). *Cómo programar en Java*. Pearson Educación.
- Díaz, Á. (2002). *La argumentación escrita*. Universidad de Antioquia.
- Fabra, U. P. (2012, June 5). El Procesamiento del Lenguaje Natural en la Recuperación de Información Textual y áreas afines. Universitat Pompeu Fabra.
- Freeman, J. B. (1991). Dialectics and the macrostructure of arguments: A theory of argument structure (Vol. 10). Walter de Gruyter.
- Gelbukh, A., & Sidorov, G. (2010). *Procesamiento automático del español con enfoque en recursos léxicos grandes* (2nd ed.).
- Goldstein, L. (2008). *Lógica: Conceptos clave en Filosofía*. Universitat de València.
- Goudas, T., Louizos, C., Petasis, G., & Karkaletsis, V. (2014). Argument extraction from news,



- blogs, and social media. In *Artificial Intelligence: Methods and Applications* (pp. 287–299). Springer.
- Groarke, L. (1996). Informal Logic. Retrieved from <http://plato.stanford.edu/entries/logic-informal/>
- Gutiérrez, S. (2005). Discurso político y argumentación. *México: Universidad Autónoma Metropolitana, Unidad Xochimilco*, 11.
- Hearst, M. A. (1992). Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th conference on Computational linguistics-Volume 2* (pp. 539–545).
- Kant, I. (2001). *Lógica*. Ediciones AKAL.
- Linguistics, A. for C. (2014). Proceedings of the First workshop on Argumentation Mining. In *52nd Annual Meeting of the Association for Computational Linguistics* (p. 108). Baltimore, Maryland, USA.
- López, F., & Romero, V. (2014). *Mastering Python Regular Expressions*. Packt Publishing Ltd.
- Martínez, M. C. (2002). *Estrategias de lectura y escritura de textos: Perspectivas teóricas y talleres*.
- Matemática: Razonamiento Y Aplicaciones 10/e*. (2006). Pearson Educación.
- Mining, 3rd workshop on Argument. (2016). Workshop Description.
- Montoro, A. F. (2013). *Python 3 al descubierto*. RC Libros.
- Nguyen, H. V., & Litman, D. J. (2015). Extracting argument and domain words for identifying argument components in texts. *NAACL HLT 2015*, 22.
- Padró, L., Collado, M., Reese, S., Lloberes, M., & Castellón, I. (2010). *FreeLing 2.1: Five Years of Open-Source Language Processing Tools*.
- Padró, L., & Stanilovsky, E. (2012). Freeling 3.0: Towards wider multilinguality. In *LREC2012*.
- Peldszus, A. (2014). Towards segment-based recognition of argumentation structure in short texts. *ACL 2014*, 88.
- Portolés, J. (2011). *Marcadores del discurso*. Editorial Ariel.
- Reyes, G. (1990). *La Pragmática Lingüística*. Editorial Montesinos.
- Rodríguez Arias, E. (2010). LOS RAZONAMIENTOS CONDICIONALES: Un enfoque psicológico. *Ciencia Y Sociedad*, 196–206.
- Sánchez Avendaño, C. (2005). Los Conectores Discursivos : Su Empleo En Redacciones De

- Estudiantes Universitarios Costarricenses, (2), 169–199.
- Sardianos, C., Katakis, I. M., Petasis, G., & Karkaletsis, V. (2015). Argument Extraction from News. *NAACL HLT 2015*, 56.
- Stab, C., & Gurevych, I. (2014a). Annotating Argument Components and Relations in Persuasive Essays. In *COLING* (pp. 1501–1510).
- Stab, C., & Gurevych, I. (2014b). Identifying Argumentative Discourse Structures in Persuasive Essays. In *EMNLP* (pp. 46–56).
- Vega Reñón, L. (2012). *Las falacias: una introducción*.
- Wellner, B., & Pustejovsky, J. (2007). Automatically Identifying the Arguments of Discourse Connectives. In *EMNLP-CoNLL* (pp. 92–101).
- Weston, A. (2011). *Las claves de la argumentación* (4th ed.). Grupo Planeta.

### Referencias en línea

- Avallone, J. (n.d.). Regexper. Retrieved April 11, 2016, from [www.regexper.com](http://www.regexper.com)
- Calzado, F. (2014.). Aprende Lógica. Retrieved September 10, 2015, from [http://escuela2punto0.educarex.es/Humanidades/Etica\\_Filosofia\\_Ciudadania/Aprende\\_logica/logica/01concbasicos/000frsetintro.html](http://escuela2punto0.educarex.es/Humanidades/Etica_Filosofia_Ciudadania/Aprende_logica/logica/01concbasicos/000frsetintro.html)
- Goyvaerts, J. (2016). Runaway Regular Expressions; Catastrophic Backtracking. Retrieved June 11, 2016, from <http://www.regular-expressions.info/catastrophic.html>
- Harada Olivares, E. (2009). Argumentos , formalización y lógica informal. Retrieved from <http://www.redalyc.org/articulo.oa?id=10411360003>
- Huerta-Cepas, J., Serra, F., & Bork, P. (2016). ETE 3: Reconstruction, analysis and visualization of phylogenomic data. *Molecular Biology and Evolution*, msw046. <http://doi.org/10.1093/molbev/msw046>
- Koutný, P. (2006, September 12). Análisis del discurso político. Masaryk University. Retrieved from [https://is.muni.cz/th/109045/ff\\_b/](https://is.muni.cz/th/109045/ff_b/)
- Lloberes, M., & Padro, L. (n.d.). Freeling shallow parser: Chunking tags for spanish. Retrieved March 3, 2016, from <http://devel.cpl.upc.edu/freeling/svn/trunk/doc/grammars/esCHUNKtags>
- Logical Fallacies. (2009). Retrieved January 28, 2015, from <http://www.logicalfallacies.info/>
- Lucia. (2013). Tipos de discursos. Retrieved June 7, 2016, from <http://tiposde.info/tipos-de-discurso/>
- Marimón Llorca, C. (2008, January 31). Temas de Análisis y Redacción de Textos (Español). Retrieved from <http://rua.ua.es/dspace/handle/10045/4023>

- Microsoft. (2016). Retroceso en expresiones regulares. Retrieved March 20, 2016, from [https://msdn.microsoft.com/es-es/library/dsy130b4\(v=vs.110\).aspx#controlling\\_backtracking](https://msdn.microsoft.com/es-es/library/dsy130b4(v=vs.110).aspx#controlling_backtracking)
- Murillo Serna, R. O. (2001). Condiciones de producción y efectos del discurso político. *Investigaciones Sociales*. Retrieved from <http://revistasinvestigacion.unmsm.edu.pe/index.php/sociales/article/view/7472>
- Padró, L., Collado, M., Reese, S., Lloberes, M., & Castellón, I. (2010). *FreeLing 2.1: Five Years of Open-Source Language Processing Tools*. Retrieved from <http://upcommons.upc.edu/handle/2117/7616>
- República, P. de la. (2015a). Estado Mayor Presidencial | Presidencia de la República | Gobierno | gov.mx. Retrieved March 3, 2016, from <http://www.gob.mx/presidencia/acciones-y-programas/estado-mayor-presidencial-14579>
- República, P. de la. (2015b). Mensaje del Presidente Enrique Peña Nieto en la Revista First | Presidencia de la República | Gobierno | gov.mx. Retrieved May 19, 2016, from <http://www.gob.mx/presidencia/articulos/mensaje-del-presidente-enrique-pena-nieto-en-la-revista-first>
- Rodríguez Jiménez, G. (2013). Noción de falacia y sofisma, (1), 1–10. Retrieved from <http://objetos.unam.mx/logica/falacias/>
- Social, S. de D. (2015). Mensaje del secretario Meade en la reunión del Consejo Directivo de la Cámara Nacional de la Industria de la Transformación (Canacintra). Retrieved May 6, 2016, from <http://www.gob.mx/sedesol/prensa/mensaje-del-secretario-meade-en-la-reunion-del-consejo-directivo-de-la-camara-nacional-de-la-industria-de-la-transformacion-canacintra>
- Strano, M., & Colosimo, B. M. (2006). Logistic regression analysis for experimental determination of forming limit diagrams. *International Journal of Machine Tools and Manufacture*, 46(6), 673–682. <http://doi.org/10.1016/j.ijmachtools.2005.07.005>
- Workshop on Argument Mining. (2016). 3er Taller sobre Minería Argumento | ACL 2016, Berlín. Retrieved from <http://argmining2016.arg.tech/>

## Anexo A. Etiquetas morfológicas empleadas por Freeling

---

El analizador morfológico de *Freeling* utiliza etiquetas para representar la información morfológica de las palabras.

En este anexo se presentan un conjunto de tablas que muestran etiquetas por categorías. La información se tomó del sitio oficial de *Freeling*: <http://nlp.lsi.upc.edu/freeling-old/doc/tagsets/tagset-es.html>

Las categorías son:

1. Adjetivos
2. Adverbios
3. Determinantes
4. Nombres
5. Verbos
6. Pronombres
7. Conjunciones
8. Interjecciones
9. Preposiciones
10. Puntuación
11. Numerales
12. Fechas y horas

Para representar a los adjetivos se sigue la siguiente nomenclatura:

Pos.	Atributo	Valor	Código
1	Categoría	Adjetivo	A
2	Tipo	Calificativo	Q
		Ordinal	O
3	Grado	Aumentativo	A

Pos.	Atributo	Valor	Código
		Diminutivo	D
		Comparativo	C
		Superlativo	S
4	Género	Masculino	M
		Femenino	F
		Común	C
5	Número	Singular	S
		Plural	P
		Invariable	N
6	Función	-	0
		Participi	P

Unos ejemplos para construir adjetivos calificativos son los siguientes:

Forma	Lema	Etiqueta
alegres	alegre	AQ0CP0
alegre	alegre	AQ0CS0
bonita	bonito	AQ0FS0
grandazo	grande	AQAMS0

Para representar a los adverbios se sigue la siguiente tabla para su codificación:

ADVERBIOS			
Pos.	Atributo	Valor	Código
1	Categoría	Adverbio	R
2	Tipo	General	G
		Negativo	N

Los siguientes ejemplos muestran su construcción:

Forma	Lema	Etiqueta
despacio	despacio	RG
ahora	ahora	RG
siempre	siempre	RG

Forma	Lema	Etiqueta
hábilmente	hábilmente	RG

En el caso de los determinantes, hay una gran variedad de ellos, éstos son: Determinantes demostrativos, posesivos, interrogativos, exclamativos, indefinidos, numerales y artículos. Éstos se presentan en la siguiente tabla:

Pos.	Atributo	Valor	Código
1	Categoría	Determinante	D
2	Tipo	Demostrativo	D
		Posesivo	P
		Interrogativo	T
		Exclamativo	E
		Indefinido	I
		Artículo	A
3	Persona	Primera	1
		Segunda	2
		Tercera	3
4	Género	Masculino	M
		Femenino	F
		Común	C
		Neutro	N
5	Número	Singular	S
		Plural	P
		Invariable	N
6	Poseedor	Singular	S
		Plural	P

Los siguientes ejemplos son para los determinantes demostrativos:

Forma	Lema	Etiqueta
aquel	aquel	DD0MS0

<b>Forma</b>	<b>Lema</b>	<b>Etiqueta</b>
aquella	aquel	DD0FS0
aquellas	aquel	DD0FP0
aquellos	aquel	DD0MP0
esa	ese	DD0FS0
esas	ese	DD0FP0

Para los nombres, su representación es:

<b>NOMBRES</b>			
<b>Pos.</b>	<b>Atributo</b>	<b>Valor</b>	<b>Código</b>
1	Categoría	Nombre	N
2	Tipo	Común	C
		Propio	P
3	Género	Masculino	M
		Femenino	F
		Común	C
4	Número	Singular	S
		Plural	P
		Invariable	N
5-6	Clasificación se- mántica	Persona	SP
		Lugar	G0
		Organización	O0
		Otros	V0
7	Grado	Aumentativo	A
		Diminutivo	D

Los siguientes ejemplos muestran cómo se implementa:

<b>Forma</b>	<b>Lema</b>	<b>Etiqueta</b>
chico	chico	NCMS000
chicas	chico	NCFP000

Forma	Lema	Etiqueta
gatito	gato	NCMS00D
oyente	oyente	NCCS000
oyentes	oyente	NCCP000
cortapapeles	cortapapeles	NCMN000
tesis	tesis	NCFN000
Barcelona	barcelona	NP000G0
COI	coi	NP000O0
Pedro	pedro	NP000P0

En el caso de los verbos se representa de la siguiente manera:

Pos.	Atributo	Valor	Código
1	Categoría	Verbo	V
2	Tipo	Principal	M
		Auxiliar	A
		Semiauxiliar	S
3	Modo	Indicativo	I
		Subjuntivo	S
		Imperativo	M
		Infinitivo	N
		Gerundio	G
		Participio	P
4	Tiempo	Presente	P
		Imperfecto	I
		Futuro	F
		Pasado	S
		Condicional	C
		-	0
5	Persona	Primera	1
		Segunda	2
		Tercera	3



Pos.	Atributo	Valor	Código
6	Número	Singular	S
		Plural	P
7	Género	Masculino	M
		Femenino	F

Como se observa en la quinta fila (posición 3) existen muchos tipos de verbos, los siguientes ejemplos muestran los verbos principales en tiempo presente indicativo.

Tiempo	VERBOS PRINCIPALES		
	Forma	Lema	Etiqueta
PRESENTE DE INDICATIVO	canto	cantar	VMIP1S0
	cantas	cantar	VMIP2S0
	canta	cantar	VMIP3S0
	cantamos	cantar	VMIP1P0
	cantáis	cantar	VMIP2P0
	cantan	cantar	VMIP3P0

Así mismo pasa con los pronombres, existen pronombres de tipo personal, demostrativos, posesivos, indefinido, interrogativo, relativo y exclamativo. La tabla que viene a continuación describe su representación en etiquetas *Eagle*.

Pos.	Atributo	Valor	Código
1	Categoría	Pronombre	P
2	Tipo	Personal	P
		Demostrativo	D
		Posesivo	X
		Indefinido	I
		Interrogativo	T
		Relativo	R
		Exclamativo	E
3	Persona	Primera	1
		Segunda	2

Pos.	Atributo	Valor	Código
		Tercera	3
4	Género	Masculino	M
		Femenino	F
		Común	C
		Neutro	N
5	Número	Singular	S
		Plural	P
		ImpersonalMInvariable	N
6	Caso	Nominativo	N
		Acusativo	A
		Dativo	D
		Oblicuo	O
7	Poseedor	Singular	S
		Plural	P
8	Politeness	Polite	P

Algunos ejemplos para los pronombres personales son:

Forma	Lema	Etiqueta
yo	yo	PP1CSN00
me	me	PP1CS000
mí	mí	PP1CSO00
nos	me	PP1CP000
nosotras	nosotros	PP1FP000
nosotros	nosotros	PP1MP000
conmigo	conmigo	PP1CSO00
te	te	PP2CS000
ti	tí	PP2CSO00

En las conjunciones solo existen de dos tipos, las coordinadas y las subordinadas:

<b>CONJUNCIONES</b>			
<b>Pos.</b>	<b>Atributo</b>	<b>Valor</b>	<b>Código&gt;</b>
1	Categoría	Conjunción	C
2	Tipo	Coordinada	C
		Subordinada	S

Ejemplos de conjunciones coordinadas son:

<b>Forma</b>	<b>Lema</b>	<b>Etiqueta</b>
e	e	CC
empero	empero	CC
mas	mas	CC
ni	ni	CC

En el caso de las interjecciones con algunos ejemplos son:

<b>INTERJECCIONES</b>			
<b>Pos.</b>	<b>Atributo</b>	<b>Valor</b>	<b>Código</b>
1	Categoría	Interjección	I

Ejemplos:

<b>Forma</b>	<b>Lema</b>	<b>Etiqueta</b>
ah	ah	I
eh	eh	I
ejem	ejem	I
ele	ele	I

Las representaciones de las preposiciones son:

<b>Pos.</b>	<b>Atributo</b>	<b>Valor</b>	<b>Código</b>
1	Categoría	Adposición	S
2	Tipo	Preposición	P
3	Forma	Simple	S
		Contraída	C

Pos.	Atributo	Valor	Código
3	Género	Masculino	M
4	Número	Singular	S

Los ejemplos se muestran a continuación:

Forma	Lema	Etiqueta
al	al	SPCMS
del	del	SPCMS
a	a	SPS00
ante	ante	SPS00
bajo	bajo	SPS00
cabe	cabe	SPS00
con	con	SPS00
a_partir_de	a_partir_de	SPS00
a_causa_del	a_causa_del	SPCMS

La representación de los signos de puntuación es muy simple:

SIGNOS DE PUNTUACIÓN			
Pos.	Atributo	Valor	Código
1	Categoría	Puntuación	F

Las etiquetas de los signos de puntuación son:

Forma	Lema	Etiqueta
¡	¡	Faa
!	!	Fat
,	,	Fc
[	[	Fca
]	]	Fct
:	:	Fd
"	"	Fe
-	-	Fg
/	/	Fh

Forma	Lema	Etiqueta
¿	¿	Fia
?	?	Fit
{	{	Fla
}	}	Flt
.	.	Fp
(	(	Fpa
)	)	Fpt
«	«	Fra
»	»	Frc
...	...	Fs
%	%	Ft
;	;	Fx
-	-	Fz
+	+	Fz
=	=	Fz

Para los números, *Freeling* lo representa de la siguiente manera:

NUMERALES			
Pos.	Atributo	Valor	Código
1	Categoría	Cifra	Z
2	Tipo	partitivo	d
		Moneda	m
		porcentaje	p
		unidad	u

Ejemplos de los números son:

Forma	Lema	Etiqueta
239	239	Z
doscientos_veinte	220	Z
un_millón	1000000	Zd
una_docena	12	Zd

Forma	Lema	Etiqueta
tres_de_cada_cuatro	3/4	Zp
seis_octavas_partes	6/8	Zp
ochenta por ciento	80/100	Zp
74_%	74/100	Zp
2000_dólares	\$_USD:2000	Zm
30_Km_por_hora	SP_km/h:30	Zu
ocho_gramos_/litro	DN_g/l:8	Zu

Por último, se tienen las fechas y horas con la siguiente nomenclatura:

FECHAS Y HORAS			
Pos.	Atributo	Valor	Código
1	Categoría	Fecha/Hora	W

Ejemplos:

Forma	Lema	Etiqueta
viernes_26_de_septiembre_de_1992	[V:26:09:1992:??:??]	W
las_tres_de_la_tarde_del_26_de_septiembre_de_1992	[?:?:26:09:1992:03.00:pm]	W
sábado por la tarde	[S:?:?:?:?:?:?:pm]	W
marzo_de_1954	[?:?:?:03:1954:?:?:?:?:?]	W
siglo_XIX	[s:xix]	W
año_1987	[?:?:?:?:1987:?:?:?:?:?]	W
cinco_de_la_mañana	[?:?:?:?:?:05.00:am]	W

## Anexo B. Etiquetas de fragmentos de texto empleadas por Freeling

Para las etiquetas de fragmentos de texto (*chunking tags*), *Freeling* maneja un conjunto de etiquetas presentados a continuación. El sitio dónde se extrajo esta información es <http://devel.cpl.upc.edu/freeling/svn/trunk/doc/grammars/esCHUNKtags>.

a-fp	adjective, adjective feminine plural
a-fs	adjective, adjective feminine singular
a-mp	adjective, adjective masculine plural
a-ms	adjective, adjective masculine singular
s-a-fp	adjective, adjective phrase feminine plural
s-a-fs	adjective, adjective phrase feminine singular
s-a-mp	adjective, adjective phrase masculine plural
s-a-ms	adjective, adjective phrase masculine singular
s-adj	adjective, adjective phrase
sadv	adverb, adverb phrase
cuantif	adverb, adverb which expresses quantification
adv-interrog	adverb, interrogative adverb
neg	adverb, negation
coord	conjunction, coordinating conjunction
conj-subord	conjunction, subordinating conjunction
data	date
grup-complex-spec-fp	determiner, complex determiner feminine plural
grup-complex-spec-fs	determiner, complex determiner feminine singular
grup-complex-spec-mp	determiner, complex determiner masculine plural
grup-complex-spec-ms	determiner, complex determiner masculine singular
j-fp	determiner, definite determiner feminine plural
j-fs	determiner, definite determiner feminine singular
j-mp	determiner, definite determiner masculine plural
j-ms	determiner, definite determiner masculine singular
espec-ms-E	determiner, definite / indefinite / demonstrative de-
terminers masculin singular whose head is a noun feminine singular	
dem-fp	determiner, demonstrative determiner feminine plural
dem-fs	determiner, demonstrative determiner feminine singular
dem-mp	determiner, demonstrative determiner masculine plural
dem-ms	determiner, demonstrative determiner masculine singular
espec-fp	determiner, determiner feminine plural
espec-fs	determiner, determiner feminine singular
espec-mp	determiner, determiner masculine plural
espec-ms	determiner, determiner masculine singular
exc-fp	determiner, exclamative determiner feminine plural
exc-fs	determiner, exclamative determiner feminine singular
exc-mp	determiner, exclamative determiner masculine plural
exc-ms	determiner, exclamative determiner masculine singular
indef-fp	determiner, indefinite determiner feminine plural
indef-fs	determiner, indefinite determiner feminine singular
indef-ms	determiner, indefinite determiner feminine singular
indef-mp	determiner, indefinite determiner masculine plural
int-fp	determiner, interrogative determiner feminine plural
int-fs	determiner, interrogative determiner feminine singular
int-mp	determiner, interrogative determiner masculine plural
int-ms	determiner, interrogative determiner masculine singular
quant-fp	determiner, numeral determiner feminine plural
quant-fs	determiner, numeral determiner feminine singular
quant-mp	determiner, numeral determiner masculine plural

quant-ms	determiner, numeral determiner masculine singular
pos-fp	determiner, possessive determiner feminine plural
pos-fs	determiner, possessive determiner feminine singular
pos-mp	determiner, possessive determiner masculine plural
pos-ms	determiner, possessive determiner masculine singular
interjeccio	interjection
grup-nom-fp	noun, nominal chunk feminine plural
grup-nom-fs	noun, nominal chunk feminine singular
grup-nom-mp	noun, nominal chunk masculine plural
grup-nom-ms	noun, nominal chunk masculine singular
grup-nom	noun, nominal chunk neuter
n-fp	noun, noun feminine plural
nom-tmp-fp	noun, noun feminine plural - time expression
nom-tmp-fs	noun, noun feminine singular - time expression
n-fs	noun, noun feminine singular
nom-fs-E	noun, noun feminine singular that goes with masculine
singular determiners (ej. el agua)	
n-mp	noun, noun masculine plural
nom-tmp-mp	noun, noun masculine plural - time expression
n-ms	noun, noun masculine singular
nom-tmp-ms	noun, noun masculine singular - time expression
sn	noun, noun phrase
sn-tmp	noun, noun phrase - time expression
w-fp	noun, proper noun feminine plural
w-fs	noun, proper noun feminine singular
w-mp	noun, proper noun masculine plural
w-ms	noun, proper noun masculine singular
num-fp	number, number feminine plural
num-fs	number, number feminine singular
num-mp	number, number masculine plural
num-ms	number, number masculine singular
numero-nopart	number, number nonpartitive (ej. un centenar de per-
sonas)	
numero-part	number, number partitive (ej. un centenar)
numero	number
morf-pron	particle, pronoun morpheme 'es'
morfema-verbal	particle, verbal morpheme 'es'
prep	preposition
grup-sp	preposition, prepositional chunk
sp-de	preposition, prepositional phrase 'de'
grup-sp-inf	preposition, prepositional phrase whose daughter is
an infinitive clause	
prepc-ms	preposition, preposition + definite determiner mascu-
line singular	
prel-adv	pronoun, adverbial relative pronoun
paton-fp	pronoun, clitic 3rd person feminine plural
paton-fs	pronoun, clitic 3rd person feminine singular
paton-mp	pronoun, clitic 3rd person masculine plural
paton-ms	pronoun, clitic 3rd person masculine singular
paton-p	pronoun, clitic 3rd person neuter plural
paton-s	pronoun, clitic 3rd person neuter singular
patons	pronoun, clitic 3rd person
relatiu	pronoun, complex relative pronoun



pdem-fp	pronoun, demonstrative pronoun feminine plural
pdem-fs	pronoun, demonstrative pronoun feminine singular
pdem-mp	pronoun, demonstrative pronoun masculine plural
pdem-ms	pronoun, demonstrative pronoun masculine singular
pindef-fp	pronoun, indefinite pronoun feminine plural
pindef-fs	pronoun, indefinite pronoun feminine singular
pindef-mp	pronoun, indefinite pronoun masculine plural
pindef-ms	pronoun, indefinite pronoun masculine singular
pinterrog-fp	pronoun, interrogative pronoun feminine plural
pinterrog-fs	pronoun, interrogative pronoun feminine singular
pinterrog-mp	pronoun, interrogative pronoun masculine plural
pinterrog-ms	pronoun, interrogative pronoun masculine singular
pinterrog-s	pronoun, interrogative pronoun neuter singular
pinterrog	pronoun, interrogative pronoun 'quÃ©'
pinterrog-p	pronoun, interrogative pronoun 'quiÃ©nes'
pnum-fp	pronoun, numeral pronoun feminine plural
pnum-fs	pronoun, numeral pronoun feminine singular
pnum-mp	pronoun, numeral pronoun masculine plural
pnum-ms	pronoun, numeral pronoun masculine singular
psubj-s	pronoun, personal pronoun 1st / 2nd person singular
psubj-fp	pronoun, personal pronoun 3rd person feminine plural
psubj-fs	pronoun, personal pronoun 3rd person feminine singular
psubj-mp	pronoun, personal pronoun 3rd person masculine plural
psubj-ms	pronoun, personal pronoun 3rd person masculine singular
paton	pronoun, personal pronoun 3rd person - non direct object ('lo': la casa es grande --> la casa lo es)
psubj	pronoun, personal pronoun
ptonic	pronoun, personal pronouns 'mi' & 'si'
pposs-fp	pronoun, possessive pronoun feminine plural
pposs-fs	pronoun, possessive pronoun feminine singular
pposs-mp	pronoun, possessive pronoun masculine plural
pposs-ms	pronoun, possessive pronoun masculine singular
pposs-ns	pronoun, possessive pronoun neuter singular
cuyo-fp	pronoun, possessive relative pronoun feminine plural
cuyo-fs	pronoun, possessive relative pronoun feminine singular
cuyo-mp	pronoun, possessive relative pronoun masculine plural
cuyo-ms	pronoun, possessive relative pronoun masculine singular
pron	pronoun
pron-fp	pronoun, pronoun feminine plural
pron-fs	pronoun, pronoun feminine singular
pron-mp	pronoun, pronoun masculine plural
pron-ms	pronoun, pronoun masculine singular
pron-ns	pronoun, pronoun neuter singular
quien-p	pronoun, relative pronoun plural 'quien'
cual-p	pronoun, relative pronoun 'qual' plural
cual-s	pronoun, relative pronoun 'qual' singular
prel	pronoun, relative pronoun 'que'
quien-s	pronoun, relative pronoun singular 'quien'
F-no-c	punctuation

F-term	punctuation, sentence terminators
vaux	verb, auxiliary verb
vaux	verb, auxiliary verb
geraux	verb, gerund as auxiliary verb
forma-ger	verb, gerund
ger	verb, gerund
gerundi	verb, gerund - periphrasis
geraux-ser	verb, gerund 'ser' as auxiliary verb
infaux	verb, infinitive as auxiliary verb
forma-inf	verb, infinitive
infinitiu	verb, infinitive - prepositional head & periphrasis
infaux-ser	verb, infinitive 'ser' as auxiliary verb
inf	verb, infinitive - verbal head
grup-verb-inf	verb, infinitive whose head is prepositional
parti-aux	verb, participle as auxiliary verb
parti-fp	verb, participle feminine plural
parti-fs	verb, participle feminine singular
parti-mp	verb, participle masculine plural
parti-ms	verb, participle masculine singular
parti	verb, participle
parti-ser	verb, participle 'ser' as auxiliary verb
parti-flex	verb, participle with gender and number
ger-pas	verb, passive gerund
inf-pas	verb, passive infinitive
vser	verb, passive verb 'ser'
verb-pass	verb, passive verb
verb	verb
grup-verb	verb, verbal chunk
v-hacer-3p	verb, verb 'fer' singular - time expression

## Anexo C. Expresión regular

```

\s((?:RG_)|(?:SPS00_)\s[\\wáéíóú]+(?:\s(?:[\\wáéíóú]+))?)|(?:CS_)(?:\s(?:[\\wáé
íóú]+))?)|(?:P00CN000_)\s[\\wáéíóú]+(?:\s(?:[\\wáéíóú]+))?)((?:Fc_)(?:P00C_
(?:VMIP_)(?:DA0F_)|(?:Fc_)(?:P00C_)(?:VMIF_)(?:VMN0_)|(?:DP1M_
(?:NCMS_)(?:AQ0C_)(?:P00C_)|(?:SPS0_)(?:DI0F_)(?:AQ0C_
(?:NCFS_)|(?:Fc_)(?:VMIF_)(?:RG_)(?:\.)|(?:Fc_)(?:VSIP_)(?:AQ0F_
(?:DI0F_)|(?:Fc_)(?:DA0M_)(?:NCMS_)(?:SPS0_)|(?:DI0M_)(?:NCMS_
(?:AQ0C_)(?:RG_)|(?:Fc_)(?:VMN0_)(?:DA0F_)(?:NCFS_)|(?:Fc_)(?:VMN0_
(?:DI0F_)(?:SPS0_)|(?:DA0M_)(?:NCMS_)(?:SPS0_)(?:DA0F_)|(?:Fc_
(?:RN_)(?:VMIP_)(?:NCFS_)|(?:Fc_)(?:SPS0_)(?:PP1M_)(?:PD0M_)|(?:Fc_
(?:VSIP_)(?:AQ0M_)(?:VMN0_)|(?:DA0F_)(?:NCFS_)(?:SPS0_
(?:CS_)|(?:Fc_)(?:RN_)(?:VSIP_)(?:VMP0_)|(?:Fc_)(?:VMIP_)(?:RG_
(?:CS_)|(?:P00C_)(?:VMIF_)(?:NCMS_)(?:AQ0M_)|(?:VSIP_)(?:AQ0C_
(?:VMN0_)(?:DA0M_)|(?:Fc_)(?:VAIP_)(?:VMP0_)(?:SPS0_)|(?:Fc_)(?:RG_
(?:P00C_)(?:VMIP_)|(?:Fc_)(?:RG_)(?:AQ0F_)(?:CC_)|(?:VMIP_)(?:NCMS_
(?:SPS0_)(?:VMN0_)|(?:Fc_)(?:RN_)(?:RG_)(?:PP3F_)|(?:Fc_)(?:VMIP_)

```

```
(?:AQ0C_) (?:CS_)|(?:Fc_) (?:VMN0_) (?:SPS0_) (?:NCFS_)|(?:Fc_) (?:DA0F_)
(?:NCFS_) (?:PR0C_)| (?:NCFS_) (?:SPS0_) (?:VAN0_) (?:PP3F_)|(?:Fc_)
(?:VMIP_) (?:VSN0_) (?:VMP0_)| (?:DD0M_) (?:NP00_) (?:P00C_)
(?:VMIP_)] [^\w]
```

## Anexo D. Manual de instalación

### Manual de instalación del usuario

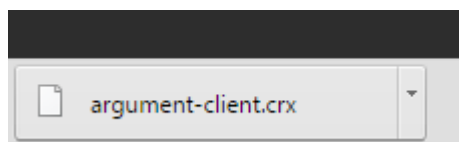
Para que el usuario haga uso del sistema para identificar argumentos en los sitios web necesita únicamente de la extensión de Google Chrome.

La dirección de descarga de la extensión es:

<https://drive.google.com/open?id=0B3ncvr8D1DQiNXZWb0dWZmdXV0E>

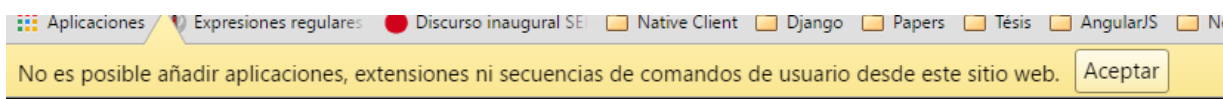
Se desplegará una página web con un botón con la opción para descargar la extensión. La extensión tiene un tamaño de 36KB en disco.

Para instalar la extensión en el navegador de Google Chrome se puede hacer de dos formas diferentes. Se presentará la primera opción porque es más fácil. Arrastrar la aplicación al navegador (en la página de extensiones ingresando la dirección: `chrome://extensions/`) para que se instale. La aplicación la puedes arrastrar desde el explorador de Windows o desde el navegador en la parte inferior (cuando se muestra la barra de descarga, ver figura).



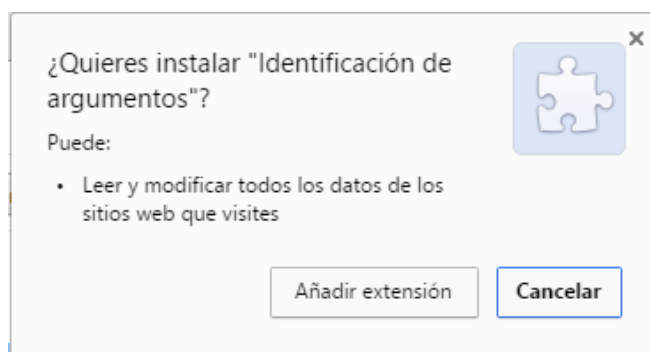
Aplicación descargada y que se visualiza en la parte inferior del navegador.

Posterior a ello, aparecerá un mensaje de advertencia del navegador, con lo cual dice: “Las extensiones, las aplicaciones y los temas pueden dañar tu ordenador. ¿Seguro que quieres continuar?”. El motivo de este mensaje es que la extensión se instala fuera de la tienda de extensiones de Chrome. Clic en “Continuar”, para instalar la aplicación. Si aparece un mensaje como se observa en la figura es que no estás en la página de extensiones del navegador. Se requiere abrir una nueva pestaña e intentar de nuevo.



Mensaje de error debido a que se intenta instalar la extensión en una página que no está permitido.

La aplicación tiene como nombre "Identificación de argumentos", así que este nombre aparecerá en un cuadro en el cual le indica los permisos que puede realizar la extensión. En la figura siguiente se observa el permiso que se requiere.



Cuadro de permisos de la extensión "Identificación de argumentos"

Clic en "Añadir extensión" para instalarlo al navegador. Al añadir la extensión aparecerá un ícono nuevo a un lado de la barra de búsqueda, esto indica que ha sido instalado correctamente y que ya puede ser usado.

La extensión solo puede ser usada en algunas páginas. Las páginas que empiezan con dirección "https" y que se encuentran en verde, no se pueden obtener sus informaciones, ya que están protegidos. Por ejemplo si ingresamos al sitio <https://www.gob.mx/presidencia/articulos/por-un-mexico-sin-corrupcion>, nos damos cuenta que no está en verde y por lo tanto la extensión funcionará sin problemas.

Al ingresar a la página de ejemplo, seleccionamos nuestra extensión para ponerlo en funcionamiento.

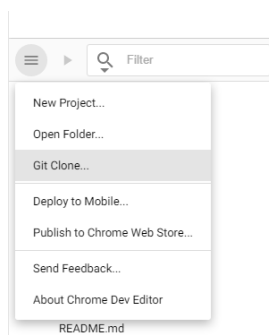
Pasados algunos segundos, si el sistema identifica argumentos, mostrará en colores las partes de un argumento, así como un cuadro flotante mostrando de una manera ordenada las premisas y conclusiones de los argumentos.

### Manual de instalación del desarrollador

Si se requiere modificar la extensión del navegador, puede realizar una copia desde el control de versiones de git: [git@git.asamblea.com:argument-client.git](https://github.com/argument-client)

La forma más fácil de realizar esto, es la extensión Chrome Dev Editor, que permite modificar extensiones y cargarlos automáticamente en el navegador.

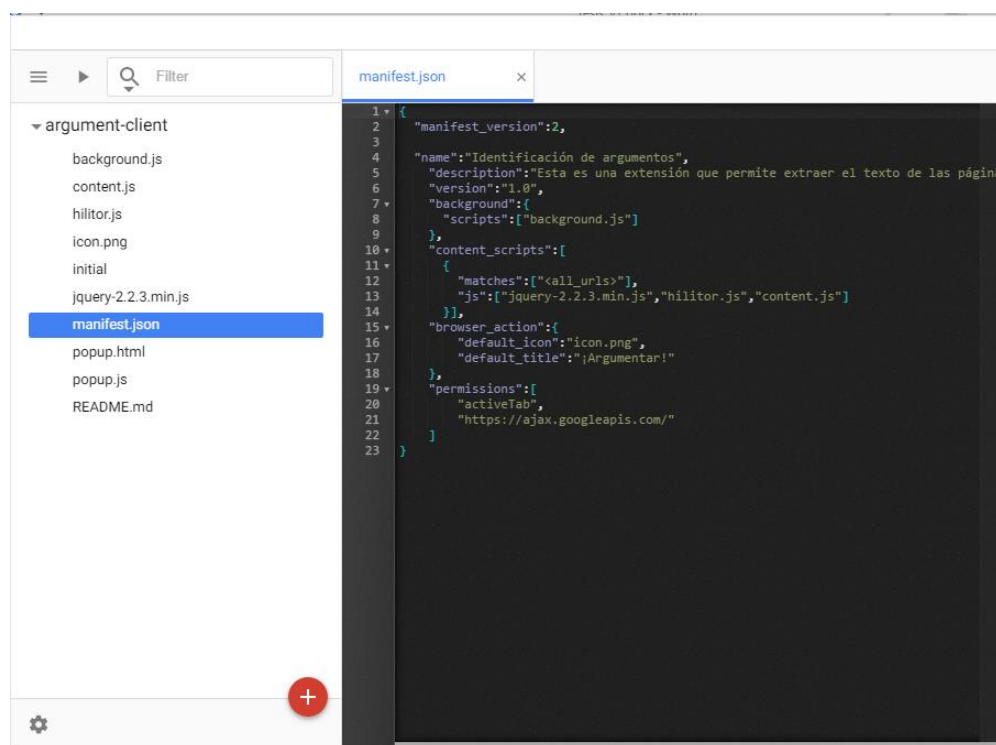
Chrome Dev Editor tiene la opción para clonar desde git un proyecto. En la figura se puede observar la opción.



Opción para obtener una copia del proyecto.

Al seleccionar la opción de “Git Clone” aparecerá una ventana donde te pide escribir la dirección del repositorio en donde se encuentra el código de la extensión.

El proyecto tiene el aspecto de la siguiente figura y contiene una serie de archivos, cada uno con un propósito específico. Para modificar el proyecto se requiere un conocimiento básico previo de extensiones de Chrome. Para mayor información de las extensiones puede consultar la extensión en la página de <https://developer.chrome.com/extensions/getstarted>.



Proyecto de argumentos cargado en Chrome Dev Editor

Para descargar el sistema del lado del servidor se requiere obtener el sistema en la dirección: <https://drive.google.com/open?id=0B3ncvr8D1DQYiYBFUIZ5UnpBYU0>

Es un archivo .zip que se descomprime, el proyecto es parte del IDE eclipse y éste a su vez es parte del *framework* Django. Se requiere instalar el plugin Python y Django en eclipse, para ello siga los pasos que están en la dirección: <http://www.pydev.org/download.html>.

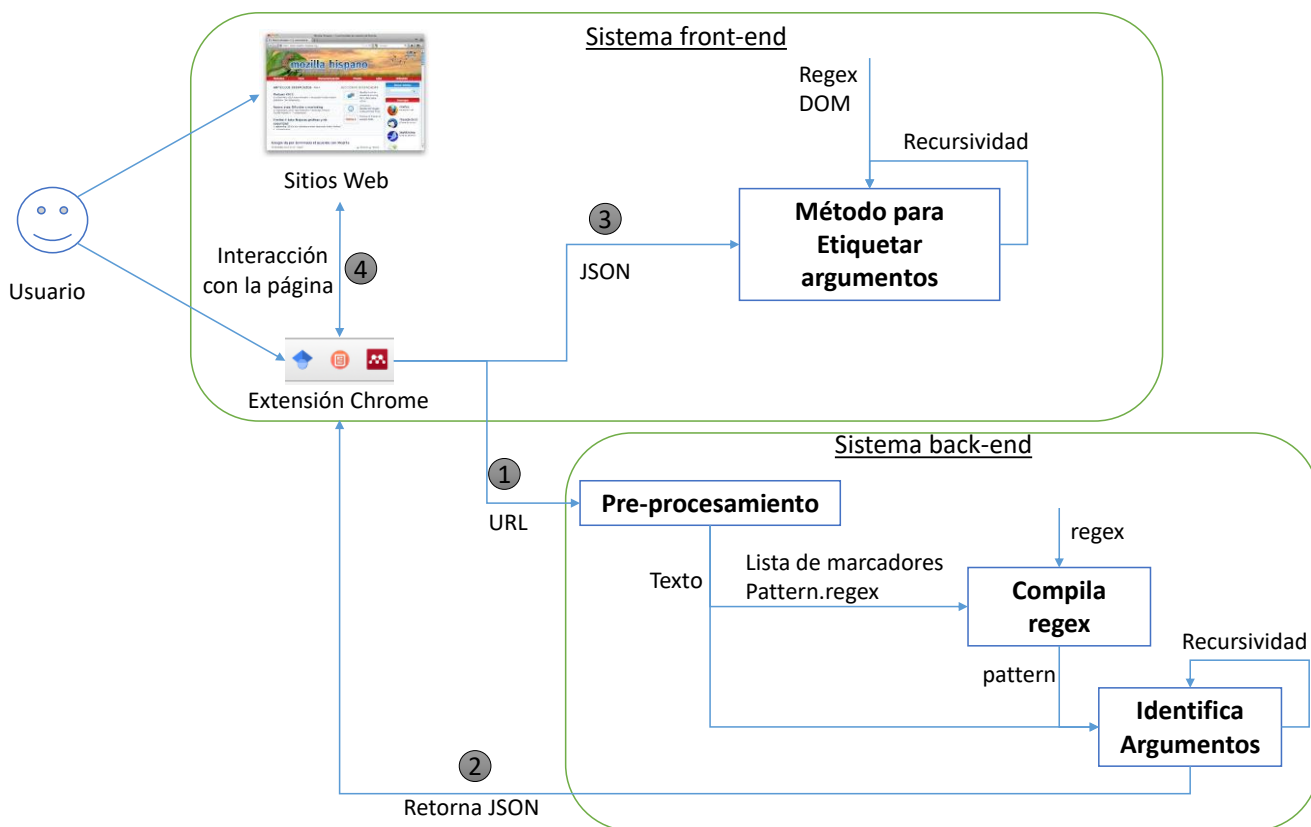
## Anexo E. Sistema identificador de argumentos en páginas web

---

El sistema está dividido en dos grandes grupos, sistema *front-end* y *back-end*. El sistema *back-end* se refiere al código desarrollado del lado del servidor, y el *front-end* es el código desarrollado del lado del cliente. El sistema *front-end* se comunica con el *back-end* realizando peticiones AJAX<sup>1</sup>. La petición AJAX se comunica con el método POST y envía un dato “*content*” con el valor URL de la página, además solicita una respuesta en formato JSON y la codificación en UTF-8. Estos detalles se pueden observar en el paso 1 de la siguiente figura, posteriormente al llegar al servidor se realiza un pre-procesamiento semejante al método de solución, es decir, el pre-procesamiento consiste en descargar la página con la URL obtenida, realiza una decodificación, guarda la página en un archivo “*resultado.html*”. Luego carga el archivo y con la librería LXML de Python realiza un limpiado de página web, con LXML se elimina la estructura de la página, los links, estilos y código JavaScript. Con la librería bs4 de Python se realiza un análisis de HTML, cargando el código restante de la página web para luego con el método *get\_text* obtener únicamente los textos de la página web, de esta manera tenemos todo el texto que la página web nos puede proporcionar. También se realiza pequeños cambios al texto para que el método de identificar argumentos no tenga inconvenientes. Estos cambios son modificar ciertos caracteres de HTML o Unicode a texto plano. Por ejemplo, existen varios tipos de caracteres que representan espacios, estos caracteres se cambian por el más común (U+0020), además los caracteres como <br> indica un salto de línea, así también las etiquetas <p> indican la existencia de un párrafo, todos ellos fueron cambiados por sus equivalentes en texto plano “saltos de línea” y “retorno de carro”.

---

<sup>1</sup> Es una técnica de desarrollo web para crear aplicaciones interactivas



Diseño arquitectónico del sistema para identificar argumentos

Para el entendimiento de la figura anterior y su explicación, se mencionará el flujo del proceso: las flechas de arriba que apuntan a la actividad (cuadro), son herramientas o recursos que se requieren para llevar a cabo el trabajo. Las flechas que apuntan desde el lado izquierdo indican la entrada de datos y el flujo, las flechas que salen desde la derecha de la actividad son salidas de datos que utilizará la misma actividad como si fuera un recurso, por último, las flechas que salen desde abajo indican la salida de datos final de la actividad.

Una vez que el método de pre-procesamiento finalice, éste le proporciona al siguiente método el texto procesado. El método "compila regex" carga las listas de marcadores y el archivo que contiene la expresión regular prediseñada, construye la expresión regular dinámicamente y la envía al siguiente método.

El método identificar argumentos, es una copia exacta de la actividad 7 de la fase 2 del método de solución, requiriendo el patrón compilado y un texto. Posteriormente el método identificar argumentos se llama así mismo cuando detecta premisas o un conjunto de premisas, al encontrar todos los sub-argumentos o argumentos devuelve un JSON con la misma estructurada presentada en la Figura 4.11, el flujo del JSON se puede apreciar en el paso 2 de la figura anterior, al llegar el resultado de vuelta a la extensión, ejecuta una secuencia de códigos de JavaScript.

El método para etiquetar argumentos crea un objeto "Hilitor" que es el encargado de encontrar en el DOM de la página web el argumento y sus componentes del argumento, para ello, se obtiene el primer argumento del JSON y navega por cada una de las etiquetas HTML, hasta coincidir el argumento con el texto de alguna etiqueta, si el primer argumento tiene sub-argumentos se realiza el mismo procedimiento mencionado previamente; así sucesivamente con los demás argumentos del JSON.

El argumento antes de ser utilizado en la búsqueda del DOM tuvo que ser procesado con anterioridad, es decir, se convirtió en un objeto regex. Existen ciertos caracteres especiales en las páginas web que imposibilitan la comparación exacta con el texto del argumento, así que se introdujo el carácter especial "\s" entre cada palabra para indicar que busque cualquier tipo de carácter de espacio que exista.

Si el argumento fue identificado en el DOM. El objeto "Hilitor", crea una etiqueta de tipo "*span*", introduce el argumento y le agrega el estilo color de fondo. Posteriormente sustituye el contenido de la etiqueta donde se encuentra el texto identificado como argumento con la etiqueta "*span*" creada previamente.

Así mismo sucede con los componentes de los argumentos, sin embargo, no así con los marcadores de conclusión y de premisas. Los marcadores a veces pueden ser palabras como "así", "mientras", "que", "si"; y estos tipos de palabras existen en multitud en los textos, así que no podemos implementar la identificación de un marcador y etiquetarlo en la página web, porque eso ocasionaría etiquetar palabras fuera de un argumento. Así que se prosiguió crear expresiones regulares y unir los marcadores con su respectiva proposición en el argumento, cada parte (marcador y cuerpo de la proposición) de la proposición es encerrado entre paréntesis para su captura individual. Si se encontraba todo el patrón de la expresión regular, se obtenía el valor de cada captura y nuevamente se creaba la etiqueta "*span*" para colocar en la etiqueta el valor de los grupos de captura de la expresión regular creada para identificar marcadores de los argumentos. De esta manera se pudo etiquetar solo aquellas palabras que están funcionando como introductores de razones o conclusiones en los argumentos.

Todos los pasos se hicieron así sucesivamente con los argumentos y sub-argumentos, y por esta razón el método para etiquetar argumentos es recursivo.

En la siguiente figura, se muestra el primer ejemplo de su funcionamiento y a continuación se detallan los pasos.

En la parte derecha superior, se encuentra una lista de íconos, cada uno de ellos representa un programa, a esto se le conoce como extensiones de navegador. Lo que está marcado en amarillo es la extensión que identifica argumentos. Al entrar a una página podemos hacer clic y en pocos segundos nuestro documento se verá subrayado por diferentes colores. Cada color representa algún componente del argumento. En la siguiente lista se presenta el significado de cada color:

- El color fucsia simboliza una proposición de tipo premisa.
- El color verde simboliza los marcadores de conclusión



- El color azul turquesa simboliza la primera conclusión, en otras palabras, la conclusión del argumento padre.
- El color amarillo simboliza que es la conclusión de un sub-argumento, eso implica que la conclusión también es premisa. Si existiera una conclusión de otro sub-argumento entonces también se etiquetaría de color amarillo.
- El color rosado simboliza el etiquetado de marcadores de premisas.
- El color gris indica un soporte, ya sea negativo o positivo para el argumento.

En la parte derecha inferior se puede apreciar la pestaña *preview*, el cual indica el JSON devuelto por el servidor.

The screenshot shows a web browser window with the URL `www.gob.mx/shcp/articulos/palabras-del-secretario-de-hacienda-en-`. The page content includes a submenú and a paragraph of text with several highlighted sections in different colors: pink, yellow, cyan, and grey. The Chrome DevTools Network tab is open, showing a list of requests. The selected request is expanded to show the 'Preview' tab, which displays a JSON object representing the response. The JSON structure includes fields like 'argument/', 'parts:', 'conclusion:', 'premise2:', 'position:', and 'subargument:'.

Ejemplo de etiquetado de argumentos desde un sitio del gobierno.

La extensión de Chrome también muestra un cuadro con los argumentos identificados, estos argumentos están representados ordenadamente, es decir, el orden de sus premisas y conclusiones, así como si aparece un sub-argumento. Dicho esto se muestra un ejemplo en la siguiente figura para su mejor entendimiento.

<p>Blog Multimedia Prensa Agenda Acciones y Programas Finanzas Públicas y Presupuesto</p>									
<p>Y quiero destacar algo que sin duda para Pedro Pablo, para su familia, es un dato relevante el día de hoy, que Pedro Pablo hace casi 20 años, en 1996, cuando iniciaba su carrera profesional se desempeñó en la Lotería Nacional, trabajó casi dos años en la Lotería Nacional, en ese entonces como secretario particular del director general. Así que de alguna manera Pedro Pablo regresa a una institución que conoce y a la cual le tiene un entrañable cariño.</p>	<table border="1"> <tr> <td></td> <td>Y quiero destacar algo que sin duda para Pedro Pablo, para su familia, es un dato relevante el día de hoy, que Pedro Pablo hace casi 20 años, en 1996, cuando iniciaba su carrera profesional se desempeñó en la Lotería Nacional, trabajó casi dos años en la Lotería Nacional, en ese</td> </tr> <tr> <td>Premisa-</td> <td>entonces como secretario particular del director general</td> </tr> <tr> <td>Conclusión</td> <td>Así que de alguna manera Pedro Pablo regresa a una institución que conoce y a la cual le tiene un entrañable cariño</td> </tr> <tr> <td>Premisa</td> <td>Ahora nos tocará también dirigir a Pronósticos para la Asistencia Pública, y</td> </tr> </table>		Y quiero destacar algo que sin duda para Pedro Pablo, para su familia, es un dato relevante el día de hoy, que Pedro Pablo hace casi 20 años, en 1996, cuando iniciaba su carrera profesional se desempeñó en la Lotería Nacional, trabajó casi dos años en la Lotería Nacional, en ese	Premisa-	entonces como secretario particular del director general	Conclusión	Así que de alguna manera Pedro Pablo regresa a una institución que conoce y a la cual le tiene un entrañable cariño	Premisa	Ahora nos tocará también dirigir a Pronósticos para la Asistencia Pública, y
	Y quiero destacar algo que sin duda para Pedro Pablo, para su familia, es un dato relevante el día de hoy, que Pedro Pablo hace casi 20 años, en 1996, cuando iniciaba su carrera profesional se desempeñó en la Lotería Nacional, trabajó casi dos años en la Lotería Nacional, en ese								
Premisa-	entonces como secretario particular del director general								
Conclusión	Así que de alguna manera Pedro Pablo regresa a una institución que conoce y a la cual le tiene un entrañable cariño								
Premisa	Ahora nos tocará también dirigir a Pronósticos para la Asistencia Pública, y								

Identificación de argumentos con Popup.

Los colores azules que aparecen en la ventana emergente indican las delimitaciones de cada argumento principal.

***cenidet***<sup>®</sup>  
*Centro Nacional de Investigación  
y Desarrollo Tecnológico*