



**I
N
A
O
E**

Análisis Textual de Argumentos en Escritos Académicos

Jesús Miguel García Gorrostieta, Aurelio López López

Laboratorio de Tecnologías del Lenguaje.
Coordinación de Ciencias Computacionales.
Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), México.

Reporte Técnico No. CCC-16-011
6 de diciembre de 2016

© Coordinación de Ciencias Computacionales
INAOE

Luis Enrique Erro 1
Sta. Ma. Tonantzintla,
72840, Puebla, México.



Índice

Índice.....	2
Resumen.....	3
1. Introducción.....	4
2. Antecedentes.....	7
2.1 Detección de argumentación.....	9
2.2 Clasificar los componentes argumentativos.....	9
2.3 Detección de relaciones entre componentes.....	11
2.4 Evaluación de la argumentación.....	13
3. Problemática.....	17
4. Preguntas, objetivos y contribuciones.....	19
4.1 Objetivo general.....	19
4.2 Objetivos específicos.....	19
4.3 Contribuciones esperadas.....	20
4.4 Hipótesis.....	20
5. Metodología.....	21
6. Trabajo realizado y resultados preliminares.....	31
6.1 Identificar los conjuntos de datos con escritos académicos de investigación.....	31
6.2 Construcción de corpus argumentativo.....	32
6.2.1 Identificar las secciones de escritos de investigación con mayor cantidad de argumentos.....	32
6.2.2 Desarrollar una guía de anotación.....	34
6.2.3 Realizar el estudio de anotación.....	35
6.3 Clasificar segmentos de texto como argumentativo o no argumentativo.....	38
6.3.1 Configuración experimental.....	38
6.3.2 Resultados experimentales.....	39
6.3.3 Muestra con un anotador.....	41
6.3.4 Muestra con tres anotadores.....	43
6.3.5 Corpus de sección “Planteamiento del Problema”.....	48
6.3.6 Corpus de sección “Justificación”.....	49
6.3.7 Conclusión de experimentos.....	51
7. Conclusiones.....	53
8. Publicaciones.....	54
Referencias.....	55
Anexos.....	59

Resumen

La argumentación en escritos académicos como tesis y propuestas de investigación es necesaria para comunicar claramente las ideas y convencer al lector de las aseveraciones presentadas. La identificación de argumentos es una tarea compleja y por otra parte, necesaria para la buena redacción de textos académicos de investigación. La elaboración de estos textos conforma un elemento clave para el desarrollo del alumno, principalmente en el caso de las tesis, las cuales son un requisito para la obtención del grado académico. Sin embargo, debido a la deficiencia de algunos estudiantes en su habilidad para argumentar hace necesaria una revisión más detallada por parte del asesor. Para apoyar al alumno y al profesor se propone utilizar técnicas de procesamiento del lenguaje natural con el objetivo de analizar automáticamente la argumentación en escritos académicos. En primer lugar se construyó un corpus anotado con los componentes, relaciones, tipos y niveles de argumentación. Posteriormente se analizaron las características léxicas y sintácticas de los escritos en el corpus para la generación de representaciones vectoriales. Con estas representaciones se utilizó aprendizaje computacional para clasificar párrafos y oraciones argumentativas. Además se propone el desarrollo de métodos para la identificación de componentes y relaciones argumentativas, necesarias para realizar un diagnóstico de la argumentación en los escritos académicos. Como contribuciones esperadas se desarrollarán una serie de métodos para la evaluación automática de argumentos con base en sus componentes y relaciones identificadas. Los resultados obtenidos a la fecha son un método para la clasificación de párrafos y oraciones argumentativas, así como la creación de una porción del corpus con argumentos anotados.

1. Introducción

La argumentación en la actualidad está a nuestro alrededor. Ya sea en medios electrónicos como foros o blogs, en los cuales las personas escriben sus opiniones y presentan elementos (pruebas) para que dichas opiniones sean aceptadas como ciertas, así como en medios escritos, tales como: periódicos, revistas, textos legales, ensayos, artículos científicos, tesis o reportes de investigación. De estos medios, en particular para los textos científicos, es necesario ofrecer argumentos razonados desarrollados a partir de evidencia claramente presentada que lleven a una conclusión consistente (Lindsay, 2011).

En investigaciones recientes se ha comenzado a estudiar el fenómeno del procesamiento automático de argumentos, donde de forma interdisciplinaria se combina la inteligencia artificial y las teorías de la argumentación con el propósito de mejorar el proceso de extracción y recuperación de información. Por ejemplo, en el ámbito legal, se busca facilitar el acceso a la jurisprudencia que brinde apoyo a un caso (Mochales & Moens, 2008, 2011; Wyner & Bench-Capon, 2007; Wyner et al, 2010). Por otra parte, en artículos científicos, particularmente biomédicos, se trata de identificar argumentos a favor o en contra de una hipótesis bajo investigación (Green, 2015). O bien en las redes sociales se pretende identificar los comentarios a favor o en contra en un debate, para observar cuál es la postura de la mayoría (Cabrio & Villata, 2012) y evaluar dichos argumentos en función de si cumplen con una estructura admisible (Park, Blake & Cardie, 2015). Por último, en otro ámbito, también se evalúa el nivel de argumentación en ensayos para calificar al alumno y ofrecerle una retroalimentación inmediata (Persing & Ng, 2015).

El acto de escribir es un proceso complejo que comprende varias etapas como la planificación, edición y revisión. En el ámbito académico, la etapa de revisión suele realizarla el profesor analizando el escrito para identificar los errores e indicarlos al alumno. Es una tarea que podría encontrar un apoyo en herramientas computacionales desde la etapa de edición del alumno, al facilitarle las indicaciones necesarias para redactar textos de calidad antes de ser corregidos. Ejemplos de estas herramientas son Criterion (Burstein, Chodorow & Leacock, 2003), Writing Pal W-Pal (Roscoe, 2014) y SWoRD (Cho & Schunn, 2007). En la actualidad, estos programas mejoran la escritura de los alumnos con una retroalimen-

tación. Sin embargo, no ofrecen un diagnóstico del uso de los componentes-relaciones de argumentos en el texto y esta dimensión del análisis textual es básica en un texto científico.

La escritura argumentativa es una parte fundamental en la redacción de escritos académicos. Esta dimensión de la escritura es utilizada en ensayos, artículos científicos y tesis para fundamentar las aseveraciones realizadas con razonamientos sólidos. La identificación de razonamientos sólidos (premisas) se observa en los trabajos de Stab & Gurevich (2014) para ensayos y Kirschner, Eckle-Kohler & Gurevych (2015) para artículos científicos. Sus investigaciones sirven de base para la construcción de herramientas computacionales que apoyen el proceso de escritura argumentativa. Sin embargo, no se han observado trabajos dirigidos a analizar de forma automática los argumentos en textos académicos más extensos y con una estructura diferente como: tesis y reportes técnicos.

El análisis automático de argumentos en reportes técnicos y tesis se convierte en una necesidad en el ámbito académico, por su capacidad de acortar los tiempos y facilitar la labor de análisis de textos largos. Los reportes técnicos y tesis se elaboran al final de curso o de una carrera universitaria como requisito de acreditación. Debido a la importancia de estos trabajos en el avance profesional del alumno, se hacen necesarios métodos que analicen los argumentos para ofrecer un diagnóstico automático.

Los métodos propuestos para lograr dicho diagnóstico consisten, primero en identificar la presencia de argumentación en el texto, posteriormente clasificar los componentes argumentativos (premisas/conclusiones), a continuación ubicar las relaciones (apoyo/ataque) entre dichos componentes y finalmente modelar la estructura argumentativa de dicho texto. Con la información obtenida de estos métodos es posible obtener un diagnóstico detallado de los argumentos.

De acuerdo a trabajos anteriores (Capaldi, 2000), un argumento puede definirse como una serie de aseveraciones (oraciones, proposiciones) que individualmente o en su conjunto apoyan a otra aseveración. La aseveración que resulta apoyada se denomina conclusión. Hay una sola conclusión para cada argumento, pero puede haber toda una serie de aseveraciones de apoyo. Las aseveraciones que brindan apoyo se denominan premisas. Entre las teorías de argumentación (Freeman, 2011; Toulmin, 1958; Walton, Reed & Macagno, 2008), el consenso para la estructura de un argumento indica que está conformada por varios componentes argumentativos, los cuales consisten en una conclusión y varias premi-

sas. En la presente investigación nos apegamos al modelo conclusión/premisa presentado por Freeman (2011).

Una representación gráfica de estructuras de argumentos ayuda a comprender la forma en que se relacionan sus componentes. Esto se hace identificando cada premisa y conclusión con una letra, que se asocian a nodos de un grafo. Posteriormente utilizando arcos dirigidos (flechas) es posible indicar las relaciones entre dichos componentes. Por ejemplo, el diagrama de un argumento sencillo tiene solo una premisa que es usada como base para apoyar la conclusión (Walton, 2005). Por ejemplo:

*Hoy en día las instituciones educativas poseen un mayor número de equipos de cómputo con Internet.*_A *Por lo tanto, más estudiantes tienen acceso a Internet.*_B



Fig. 1. Diagrama de argumento sencillo

Como se observa en la figura 1, la primera oración es la premisa (A) que apoya la conclusión en la segunda oración (B con texto subrayado). El argumento sencillo es aquel donde la premisa proporciona el fundamento para tomar como cierta la conclusión presentada.

El análisis computacional de dichos argumentos involucra varios aspectos, desde la identificación de conclusiones y premisas, hasta determinar cuáles son las relaciones entre dichos elementos. Con esta información es posible indicarle al estudiante la falta de conclusiones en sus escritos o bien la falta de premisas de apoyo para dichas conclusiones. Por estas razones, el propósito de esta investigación es proponer una serie de métodos para la extracción y evaluación de la argumentación en escritos académicos de investigación que ayuden a los estudiantes en su redacción. Como avance en la investigación propuesta, se presentan resultados preliminares en los cuales se logra clasificar párrafos y oraciones argumentativas con niveles similares a los reportados en el estado del arte, además se realizó la anotación de una porción del corpus.

El documento de propuesta se organiza de la siguiente manera. En la sección 2, se detallan los trabajos relacionados con esta investigación. En la sección 3, se describe la

problemática. En la sección 4 se presentan las preguntas de investigación, objetivos y las principales contribuciones. En la sección 5, se describe la metodología a seguir por la presente investigación. En la sección 6 se reportan los experimentos realizados y los resultados alcanzados hasta el momento. Finalmente, en la sección 7, se concluye con observaciones finales.

2. Antecedentes

El análisis computacional de argumentos se ha realizado en ensayos, textos legales y artículos científicos. En la literatura se ha observado que dicho análisis es altamente dependiente del tipo de documento utilizado. Por ejemplo, en el caso de los documentos legales, se observan textos muy estructurados para cumplir con los requisitos de las cortes de justicia. Sin embargo, en ensayos académicos se tiene más libertad en cuanto a la estructura de los argumentos, lo cual complica la tarea de análisis.

A continuación se presentan diversas investigaciones que han tratado las tareas relacionadas al análisis, extracción y evaluación de argumentos en diferentes tipos de documentos, dichas tareas son:

- creación del corpus de argumentación
- detección de argumentación
- clasificación de tipo de componente argumentativo
- identificación de relación argumentativa
- determinación de nivel de argumentación
- clasificación de tipo de argumento

El primer paso de la tarea de análisis de textos consiste en contar con un corpus anotado que permita validar el desempeño del método propuesto. Como se ha observado en la literatura, la mayoría de los investigadores del área de análisis de argumentos se dan a la tarea de crear sus corpus anotados con un determinado esquema de argumentación. En la literatura se encuentran pocos corpus disponibles con argumentos anotados. Uno de los más utilizados entre los investigadores para identificar la presencia de argumentos es el corpus Araucaria (Katzav, Reed & Rowe, 2004). Dicho corpus cuenta con varios tipos de documentos (p. ej. registros de parlamento, periódicos, resúmenes judiciales y foros de discusión

en idioma inglés) para el cual se tienen anotadas sus premisas y conclusiones así como el esquema de argumentación utilizado. Sin embargo no se indica un nivel de acuerdo entre anotadores en dicho estudio y esto le resta fiabilidad.

La creación del corpus para cada investigación se observa en diferentes tipos de texto así como en diferentes dominios. En Mochales & Moens (2008) se construyó un corpus al anotar 10 documentos legales del corpus ECHR (European Court of Human Rights) al realizar la anotación de premisas y conclusiones, en dicho corpus se obtiene un nivel de acuerdo entre los dos anotadores del 80% con indicador Kappa (ver definición en anexo A). Posteriormente, Mochales & Moens (2011) incrementaron el número de anotadores a tres y el número de documentos en su corpus a 47, lo cual produjo una disminución del nivel de acuerdo entre sus anotadores a un valor de Kappa del 75%. Es importante indicar que al tratarse de textos legales, la estructura está muy bien definida lo cual facilita el proceso de anotación e incrementa el nivel de acuerdo entre los anotadores.

Por otro lado, en el trabajo de Stab & Gurevych (2014), se utilizan 90 ensayos persuasivos de temáticas elegidas al azar, anotados por tres participantes, quienes indicaron los componentes argumentativos con un nivel de acuerdo de Fleiss Kappa (ver definición en anexo B) para: conclusión mayor un 83% (postura del autor con respecto al tema analizado), premisas un 70% y conclusiones un 65%. Además, el porcentaje de acuerdo en las relaciones entre componentes argumentativos de ataque fue de 80% y 81% en los casos de apoyo.

En la investigación de Kirschner, Eckle-Kohler & Gurevych (2015) se crea un corpus con 24 artículos científicos en educación para las secciones de introducción y discusión, anotado por cuatro participantes para los componentes del argumento: premisas y conclusiones, así como las cuatro relaciones de dichos componentes argumentativos (apoyo, ataque, secuencia y detalle), obteniendo un nivel de acuerdo promedio de Fleiss Kappa del 41%. Con esto se observa que obtener niveles de acuerdo aceptables entre anotadores en textos científicos es una tarea compleja, la cual recae en una adecuada guía de anotación, así como el seguimiento regular a los anotadores durante la construcción del corpus. Para nuestra investigación, el tipo de documento más cercano son los artículos científicos ya que las tesis y propuestas de investigación de los alumnos comparten una estructura similar con la diferencia de que dichas propuestas son más extensas.

2.1 Detección de argumentación

Una vez construido el corpus es necesario analizar la tarea de detectar la argumentación en párrafos, oraciones o cláusulas, con la finalidad de identificar la presencia de premisas o conclusiones. Para ello investigadores como Moens et al (2007), realizan la clasificación de oraciones argumentativas y no argumentativas en el corpus Araucaria. Representan oraciones con características como combinaciones de pares de palabras, verbos y las estadísticas de texto, y utilizando un clasificador de Bayes reportan un 73,75% de exactitud. Además Mochales & Moens (2011) utiliza el corpus de textos legales ECHR con 47 documentos anotados, donde se clasifica como argumentativo o no las cláusulas (sub-oración) de las oraciones utilizando un clasificador de entropía máxima, y reportan una exactitud del 80% para dicha tarea. Es importante hacer notar que los textos legales cuentan con una estructura particular que les permite a los abogados identificar claramente los argumentos.

Por otro lado, la identificación de párrafos argumentativos es investigada por Florou et al. (2013) quienes utilizan 5 conjuntos de categorías de argumentos (justificación, explicación, deducción, refutación y condicional) y características basadas en el modo y tiempo de los verbos. En dicho trabajo se consigue identificar segmentos de texto con argumentación, empleando un algoritmo de clasificación de árbol de decisión J48, reportando una medida F de 76.4%.

Otro enfoque para la identificación de argumentación en segmentos de texto, lo presentan Goudas et al. (2014), los cuales construyen un corpus a partir de 204 documentos recolectados de redes sociales, que son anotados con sus premisas. Ellos utilizan características estructurales, léxicas, contextuales y gramaticales para representar cada oración. Empleando un clasificador de regresión logística reportan una medida F de 77%.

Por otra parte, Sardinós et al. (2015) identifican segmentos de texto que corresponden a componentes argumentativos. Se utilizan etiquetas de partes de la oración, lista de palabras claves y representaciones distribucionales para representar los textos. Utilizando CRF (Conditional Random Fields) reportan una medida F del 32.21%.

2.2 Clasificar los componentes argumentativos

Una vez identificada la presencia de argumentación en segmentos de texto, el siguiente paso es extraer y clasificar los componentes argumentativos presentes en dicho texto (p. ej. premisa, conclusión). Los investigadores Mochales & Moens (2011) y Goudas et

al. (2014) como se indicó anteriormente, realizan en primer lugar una segmentación del texto argumentativo, para posteriormente clasificar cada segmento en su respectivo componente argumentativo. Mochales & Moens (2011) utilizan una máquina de soporte vectorial para clasificar cada cláusula con una medida F para premisas de 68% y conclusiones de 74%. Aquí puede verse que nuevamente se aprovecha la estructura de los textos legales. Goudas et al. (2014) utilizan campos aleatorios condicionales (*Conditional Random Fields o CRF*) para identificar el fragmento del texto que corresponde a una premisa (cada palabra se identifica como inicio o parte de una premisa), y con esto reportan una medida F de 42%.

Para ensayos persuasivos, en el trabajo de Stab & Gurevych (2014) se utilizó una máquina de soporte vectorial (SVM) para la clasificación de componentes argumentativos como: sin argumentación, premisa, conclusión y conclusión mayor. Utilizaron características estructurales, léxicas, sintácticas y contextuales para representar cada proposición, con las cuales reportan una exactitud del 77.3%. Utilizando las mismas características, Persing & Ng (2016) realizan la misma clasificación de componentes, reportan una medida F (ver definición en anexo C) de 57.2% utilizando un clasificador de entropía máxima MALLET. Además, en un estudio realizado con el mismo corpus, Nguyen & Litman (2015) realizan esta misma clasificación de componentes argumentativos con un SVM. Utilizaron palabras del dominio y argumentativas extraídas de ensayos persuasivos sin etiquetar utilizando LDA. Reportan una exactitud del 79% en esta tarea.

Otra forma de clasificar los argumentos consiste en utilizar artículos científicos, como se observa en el trabajo de Teufel (1999) que emplea zonas de argumentación (meta, antecedentes, trabajo propio, contraste, bases) con un corpus de 80 artículos científicos del área de computación. Utiliza un clasificador Naive Bayes para obtener una precisión en la identificación de argumentos de los tipos: meta 57%, contraste 57% y otros trabajos 62%.

Otro acercamiento a la identificación de premisas se ha conseguido aplicando técnicas de análisis de sentimientos. Villalba & Saint-Dizier (2012) identifican estructuras del discurso tales como justificación, elaboración e ilustraciones que apoyan las opiniones en un corpus de reseñas de hoteles y restaurantes. Utilizaron reglas de extracción de argumentos con características léxicas como términos que expresan polaridad, adverbios con inten-

sidad y verbos del dominio para identificar estructuras del discurso. Reportan una precisión del 92% y recuerdo del 86% en la identificación de justificaciones.

Rooney, Wang & Browne (2012) realizan la identificación de componentes en el corpus Araucaria con 662 documentos. Utilizan un SVM con una función de kernel de secuencia normalizado para clasificar secuencias como premisa, conclusión o ninguna. Además, indican que es aplicable para cualquier tipo de documento y temática al no depender de una selección de características. Reportan una exactitud del 65% en identificación de componentes. Para la representación de las oraciones utilizaron etiquetas de partes de la oración, lemas y palabras.

2.3 Detección de relaciones entre componentes

Después de identificar los tipos de componentes argumentativos y sus relaciones, es posible obtener una estructura argumentativa. En el caso de textos legales se observa en el trabajo de Mochales & Moens (2009, 2011), en el cual se utiliza una Gramática Libre de Contexto (CFG Context-Free Grammar) para dar una forma de árbol a la estructura de argumentos. Las CFGs no solo son utilizadas para identificar relaciones sino también para clasificar premisas y conclusiones en textos legales a nivel oración, como se presenta en trabajo de (Mochales & Moens, 2008) obteniendo un 59% de precisión en la clasificación de premisas y 61% de precisión para la clasificación de conclusiones. Dicha gramática fue creada a partir del análisis de 10 documentos de textos legales y se probó con 20 documentos de textos legales.

Por otro lado, utilizando el aprendizaje computacional para identificar relaciones entre componentes, Stab & Gurevech (2014) logran identificar pares de componentes con una relación de apoyo. Para generar el conjunto de entrenamiento se crearon todos los posibles pares entre componentes. Posteriormente se indicaron como apoyo o no-apoyo, obteniendo 989(15.6%) pares de apoyo y 5,341(84.4%) pares de no-apoyo. Los atributos utilizados para representar a dichos pares fueron: estructurales, tales como número de tokens y número de signos de puntuación; léxicos, tales como son las combinaciones de pares de palabras, primeras palabras y palabras modales; sintácticas, donde se tienen reglas de producción; e indicadores como marcadores argumentativos con características binarias de aparición, y predicción de tipo de componentes ya sea premisa, conclusión o conclusión mayor. Poste-

riormente, utilizando un clasificador SVM reportan una exactitud de 86.3% y una medida F1 para la identificación de pares de componentes con una relación de apoyo de 51.9%. Utilizando las mismas características, Persing & Ng (2016) identifican relaciones de ataque y apoyo entre componentes de argumentos. Reportan una medida F de 38.8% para dicha tarea con el clasificador de entropía máxima MALLET.

Un enfoque similar se observa en el trabajo de Carstens & Toni (2015), en el cual se utiliza también aprendizaje computacional con un conjunto de características para representar pares de oraciones. Las características utilizadas son bolsa de palabras, atributos que relacionan dos oraciones como: similitud basada en WordNet, medidas de distancia de edición y medidas de implicación textual. Para características de sentimientos se consideran listas de palabras, palabras de marcadores de discurso y puntuación de sentimiento (p. ej. SentiWordNet). Utilizando el clasificador Random Forest identifican relaciones entre dos oraciones como: apoyo, ataque o ninguna. Reportan una exactitud del 77.5% utilizando un corpus con 854 pares de mensajes pertenecientes a medios sociales. Sin embargo, no reportan un nivel de acuerdo para el estudio de anotación, ya que aún está en curso, así como la falta de un análisis de las características utilizadas para representar a cada par de oraciones.

Otro método para identificar relaciones se propone en el trabajo de (Lawrence & Reed 2014), en el cual se utiliza la similitud semántica entre proposiciones secuenciales del mismo párrafo para determinar si tiene una conexión (apoyo / conflicto). Para obtener la similitud entre dos proposiciones utilizan Wordnet para determinar la distancia de cada palabra de la primera proposición con cada palabra de la segunda proposición. Esta similitud es inversamente proporcional al número de nodos a lo largo del camino más corto entre los synsets¹ de dichas palabras. Se toma al mayor de estos valores con el fin de vincular una palabra en la primera proposición con otra en la segunda proposición. Finalmente se calcula el promedio de los valores de cada palabra para obtener una puntuación entre 0 y 1. Dicha similitud se utiliza en conjunto con un umbral, el cual determinan heurísticamente y asignan en 0.2 para determinar si se encontró una conexión entre las dos proposiciones. Reportan una precisión del 82%, con un recuerdo del 56% y una medida F1 de 67% para la identificación de una conexión entre dos proposiciones. El conjunto de datos utilizado se basa en el corpus Araucaria (Reed et al. 2008). Utilizan solamente 78 argumentos completos,

¹ Son conjuntos de sinónimos que forman parte de la estructura de relaciones semánticas de WordNet.

con un total de 404 proposiciones. Los tipos de esquemas de argumentación usados son: opinión de experto y consecuencia positiva.

Otro enfoque para identificar las relaciones entre argumentos es la implicación textual (*Textual Entailment*). En el trabajo de Cabrio & Villata (2012) se observa un análisis sobre textos de debates tomados de Dbpedia, de los cuales se utilizaron 100 pares para entrenar la herramienta de implicación textual EDITS y 100 pares para probarla. Se midió la implicación entre 2 oraciones y si había implicación, se clasificaba como de apoyo, en caso contrario de ataque, con lo que se obtuvo un 75% de exactitud para dicha tarea de clasificación.

2.4 Evaluación de la argumentación

Otros trabajos han tocado el tema de la evaluación de la argumentación para retroalimentar numéricamente al alumno. La investigación de argumentos en ensayos ha tenido una amplia atención, ya sea para la identificación de argumentos o para la evaluación de la fuerza de argumentos y persuasión. En la investigación de Persing & Ng (2015) se realiza la evaluación de la fuerza de los argumentos en ensayos y se utiliza una escala de 4 niveles. El nivel cero indica que el texto no tiene argumentación y los 3 niveles superiores se utilizan para determinar si la argumentación es fuerte, media o débil. Para asignar la calificación se utilizó regresión lineal implementado en el software LIBSVM, con el cual un 61.8% de las veces asigna la calificación errónea, y el 38.2% de las veces la calificación correcta. A pesar de evaluar los ensayos a partir de características léxicas (p. ej. frases de transición), sintácticas (p. ej. n-grams de etiquetas de POS), postura con relación al tema, errores en argumentos y predicción de componentes argumentativos, su corpus de 1000 ensayos solo contiene las calificaciones de cada ensayo. Al realizar la predicción de componentes argumentativos sin un “ground truth”, no es posible validar dicha predicción.

Además la evaluación de ensayos argumentativos se ha realizado aplicando técnicas heurísticas para la identificación de argumentos, como se observa en Ong, Litman y Brusilovsky (2014), donde se hace con un sistema basado en 8 reglas heurísticas para identificar argumentos y 5 reglas para la evaluación de ensayos. Las etiquetas empleadas para las oraciones en el corpus fueron: *Opposes*, *Supports*, *Citation*, *Claim*, *Hypothesis* y *Current Study*. Las reglas heurísticas se crearon a partir del análisis intuitivo de 9 ensayos de un

corpus de 52 ensayos. Dicho corpus fue calificado en una escala de 5 puntos. En esta investigación, la falta de la anotación de cada componente argumentativo en el corpus impide realizar una validación clara de las reglas heurísticas y se observa que el sistema califica optimistamente los ensayos al tener un promedio en las calificaciones asignadas automáticamente de 3.42, mientras que los expertos indicaban un promedio en las calificaciones de 3.03.

Otro enfoque para la evaluación de argumentos en textos académicos es su organización (estructura del escrito) con el análisis de la secuencia de funciones discursivas y argumentativas observadas en párrafos (p. ej. con argumentos o sin argumentos) u oraciones (p. ej. premisas, conclusiones, ninguna). En la investigación de Persing et al. (2010) se estima la calificación de la organización de ensayos. Para ello realizan la identificación de párrafos con las siguientes funciones del discurso: introducción, cuerpo, conclusión y refutación. Para la identificación de oraciones en cada párrafo indican las funciones discursivas como: tema, transición, tesis, idea principal, elaboración, apoyo, conclusión, refutación, solución y sugerencia. Para realizar la tarea de identificación se basaron en palabras claves, en características estructurales y en el empalme de las palabras de contenido de la oración del tema. Para evaluar un ensayo, primeramente tomaron la secuencia de funciones del discurso identificadas, posteriormente buscaron los k-vecinos más cercanos en el conjunto de entrenamiento y finalmente se tomaron las calificaciones de dichos vecinos como base para realizar el cálculo de la calificación del ensayo analizado. Para realizar el cálculo se utilizó un regresor SVM con tres kernels: lineal, de cadena y de alineación. Combinando los tres kernels reportan una exactitud del 48.3% para la estimación de la calificación de la estructura de ensayos con una validación cruzada a 5 pliegues. El conjunto de datos utilizado para el experimento fue de 1003 ensayos del corpus ICLE (International Corpus of Learner English). Un enfoque similar se observa en el trabajo de Wachsmuth et al. (2015), en el cual se modela el flujo de patrones de sentimientos como una forma para la revisión de la argumentación en textos de opinión web. Con estos trabajos se observa que considerar la estructura del escrito para realizar una estimación de su calificación es un enfoque adecuado.

Por otra parte la clasificación del tipo de argumentación es otra tarea que se ha considerado para identificar la estructura de los argumentos. El uso de esquemas de argumentación para identificar el tipo de argumentación se observa en Song et al. (2014) para la

anotación y para la identificación de oraciones que incluyen preguntas críticas de los esquemas: Político, Causal y Ejemplo. El corpus utilizado consistió en 300 ensayos de 2 temas diferentes anotados por 4 anotadores con conocimientos de lingüística. El acuerdo entre anotadores fue bueno obteniendo un valor de Kappa desde 54.9% a 84.8% a excepción de la “definición del plan” la cual tuvo un valor de Kappa de 34.6%. Se utilizó aprendizaje automático con la herramienta SKLL usando el modelo de regresión logística penalizada L2 para la clasificación binaria de oraciones que incluyen preguntas críticas, con un valor de Kappa de 47.8%.

Como se ha observado en la mayoría de los trabajos analizados un elemento importante para la identificación de argumentos son los marcadores argumentativos o marcadores del discurso (p. ej. “por lo tanto”).

Utilizando marcadores del discurso como atributos para identificar conclusiones o premisas, Eckle-Kohler, Kluge & Gurevych (2015), trabajan sobre 88 documentos de noticias en alemán de temas relacionados a la educación e identificaron que la mayoría de los argumentos consistían en premisas y conclusiones adyacentes, relacionadas por una relación de ataque o apoyo. Para esto, se genera una representación vectorial con el grupo de marcadores discursivos de las 300 palabras más frecuentes de clase cerrada obtenidas del corpus Tiger, y utilizando un clasificador Naive Bayes se obtiene una medida F de 70.86% con respecto al anotador 1. Con esto se observa la importancia de los marcadores argumentativos y de las palabras vacías, para caracterizar textos argumentativos.

Como se puede observar son varias las técnicas utilizadas para tratar las tareas relacionadas a la identificación y evaluación de argumentos, tales como: reglas heurísticas, gramáticas libres de contexto (CFG), campos aleatorios condicionales, métodos de aprendizaje (p. ej. naive Bayes, SVM, EM), similitud textual e implicación textual. En nuestra investigación se propone un método el cual toca cada una de estas tareas con el propósito de brindar recomendaciones con respecto a la argumentación en los escritos de propuestas de investigación o tesis realizados por estudiantes de licenciatura. En dichas recomendaciones se observará información relativa a la identificación de texto argumentativo o no, identificación de premisas y conclusiones, identificación de relaciones entre dichos componentes, evaluación del nivel de argumentación e identificación el tipo de argumento empleado por el alumno: autoridad, causal, ejemplo, analogía o comparación. Como se observó, las dife-

rentes investigaciones exploran estas tareas pero no en forma global, ni para ofrecer sugerencias o recomendaciones al alumno.

Tabla 1. Resumen de trabajos relacionados

Trabajo	Acercamiento	Tarea	Tipo documento	Corpus	Métrica
(Florou et al. 2013)	Marcadores y verbos (DT)	1*	Mensajes foro debates	677 segmentos de texto	Medida F de 76.4%
(Moens et al. 2007)	Estructurales, léxicas y sintácticas (NB)	1*	Noticias, legal y foros	Araucaria (1899 oraciones)	Exactitud de 73.75%
(Nguyen & Litman 2015)	Palabras del dominio y argumentativas (LDA-SVM)	1	Ensayos	90 ensayos (1,673 oraciones)	Exactitud del 80%
(Stab & Gurevych, 2014)	Estructurales, léxicas, sintácticas y contextuales (SVM)	1+2	Ensayos	90 ensayos (1,673 oraciones)	Exactitud del 77.3% /86.3%
(Persing & Ng, 2016)	Estructurales, léxicas, sintácticas y marcadores (MaxEnt)	1+2	Ensayos	90 ensayos (1,673 oraciones)	Exactitud del 57.2% / 38.8%
(Mochales & Moens, 2011)	Estructura (CFG)	1+2	Textos legales	ECHR (763 premisas y 304 conclusiones)	Medida F : 67.27%(C)/64%(P) Exactitud(R): 60%
(Carstens & Toni, 2015)	Relaciones y opinión (RF)	2	Noticias	854 oraciones	Exactitud del 77.5%
(Cabrio & Villata, 2012)	Implicación textual	2	Debates	Debatepedia 200 pares componentes	Exactitud del 75%
(Ong et al. 2014)	Reglas heurísticas	(1)+3	Ensayos	52 ensayos calificación	Correlación rho 0.997
(Persing et al. 2010)	Secuencias de funciones discursivas (SVM)	3	Ensayos	1003 ensayos calificación	Exactitud del 48.3%
(Persing & Ng, 2015)	Léxicas, sintácticas, postura errores y componentes (SVM)	(1)+3	Ensayos	846 ensayos calificación	Exactitud de 38.2%
Propuesta	Estructurales, léxicas, sintácticas, semánticas y contextuales (método a proponer)	1+2+3	Tesis	930 párrafos / 1910 oraciones (a la fecha)	Exactitud de 77%

En la tabla 1 se presenta de forma resumida la comparativa de algunos de los trabajos mencionados y en el último renglón nuestra propuesta. La propuesta se diferencia del resto al considerar las tres tareas para ofrecer un mejor análisis de la argumentación (tarea 1 – clasificar componentes argumentativos, tarea 2 – detección de relaciones entre componentes, tarea 3 - evaluación de la argumentación). Además se trabaja con documentos académicos

micos como son las tesis y a la fecha se cuenta con un total de 930 párrafos anotados lo cual supera a las colecciones de otros autores.

3. Problemática

La presente investigación busca analizar y evaluar automáticamente el texto argumentativo en escritos académicos como lo son las propuestas de investigación o tesis de alumnos de nivel licenciatura. La redacción de este tipo de documentos implica que los alumnos plasmen sus ideas apropiadamente. Por lo regular se realizan numerosas revisiones a dichos escritos antes de obtener una versión final. Los profesores son los encargados de revisar la estructura, escritura y contenido de estos documentos.

Argumentar significa fundamentar con pruebas y evidencia lo que se pretende probar para persuadir al lector a través de razonamientos convincentes de la validez de nuestra aseveración (Argudin, 2005). Una buena argumentación en los escritos académicos es importante para transmitir claramente las ideas y convencer al lector de las afirmaciones realizadas. Por lo tanto, la característica argumentativa en los trabajos académicos es importante para la aceptación de la información de los mismos.

La identificación de argumentos es una tarea compleja que involucra conocer primeramente cuál es la aseveración (conclusión) propuesta por el estudiante y si dicha aseveración cuenta con evidencias (premisas) que sirvan de apoyo. Por ejemplo, en el elemento justificación de un trabajo de licenciatura del corpus a utilizar, se observa el siguiente párrafo anotado:

“[Dentro del departamento de Ingeniería en Sistemas Computacionales en la UDLA, se necesita facilitar el manejo de los asuntos departamentales]/C1 [ya que pueden llegar a consumir demasiado tiempo]/P1 [debido a que se tiene una gran cantidad de información a procesar]/P2”

La aseveración (conclusión) en el ejemplo se indica con C1, la cual es apoyada por la premisa P1, que a su vez es apoyada por la premisa P2. Dichas premisas utilizan los marcadores argumentativos para indicar explícitamente el apoyo: “ya que” y “debido a que”. Con el apoyo de dichas premisas se entiende la razón por la cual es válida la aseveración

realizada por el alumno al indicar la necesidad de facilitar el manejo de los asuntos del departamento.

Es importante hacer notar que una característica de los textos académicos es la argumentación explícita, pues conviene marcar que en la exposición hay un razonamiento, ya que estamos en el terreno de la exposición de conocimientos académicos (López, 2003). Es por esta razón que los marcadores argumentativos (p. ej. “ya que” y “debido a que”) se presentan por lo regular en las propuestas de investigación y tesis para indicar puntualmente la relación de apoyo que ofrece una determinada premisa. Sin embargo, muchas veces encontramos trabajos con argumentos débiles, los cuales a falta de premisas de apoyo y una adecuada estructura terminan siendo confusos.

La tarea de identificación de argumentos no es trivial debido a la vaguedad y ambigüedad que se observa en el lenguaje natural, es difícil identificar si un segmento de texto encaja en el patrón con el cual debe cumplir un argumento. Por ejemplo, un alumno podría fácilmente confundir un argumento con una explicación. Esto se observa si se utilizan únicamente palabras claves (p. ej. “debido a”) para identificar argumentos, ya que podríamos recuperar textos en los cuales se brinda una explicación, sin embargo no se argumenta ninguna postura por parte del autor. Otro ejemplo son los argumentos del tipo de autoridad, un tipo comúnmente observado en los trabajos de investigación. En estos argumentos se debe tener una premisa y una conclusión, donde la premisa que apoya a dicha conclusión incluye la referencia a un artículo, libro u experto. Sin embargo, si se buscarán en la red todos los documentos con una referencia, no necesariamente todos serían argumentativos ya que muchos de ellos serían explicaciones u otro tipo de textos. Si bien son una herramienta para identificar el tipo de argumento (p. ej. de autoridad) no necesariamente identifican la argumentación. Es por ello que es necesario un análisis tanto de las características léxicas como estructurales, morfosintácticas y contextuales para lograr identificar los patrones asociados con los componentes y relaciones de los argumentos.

Esta investigación busca ayudar a los alumnos en la identificación y evaluación de los argumentos contenidos en sus tesis y propuestas de investigación desde un enfoque computacional. De acuerdo a estudios realizados como Bañalez et al. (2015), se observa una deficiencia en la habilidad de escribir argumentos entre algunos estudiantes, dado que es una tarea compleja. Si se obtiene una estructura de argumentos fuerte se proporciona

claridad y soporte a las afirmaciones presentadas por el alumno. Con esto se puede apoyar el proceso de revisión realizado por el profesor. Además, en la revisión realizada a la literatura no se ha encontrado un método para identificar y evaluar párrafos argumentativos en tesis y propuestas de investigación que brinde apoyo tanto al alumno como al profesor en los procesos de edición y revisión.

4. Preguntas, objetivos y contribuciones

En la presente propuesta se pretende contestar a las siguientes preguntas de investigación:

¿De qué manera se pueden aplicar las técnicas de procesamiento del lenguaje natural para la identificación automática de los componentes argumentativos en escritos académicos de investigación?

¿Qué enfoques de procesamiento del lenguaje natural pueden ser aplicados para identificar las relaciones entre componentes argumentativos y cómo se puede modelar la estructura de estas relaciones?

¿Cómo se puede evaluar automáticamente la argumentación existente en escrito académico de investigación considerando sus componentes y relaciones?

4.1 Objetivo general

Desarrollar métodos para extraer y evaluar los argumentos en escritos académicos de investigación aplicando técnicas de procesamiento del lenguaje natural para generar representaciones que permitan modelar su información lingüística argumentativa, alcanzando niveles de acuerdo aceptables comparados con los revisores humanos (anotadores), con la finalidad de apoyar al alumno en la escritura de su texto, así como de asistir al profesor en su proceso de revisión.

4.2 Objetivos específicos

- Anotar y analizar un corpus de escritos académicos de investigación para identificar sus características argumentativas.
- Desarrollar un método para obtener la representación de componentes argumentativos con características que permitan su identificación y extracción.

- Formular una técnica para determinar las relaciones entre los componentes argumentativos y modelar su estructura.
- Diseñar e implementar un procedimiento para evaluar el nivel de argumentación en los textos de estudiantes que aproveche la información de las representaciones generadas a partir de la identificación de componentes y relaciones argumentativas.

4.3 Contribuciones esperadas

A través de esta investigación doctoral se espera obtener las siguientes contribuciones:

- Un método para representar los escritos académicos de investigación para la identificación y extracción de sus componentes argumentativos.
- Una técnica para la determinación y modelado de las relaciones entre componentes argumentativos basada en la identificación de componentes argumentativos en textos de estudiantes.
- Un procedimiento para la evaluación de textos argumentativos en escritos académicos de investigación basado en las características de sus componentes y relaciones.
- Una colección anotada de escritos académicos de investigación para el análisis de argumentos.

4.4 Hipótesis

La evaluación automática de párrafos argumentativos en escritos académicos de investigación es posible utilizando determinadas características léxicas, sintácticas, semánticas, contextuales y estructurales de sus componentes y relaciones, con las técnicas actuales de procesamiento del lenguaje natural.

5. Metodología

En esta sección se presenta la metodología propuesta para alcanzar los objetivos planteados. La metodología se describe para las siguientes tareas:

1. Construcción de un corpus argumentativo de escritos académicos de investigación (propuestas de investigación y tesis).
2. Desarrollo de un método para la identificación y extracción automática de componentes argumentativos.
3. Formulación de una técnica para determinar las relaciones entre componentes argumentativos y su estructura.
4. Diseñar un procedimiento para la evaluación de nivel argumentativo de los textos de investigación de estudiantes.

Estas tareas y su metodología se desarrollan a detalle a continuación:

Tarea 1. Construcción de corpus argumentativo de escritos académicos de investigación (propuestas de investigación y tesis)

La metodología a seguir para la tarea 1 es:

1. Revisión de trabajos relacionados con la construcción de corpus argumentativo.
2. Investigar las características de las guías de anotación desarrolladas por otros autores.
3. Desarrollar nuestra guía de anotación a partir de las propuestas de otros autores y las características propias de los escritos académicos.
4. Realizar pruebas preliminares de anotación entre diferentes candidatos para validar la correcta detección de los elementos.
5. Realizar el estudio de anotación con los anotadores seleccionados.
6. Evaluar el nivel de acuerdo entre anotadores.

Las sub-tareas para la tarea 1 son:

- a) **Identificar los conjuntos de datos con escritos académicos de investigación en español de diferentes niveles académicos.** Estos pueden ser colecciones de tesis de nivel superior (p. ej. licenciatura, maestría y doctorado) o bien reportes técnicos, los cuales tienen una estructura similar. El propósito es encontrar colecciones en donde

puedan existir patrones argumentativos para caracterizarlos a través de su información textual.

1. Existen diferentes recursos lingüísticos utilizados por investigadores del área de minería de argumentos, los cuales en su mayoría son de creación propia para un tipo de dominio (p. ej. educación, legal, biomédica) y tipo de documento en particular (p. ej. ensayos, textos de la corte, textos de foros). Una de las colecciones más utilizadas entre los investigadores para identificar la presencia de argumentos es el corpus Araucaria (Katzav, Reed & Rowe, 2004), el cual cuenta con varios tipos de documentos (p. ej. registros del parlamento, periódicos, resúmenes judiciales y foros de discusión) para el cual se tienen anotados premisas y conclusiones así como el esquema de argumentación utilizado. Sin embargo, en dicho corpus no se informa de un nivel de acuerdo entre anotadores, lo cual hace poco fiable el estudio. Además, en dicha colección se carece de escritos académicos de investigación con información de sus componentes, relaciones y evaluación argumentativa, los cuales son el objeto de estudio de la presente investigación. Por dicha razón es necesario construir una colección con textos de investigación de estudiantes para realizar una correcta evaluación de los enfoques planteados en la investigación doctoral.

b) **Construcción de un corpus argumentativo.** Para construir el corpus se planean realizar las siguientes actividades:

1. Identificar colecciones de tesis o reportes de investigación de nivel superior en idioma español. La colección contendrá trabajos de nivel TSU, licenciatura, maestría y doctoral. La razón para considerar diferentes niveles académicos es para utilizar a los trabajos de nivel doctoral y maestría como punto de comparación (mejores escritos) con respecto a los trabajos de nivel licenciatura y TSU.
2. Identificar las secciones de escritos académicos de investigación con mayor cantidad de argumentos. Para realizar la selección se propone utilizar marcadores argumentativos para encontrar párrafos con argumentos (Lawrence & Reed 2014). Con esto se busca encontrar secciones que cuenten con un mayor número de razonamientos por parte del autor. En el tipo de documentos de artículos

científicos se utilizan las secciones de introducción y discusión (Kirschner, Eckle-Kohler & Gurevych, 2015).

3. Desarrollar una guía de anotación con ejemplos de diferentes tipos de estructuras de argumentos. Indicar en la guía el esquema de anotación para la identificación de los aspectos argumentativos concernientes al presente estudio.
4. Realizar el estudio de anotación de las secciones con mayor argumentación con la ayuda de profesionales con antecedentes en estudios en lingüística, filosofía o áreas similares. Seleccionar por lo menos 2 anotadores para realizar el estudio, con la finalidad de determinar un nivel de acuerdo entre anotadores. El resultado del estudio será: un conjunto de componentes argumentativos identificados como premisa o conclusión, un conjunto de pares de componentes argumentativos con relación de ataque o apoyo, un conjunto de párrafos evaluados con un nivel de argumentación de 0 a 3 y un conjunto de párrafos con el tipo de argumentación empleada, tales como: ejemplo, analogía, autoridad, causal o comparación.

Tarea 2. Desarrollo de un método para la identificación y extracción automática de componentes argumentativos.

La metodología a seguir para la tarea 2 es:

1. Identificar los métodos para la extracción de argumentos en textos similares a los escritos académicos de investigación.
2. Analizar las características y técnicas utilizadas por otros autores aplicables a los escritos académicos en español.
3. Seleccionar un conjunto de características y técnicas adecuados para la detección de argumentación y extracción de componentes argumentativos (premisa / conclusión).
4. Evaluar las representaciones y métodos con respecto al corpus creado.

En esta tarea emplearemos un enfoque de aprendizaje supervisado, utilizando nuestra colección anotada se construye una representación con las características más significativas de los textos (párrafos, oraciones y cláusulas). Se evalúa la tarea utilizando un clasificador para verificar sus medidas de desempeño. Con la finalidad de analizar la argumentación en

diferentes niveles de granularidad textual, se implementarán métodos para la detección de argumentación en los siguientes niveles: párrafo, oración y cláusula. Posteriormente se procederá a la identificación de componentes argumentativos como: premisa, conclusión o ninguno. A continuación se presentan las sub-tareas a realizar:

- a) **Analizar los métodos para la identificación de argumentos en textos similares a los escritos académicos de investigación.** Los tipos de textos cercanos son: ensayos (Persing & Ng, 2016; Stab & Gurevech, 2014), artículos científicos (Teufel & Moens, 2002), textos legales (Moens et al. 2007; Mochales & Moens 2008).
- b) **Clasificar los párrafos, oraciones y cláusulas como argumentativos o no argumentativos.** Para esta tarea se consideran los siguientes tipos de características: a) estructural: en la cual se considera la longitud, ubicación y puntuación, b) morfosintáctica: en la cual se considera que al momento de argumentar se utilizan ciertos modos y tipos de los verbos, c) léxicas: en las cuales se considera las secuencias de palabras (n-gramas), así como la identificación de partes de la oración como los verbos, adverbios y auxiliares modales, d) marcadores argumentativos: son un conjunto de expresiones utilizadas por ejemplo para justificar un razonamiento dentro de un argumento (p. ej. "debido a ello"), e) contextual: considerar las características de las oraciones adyacentes.
- c) **Clasificar las unidades argumentativas como premisas, conclusiones o ninguna.** Los enfoques observados para construir representaciones para la clasificación de componentes argumentativos aprovechan las características consideradas en el método de clasificación anterior y adicionan otras características como el número de sub-cláusulas y profundidad del árbol sintáctico, la ubicación absoluta en la sección, o las estructuras retóricas identificadas en las oraciones adyacentes.

Tarea 3. Formulación de una técnica para determinar las relaciones entre componentes argumentativos y su estructura.

La metodología a seguir para la tarea 3 es:

1. Identificar los métodos para el modelado de las relaciones en argumentos.

2. Analizar las técnicas y características utilizadas por otros autores aplicables a escritos académicos en español (relacionales y atributos del texto).
3. Seleccionar un conjunto de características y técnicas adecuados para la representación de pares de segmentos de texto para el modelado de las relaciones entre componentes.
4. Evaluar las representaciones y métodos en el corpus construido.

Para la tarea 3, inicialmente es necesario realizar un emparejamiento de los componentes argumentativos. Posteriormente se requiere generar representaciones para los componentes argumentativos y determinar las relaciones de ataque o apoyo existente entre los mismos. Finalmente una vez identificadas las relaciones, hay que construir una estructura de grafo o árbol para modelar al argumento analizado. Las sub-tareas a realizar son las siguientes:

a) **Identificar relaciones entre componentes.** En esta tarea, consideramos los siguientes enfoques como los mejores candidatos:

1. **Clasificar la relación de pares de componentes argumentativos utilizando las características textuales de cada componente individual,** para esto es necesario crear una representación vectorial de las características lingüísticas más representativas para identificar las relaciones de apoyo o ataque (Stab & Gurevych, 2014). Se generan todos los posibles pares de componentes argumentativos para cada párrafo. Además, se crea una representación para cada componente de los pares a analizar y posteriormente se unen para realizar la tarea de clasificación.
2. **Identificar relaciones entre pares de componentes argumentativos utilizando medidas de similitud** para identificar el apoyo de un componente a otro. Dentro de dicho acercamiento se utiliza similitud textual y distancia de edición. Con medidas de similitud léxica como: similitud coseno, solapamiento de palabras y coeficiente Jaccard; Medidas de distancia de edición como: Levenshtein y Jaro-Winkler. Así como utilizando herramientas como EDITS (Kouylekov & Negri, 2010) de código abierto utilizada para el reconocimiento de implicación textural entre dos proposiciones (Cabrio & Villata, 2012). Otro acercamiento es calculando la similitud semántica entre dos proposiciones utilizando recursos ex-

ternos (p. ej. Wordnet, Polyglot-word2vec-contexto, LDA-tópicos) para determinar la distancia de cada palabra de la primera proposición con cada palabra de la segunda proposición (Lawrence & Reed 2014).

- b) **Construcción de estructura de Argumento.** Partiendo de las relaciones obtenidas de la sub-tarea anterior, diseñar un algoritmo para la construcción de la estructura del argumento por párrafo. Las formas utilizadas para dicha estructura son: árboles (Peldszus & Stede, 2015; Stab & Gurevych, 2016) y grafos (Petasis & Karkaletsis, 2016)

Tarea 4. Diseñar un procedimiento para la evaluación del nivel argumentativo de los escritos académicos de investigación.

La metodología a seguir para la tarea 4 es:

1. Identificar los métodos y características para la evaluación de argumentos en textos similares a los escritos académicos en español.
2. Analizar las características y técnicas que aprovechen la estructura del argumento para la evaluación de la argumentación en dichos escritos.
3. Evaluar el desempeño de los métodos propuestos utilizando el corpus creado.

En la tarea 4, la evaluación se realiza a nivel de párrafo para determinar un nivel de argumentación. La escala a utilizar es del 0 al 3, en donde se indica 0 para ausencia de argumentación, 1 para un argumento débil, 2 para un argumento medio y 3 para un argumento fuerte. Para realizar este paso nos apoyamos en la información de los componentes identificados y las relaciones detectadas. Se genera una representación de las características de los componentes y se propone utilizar un algoritmo de aprendizaje para estimar la calificación de los argumentos: SVM (Persing & Ng, 2015), Bayesiano (Chen et al. 2010), basado en reglas (Ong, Litman & Brusilovsky, 2014). Las sub-tareas a completar son las siguientes:

- a) **Clasificar el tipo de argumentación en el párrafo** utilizando técnicas de aprendizaje computacional tales como: árboles de decisión (Feng & Hirst, 2011) o Naive Bayes (Lawrence & Reed, 2014). El objetivo es identificar el tipo de argumento, que puede ser: ejemplos, analogía, autoridad, causal y comparación.

b) **Diseñar un método para para estimar el nivel de argumentación** en el párrafo. Existen métodos de regresión que utilizan algoritmos de aprendizaje para realizar dicha estimación (p. ej. regresor SVM) (Persing & Ng, 2015). Se analizarán las características estructurales, léxicas, sintácticas y gramaticales más representativas para la tarea. Particularmente se considera realizar un análisis de los componentes argumentativos identificados (premisas/conclusiones), relaciones entre dichos componentes (ataque/apoyo/ninguno), tipo de argumento identificado, errores frecuentes al escribir argumentos (p. ej. conclusión sin premisa) y estructura del argumento (Persing et al. 2010). Además de considerar representaciones basadas en tópicos y contexto como LSA (Latent Semantic Analysis) (Venegas, 2003) y Word2Vec (Mikolov et al., 2013).

Las tareas descritas se presentan gráficamente en el diagrama de la Figura 2. Se observan las 4 tareas principales, las cuales incluyen las actividades más representativas.

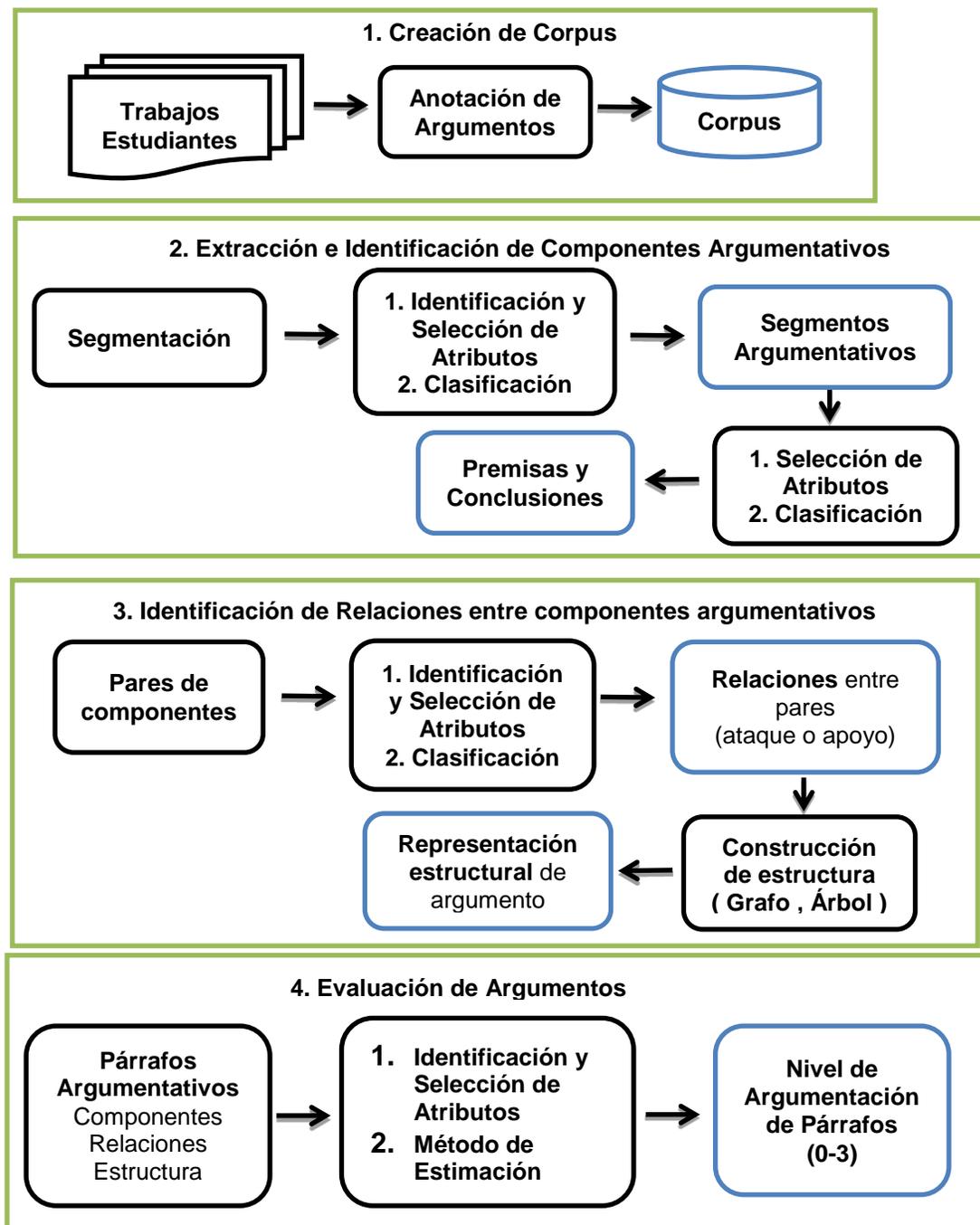


Fig. 2. Diagrama de metodología

Modelo de Análisis de Argumentos

Nuestro acercamiento a la solución sería basándonos en ciertos procesos de la metodología empleada en minería de argumentos (Peldszus y Stede, 2013). En la Figura 3, se esquematiza lo que podría ser el proceso para el análisis del texto del estudiante.

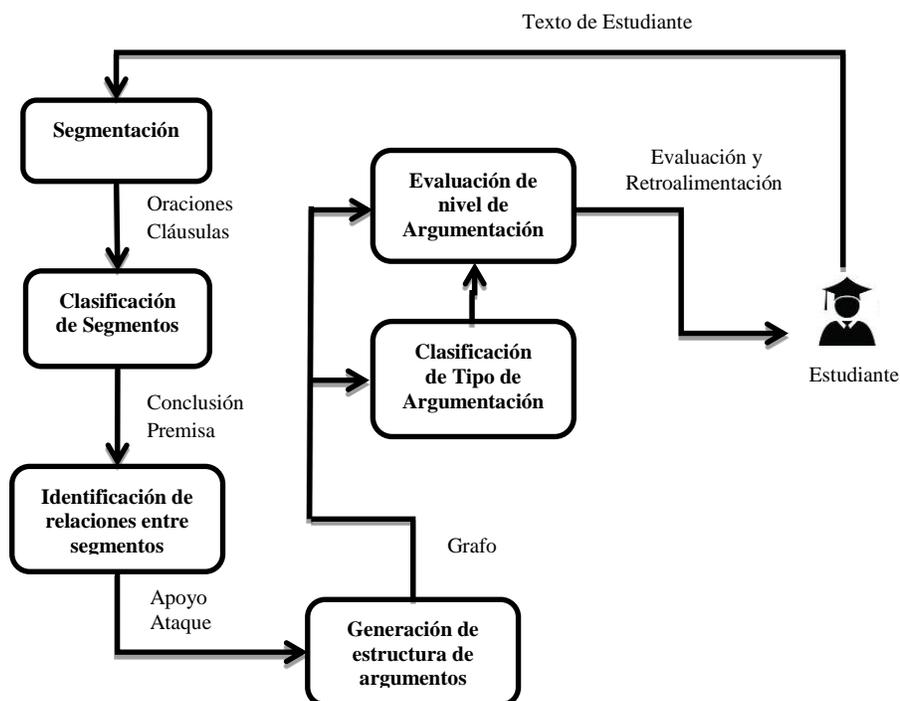


Fig. 3. Modelo de Análisis de Argumentos

Primeramente identifica los párrafos argumentativos y posteriormente segmenta el texto en unidades mínimas de argumentación para textos académicos, que de acuerdo a la revisión realizada al corpus, la segmentación sería en cláusulas u oraciones. En segundo lugar es necesario obtener una representación vectorial de los segmentos para posteriormente clasificar cada segmento de acuerdo a su rol argumentativo, ya sea una premisa o conclusión. Para realizar esta tarea es posible emplear un clasificador (SVM, NB, DT, LR), entrenado previamente con el corpus. A continuación, para identificar las relaciones entre segmentos de texto en un párrafo, es necesario emparejar premisas y conclusiones para detectar el tipo de relación que manifiestan, nuevamente utilizando un clasificador para dicha tarea. Una vez identificadas las relaciones se modelaran en una estructura (grafo o árbol) adecuada para las características de los escritos académicos. Posteriormente se utilizará un clasificador para identificar el tipo de argumentación presente en el párrafo aprovechando

la estructura generada. Finalmente, aprovechando la estructura obtenida por cada párrafo se generará una nueva representación vectorial que capture las características argumentativas del párrafo para estimar el nivel argumentativo del mismo utilizando un algoritmo de aprendizaje computacional. Nuestro reto es desarrollar una serie de métodos que nos permitan construir representaciones vectoriales adecuadas para las características léxicas, sintácticas, semánticas y contextuales del texto académico, las cuales brinden la información necesaria para identificar satisfactoriamente los componentes argumentativos (premisas y conclusiones), así como sus relaciones (ataque y apoyo), además del nivel y tipo de argumentación de cada uno de los párrafos. El producto final es proveer una retroalimentación completa al alumno para apoyarlo en el mejoramiento de la argumentación presente en sus textos.

6. Trabajo realizado y resultados preliminares

En este apartado se describe el trabajo realizado del periodo de agosto 2015 a junio 2016.

1. Identificar los conjuntos de datos con escritos académicos de investigación en español de diferentes niveles académicos (parte de la primer tarea indicada en la metodología, inciso a). Para llevar a cabo esta sub-tarea se utilizó el recurso lingüístico *Coltypi* (Colección de tesis y propuestas de investigación) (González-López & López-López, 2015) creado en el laboratorio de Tecnologías del Lenguaje del INAOE. Las tesis y reportes técnicos de dicho recurso son de diferentes niveles académicos del área de computación y TICs.
2. Construcción de un corpus argumentativo (parte de la primer tarea de la metodología, inciso b). Para esta sub-tarea, una vez identificada la colección a ser utilizada, se procedió a analizar la colección para identificar las secciones que por su naturaleza resultan ser las más argumentativas. Posteriormente se elaboró una guía de anotación. Con dicha guía se anotó una muestra para su análisis en los siguientes pasos de la metodología.
3. Realizar un primer estudio de anotación con lo cual se completó esta tarea por parte de un anotador de las secciones de planteamiento del problema y justificación (parte de la primer tarea de la metodología, inciso b).
4. Clasificación de párrafos y oraciones como argumentativos o no argumentativos (parte de la segunda tarea indicada en la metodología, inciso a y b). Utilizando la muestra anotada del punto anterior, se construyeron diferentes representaciones vectoriales para analizar las características léxicas y sintácticas de los párrafos para la tarea de identificación de párrafos y oraciones argumentativas. Posteriormente se analizaron las secciones de planteamiento del problema y justificación anotados.

6.1 Identificar los conjuntos de datos con escritos académicos de investigación

Al revisar los corpus disponibles con argumentos anotados identificamos al corpus Araucaria (Reed et al., 2008). Sin embargo, este recurso se encuentra en idioma inglés y cuenta con información de documentos como noticias, legales y foros de discusión, los cuales no son útiles para el propósito de la presente investigación. Otro recurso disponible a la comunidad científica está basado en ensayos académicos (Stab & Gurevych, 2014). Sin

embargo, dicho recurso también se encuentra en idioma inglés. Otras diferencias se observan en la estructura, la cual es diferente a la utilizada en las tesis, así como la manera informal y subjetiva en la cual se encuentran redactados dichos ensayos. Finalmente se identificó la colección de documentos *Coltypi* (González-López & López-López, 2015) la cual está constituida por un total de 468 tesis y propuestas de investigación en idioma español de nivel pregrado (TSU y Licenciatura), así como posgrado (Maestría y Doctorado) del área de computación y TICs (Tecnologías de la Información y Comunicación). Por las características observadas en la colección se decidió utilizarla para el presente estudio. La cantidad de documentos por nivel académico se detalla en la tabla 2.

Tabla 2. Documentos por nivel académico de colección *Coltypi*

Grado	Documentos
Doctorado	59
Maestría	181
Licenciatura	150
TSU	78
Total	468

6.2 Construcción de corpus argumentativo

Con la colección identificada se procedió a realizar los siguientes pasos:

6.2.1 Identificar las secciones de escritos de investigación con mayor cantidad de argumentos

Al realizar una revisión de artículos de argumentación académica, se encontró que las secciones de planteamiento del problema, justificación y conclusiones son consideradas altamente argumentativas por lo que se procedió a realizar un análisis de las mismas en el corpus para corroborar dicha aseveración (López, 2003).

En el trabajo de (Lawrence & Reed, 2014), se observa el uso de marcadores argumentativos para identificar texto argumentativo con una precisión del 89% y un recuerdo del 4%. No obstante que se recupera una pequeña parte de los datos, la precisión es suficiente como para dar una idea de la presencia de argumentación en cada sección. Como se observa en la tabla 3, fue en la sección de conclusiones en la cual se encontraron 731 párrafos con argumentación, lo cual indica una proporción del 61% con respecto al total de párrafos en los trabajos de nivel licenciatura. Por otra parte se observa que en la sección de

justificación de nivel doctorado un 61% de los párrafos incluye argumentación. Para la sección de TSU se observa que la sección con mayor número de párrafos con argumentación es la de planteamiento del problema con un 62% de párrafos argumentativos. Con esto se evidencia la gran cantidad de argumentos en los diferentes niveles académicos.

Al analizar el corpus se observa que cada sección contiene en promedio 11 oraciones por elemento. Cada oración contiene 35 palabras en promedio de un total de 398 palabras por sección. El tamaño de las oraciones para nivel licenciatura es de 38 palabras por oración lo cual dificulta su lectura, en contraste con el nivel doctoral en donde se tiene un promedio de 30 palabras por oración. Basados en esto consideramos a los escritos doctorales mejores y se consideran como una buena referencia.

Tabla 3. Párrafos con argumentación en corpus

		Con Argumentos	Total	Porcentaje
Doctorado	Planteamiento del Problema	124	245	51%
	Justificación	56	92	61%
	Conclusión	194	449	43%
Maestría	Planteamiento del Problema	206	392	53%
	Justificación	203	375	54%
	Conclusión	707	1414	50%
Licenciatura	Planteamiento del Problema	150	269	56%
	Justificación	180	313	58%
	Conclusión	446	731	61%
TSU	Planteamiento del Problema	95	153	62%
	Justificación	99	212	47%
	Conclusión	205	363	56%

En la tabla 4 se muestra el resumen de la cantidad de párrafos y oraciones con argumentación para las secciones de planteamiento del problema, justificación y conclusiones. En esta tabla se observa a la sección de conclusiones en la cual se encontraron 1552 párrafos con argumentación, lo cual indica una proporción del 54% con respecto al total de párrafos en dicha sección. Para las secciones de justificación y planteamiento del problema se identificaron proporciones similares en sus párrafos argumentativos con 54% y 52% respectivamente.

Tabla 4. Párrafos y oraciones con argumentación en corpus

Secciones	Párrafos con argumentación	Total de Párrafos	Oraciones con argumentación	Total de Oraciones
Planteamiento del problema	575	1059	777	2176
Justificación	538	992	653	1696
Conclusiones	1552	2957	1972	5446
Total	2665	5008	3402	9318

6.2.2 Desarrollar una guía de anotación

Para la realización del estudio de anotación se elaboró una guía de anotación. En dicha guía se describen las diferentes estructuras argumentativas. Para el esquema de anotación consideramos dos componentes argumentativos, de acuerdo al modelo conclusión/premisa de Freeman (2011), así como los tipos de relación entre componentes: apoyo y ataque. Además, se definieron los tipos de argumentación y una escala para establecer el nivel de argumentación. Para cada una de estas secciones, se proporcionaron ejemplos de textos de tesis y propuestas de investigación debidamente anotados para guiar a los anotadores en su tarea. Al final de dicha guía se incluye el procedimiento que deben seguir los anotadores para realizar la tarea.

Tabla 5. Nivel de argumentación en párrafo

Nivel	Escala	Descripción
0	No tiene	No hay argumento.
1	Débil	No es un argumento completo. Se tiene una conclusión sin premisa.
2	Media	Argumento con un razonamiento
3	Fuerte	Argumentos con dos o más razonamientos (apoyo o ataque)

El procedimiento en la guía de anotación indica los siguientes pasos. Primeramente se indica al anotador leer el título y el objetivo de la tesis. Posteriormente una vez comprendido el contexto de la tesis y el tema del cual trata la tesis, se procede a identificar si el texto incluye alguna conclusión o aseveración. A continuación se identifican las ideas que apoyan a dicha conclusión y se anotan como premisas de apoyo. Se indicó a los anotadores que de preferencia anotarán las oraciones o cláusulas correspondientes a una conclusión o premisa. Para indicar una premisa se utilizaron corchetes agregando al final /P, esto es:

[texto de premisa]/P. Para el caso de las conclusiones, de forma similar se encerró el texto con corchetes finalizando con /C, es decir: [texto de conclusión]/C. Finalmente se solicitó al anotador indicar el tipo de argumento observado en el párrafo, de entre los más utilizados por los estudiantes, que son: ejemplos, analogía, autoridad, causal y comparación (Weston, 2005). Además los anotadores deben calificar los párrafos de acuerdo a su nivel de argumentación, apoyados en los criterios de la tabla 5.

Como se observa en la tabla 5, se asigna una calificación de cero (0) en caso de que no se cuente con ningún texto argumentativo, es decir se trata quizás de una descripción o bien de una definición. Para cuando se identifica alguna argumentación la escala varía del uno a tres, con valor de uno (1) en caso que el alumno haya escrito una conclusión es decir alguna aseveración o presunción pero sin una premisa que apoye sus afirmaciones, para el caso del nivel dos (2) el alumno presenta una conclusión con una premisa de apoyo, en el caso de tenerse dos o más premisas de apoyo se considera el argumentos como fuerte y se asigna un nivel de tres (3). De esta forma se evaluaron los párrafos en el corpus analizado.

La anotación se realizó con la ayuda de profesionales con antecedentes en estudios en lingüística, filosofía o áreas similares. Con este esquema se les indicó a los anotadores la forma de identificar las estructuras comunes de argumentos, así como las conclusiones y premisas, además de las relaciones entre sus elementos. Con la identificación de dicha información es posible indicarle al estudiante la falta de conclusiones en sus escritos o bien la falta de premisas de apoyo para dichas conclusiones, utilizando un sistema de soporte a la escritura

6.2.3 Realizar el estudio de anotación

Muestra con un anotador

En la primera etapa del estudio de anotación se utilizó una muestra de 46 secciones de planteamiento del problema, justificación y conclusión. La muestra de 224 párrafos se anotó primeramente por un anotador con conocimientos en formulación y revisión de argumentos utilizando la guía de anotación desarrollada. En la tabla 6 se presenta la distribución de instancias entre las clases, en la cual se observa que 127 párrafos son argumentativos con una proporción del 56.7% y con 97 párrafos sin argumentación el 43.3%.

Tabla 6. Distribución de instancias entre clases

Párrafos con argumentos	Párrafos sin argumentos
127 (56.7%)	97 (43.3%)

Los argumentos identificados en la muestra son de tipo: ejemplos, analogía, autoridad, causal y comparación. Como se observa en la Tabla 7 el tipo de argumento más abundante son los argumentos de tipo causal con 95 párrafos, seguido de los argumentos por autoridad con 20 párrafos, ello debido a que generalmente en los párrafos académicos se pretende apoyar una idea de forma causal y además basándose en citas de autores reconocidos del área para proporcionar mayor soporte.

Tabla 7. Tipos de Argumentos en Muestra

	Párrafos	Porcentaje
Autoridad	20	14%
Ejemplo	18	13%
Causal	95	66%
Comparación	3	2%
Analogía	6	4%

Al revisar los argumentos por autoridad en la muestra se observan las siguientes características. Dichos argumentos se apoyan en otras referencias —personas, organizaciones u obras de referencia más documentadas— para que nos expliquen gran parte de lo que necesitamos saber sobre el mundo (Weston, 2005). Este tipo de argumentos son comunes en el ámbito académico, ya que generalmente el autor se apoya en las publicaciones realizadas por investigadores del área para sustentar sus aseveraciones. A continuación se presentan un ejemplo de este tipo identificado en trabajos académicos, en el cual se puede observar que se hace referencia a un autor de donde se toma la información de la premisa P que apoya a la conclusión C.

Ejemplos:

[Actualmente nos encontramos en la denominada Sociedad Red]/C que, [de acuerdo a Castells (2000), es una sociedad que se generó a partir de la revolución tecnológica de la información y el florecimiento de las redes sociales.]/P

Muestra con tres anotadores

Posteriormente se realizó la anotación por parte de 2 anotadores más con conocimientos en formulación y evaluación de argumentos. Se anotaron 226 párrafos, con lo cual se incluyeron 2 párrafos más para su análisis. Con la información de tres anotadores se identificaron los párrafos en los cuales la mayoría indicaban la existencia de argumentación. En la tabla 8 se presenta la distribución de instancias entre las clases, en la cual se observa que 141 párrafos son argumentativos con una proporción del 62.4% y con 85 párrafos sin argumentación el 37.6%. Como se observa se identificaron un mayor número de párrafos argumentativos considerando la votación mayoritaria de los anotadores, con 14 párrafos más.

Tabla 8. Distribución de instancias entre clases

Párrafos con argumentos	Párrafos sin argumentos
141 (62.4%)	85 (37.6%)

La tarea de identificación de párrafos argumentativos en textos académicos es compleja, esto se observa al analizar el nivel de acuerdo con la medida Kappa (Cohen, 1960) alcanzado entre anotadores presentado en la tabla 9, dicho nivel cae en el rango “justo” (fair 0.21-0.40) para dos pares de anotadores (Landis y Koch, 1977). Con lo cual constatamos la dificultad de dicha tarea.

Tabla 9. Nivel de acuerdo Kappa entre anotadores

Anotador 1 y 2	Anotador 1 y 3	Anotador 2 y 3
0.140	0.243	0.251

Al analizar las oraciones se obtuvo un total de 403 oraciones, de las cuales el 69% son argumentativas, como se observa en la tabla 10.

Tabla 10. Distribución de instancias entre clases

Oraciones con argumentos	Oraciones sin argumentos
278 (69%)	125 (31%)

Anotación a sección planteamiento del problema

Como parte del análisis a la sección de “Planteamiento del problema” se seleccionaron 100 trabajos para su anotación. Se anotaron 397 párrafos de los cuales, como se observa en la tabla 11, el 51.4% son argumentativos y el restante 48.6% fueron anotados como sin argumentos. Además se observa un total de 905 oraciones, con 58.2% argumentativas.

Tabla 11. Distribución de instancias entre clases

Párrafos	
Con argumentos	Sin argumentos
204 (51.4%)	193 (48.6%)
Oraciones	
Con argumentos	Sin argumentos
527 (58.2%)	378 (41.8%)

Anotación a sección justificaciones

Como parte del análisis a la sección de “Justificación” se seleccionaron 100 trabajos para su anotación. Se anotaron 307 párrafos de los cuales, como se observa en la tabla 12, el 56% son argumentativos y el restante 44% fueron anotados como sin argumentos. Además se observa el total de oraciones con 602 oraciones, de las cuales 372 oraciones tienen argumentos lo cual corresponde al 61.8%.

Tabla 12. Distribución de instancias entre clases

Párrafos	
Con argumentos	Sin argumentos
172 (56%)	135 (44%)
Oraciones	
Con argumentos	Sin argumentos
372 (61.8%)	230 (38.2%)

6.3 Clasificar segmentos de texto como argumentativo o no argumentativo

6.3.1 Configuración experimental

Se abordó el problema como clasificación binaria, es decir para cada texto se identificó si es argumentativo o no es argumentativo. Para determinar la eficacia de cada clasificador, se utilizó una validación cruzada a 10 pliegues.

Para realizar el proceso de clasificación, se utilizó el software de minería de datos Weka (Hall et al., 2009). Los clasificadores utilizados son *Support Vector Machine* (SVM) (Stab & Gurevych, 2014), *Naive Bayes* (NB)(Moens et al. 2007) y *Decision Tree* J48 (DT)(Florou et al. 2013), debido a que estos clasificadores han sido empleados anteriormente en minería de argumentos. A continuación se presentan los resultados experimentales de la tarea de clasificación para los corpus creados. La razón por la cual se utilizaron diferentes corpus es para analizar las características de cada sección en cuestión.

6.3.2 Resultados experimentales

El objetivo de los experimentos es analizar la eficacia de las diferentes representaciones propuestas para la tarea de clasificación de párrafos y oraciones como argumentativos o no argumentativos (parte de la segunda tarea de la metodología, inciso a y b).

Las primeras representaciones vectoriales utilizadas para la identificación de argumentación fueron principalmente léxicas, con la frecuencia de n-gramas de palabras consecutivas de longitud de 1-3. Se construyeron las tres representaciones con unigramas, bigramas y trigramas de palabras los cuales incluyen las palabras vacías y puntuación.

Posteriormente al analizar los trabajos relacionados con la clasificación de textos argumentativos observamos a investigadores como Florou et al. (2013) y Moens et al. (2007) los cuales identifican textos argumentativos utilizando, por ejemplo: conjuntos de categorías de argumentos, características basadas en la frecuencia del modo y tiempo de los verbos, estadísticas del texto (longitud de oración, promedio de longitud de oración, puntuación), combinación de pares de palabras y uso de verbos principales. Es importante indicar que las características descritas usadas por los investigadores antes mencionados, se aplicaron en lenguajes diferentes al español y tipos de textos diferentes a tesis de estudiantes, por tal motivo las herramientas y estructura de los textos son diferentes. Se crearon representaciones aprovechando dichas características para analizar su efectividad en cada una de nuestras colecciones.

Se utilizaron las siguientes características para crear las representaciones vectoriales:

- Unigramas
 - o Son todas las palabras en el segmento de texto. Se incluyen las marcas de puntuación como (. ; : ,).

- Bigramas
 - Son pares de términos sucesivos ya sean palabras o marcas de puntuación.
- Trigramas
 - Son secuencias de tres términos sucesivos sean palabras o marcas de puntuación.
- Verbos principales
 - Los verbos son detectados con un etiquetado previo de partes de la oración. Ejemplos de verbos principales son: deber, emplear, ver, poner, etc. Para el etiquetado de partes de la oración se utilizó la herramienta de análisis de lenguaje FreeLing (Padró & Stanilovsky, 2012).
- Pares de palabras:
 - Se generan todas las posibles combinaciones entre pares de palabras en una oración. La intención es capturar más información que los bigramas.
- Estadísticos del Texto (3 características):
 - Se considera la longitud de la oración, longitud promedio de las palabras de la oración y el número de marcas de puntuación (. , : ;).
- Marcadores argumentativos (2 características):
 - Estos son palabras claves, las cuales nos ayudan a identificar los elementos de un argumento. Se crearon dos conjuntos de palabras claves para premisas y conclusiones (Capaldi, 2000). Se calculó la frecuencia de ocurrencia de cada palabra clave, posteriormente se sumaron las frecuencias en cada conjunto para obtener un valor global para el conjunto premisas y otro para conclusiones. Algunos marcadores para identificar las premisas son: *ya que, puesto que, puesto, pues, como, en tanto que, dado que, viendo que, debido a, de acuerdo con, por cuanto*, etc. Y para las conclusiones: *por lo tanto, por ende, de ahí que, en consecuencia, como resultado, llegamos a la conclusión, lo cual apunta a la conclusión de que, lo cual muestra que*, etc.
- Categorías de argumentos (5 características):
 - Se utilizaron cinco conjuntos de patrones de palabras para obtener la frecuencia de ocurrencia de cada uno de los siguientes tipos de argumentos: justificación, explicación, deducción, refutación y condicional. Dichos con-

juntos se crearon a partir de la revisión de diferentes fuentes (Avendaño, 2005; Briz et al. 2008; Rodríguez, 2009).

- Frecuencia relativa de tiempo y modo de verbos (12 características)
 - Se obtuvieron las frecuencias para cada tiempo (Presente - P, Pasado - S, Imperfecto - I, Condicional - C, Futuro - F, No tiene - 0) y modo (Participio - P, Gerundio - G, Indicativo - I, Subjuntivo - S, Imperativo - M, Infinitivo - N) presentes en la colección de documentos Coltypi.
- Frecuencia relativa de la combinación de modo/tiempo de los verbos (11 características)
 - Se obtuvo la frecuencia relativa de las 11 combinaciones identificadas en la colección de documentos Coltypi para cada segmento de texto (Participio sin tiempo - P0, Gerundio sin tiempo - G0, Indicativo presente IP, Indicativo Pasado - IS, Subjuntivo Presente - SP, Indicativo Imperfecto - II, Subjuntivo Imperfecto SI, Imperativo sin tiempo - M0, Infinitivo sin tiempo - N0, Indicativo Condicional - IC, Indicativo Futuro - IF).
- Indicador binario de aparición de tiempo y modo de verbos (12 características)
 - Para cada uno de las 6 posibilidades de tiempos de verbos identificadas en la colección de documentos Coltypi y 6 modos posibles, se indicó la aparición o ausencia de dicha característica.
- Características de verbos más frecuentes (3 características):
 - Se identificó el tiempo más frecuente (P, S, I, C, F, 0), el modo más frecuente (P, G, I, S, M, N) y la combinación modo/tiempo más frecuente (P0, G0, IP, IS, SP, II, SI, M0, N0, IC, IF).

6.3.3 Muestra con un anotador

Como primer paso, se analizó la muestra anotada por parte de un experto con conocimiento en la construcción y elaboración de argumentos. La proporción de párrafos argumentativos es del 56.7% y para los párrafos no argumentativos del restante 43.3%. Las representaciones vectoriales utilizadas para la identificación de argumentación en párrafos fueron principalmente léxicas con la frecuencia de n-gramas de palabras consecutivas de longitud de 1-3. Se construyeron las tres representaciones con unigramas, bigramas y trigramas de palabras, las cuales incluyen las palabras vacías y puntuación. En la tabla 13 se

muestran los resultados de estas representaciones con los distintos algoritmos de clasificación.

Como se puede observar en la tabla 13 el clasificador SVM es el clasificador que presenta una mejor exactitud, precisión, recuerdo y medida F1 en comparación con NB y DT, identificando párrafos con argumentos con un 71.9% de exactitud y un 81.28% de precisión.

Tabla 13. Resultados en clasificación de párrafos argumentativos

Unigramas				
	Precisión	Recuerdo	Medida F1	Exactitud
Support Vector Machine	79.55%	69.42%	72.64%	70.99%
Naïve Bayes	73.27%	65.58%	68.35%	66.69%
Decision Tree	62.76%	56.15%	53.10%	56.48%
+Bigramas (n-grams 1-2)				
	Precisión	Recuerdo	Medida F1	Exactitud
Support Vector Machine	80.52%	67.95%	72.20%	71.59%
Naïve Bayes	71.99%	67.95%	68.75%	66.25%
Decision Tree	60.92%	55.38%	57.46%	55.01%
+Trigramas (n-grams 1-3)				
	Precisión	Recuerdo	Medida F1	Exactitud
Support Vector Machine	81.26%	66.99%	72.44%	71.90%
Naïve Bayes	74.67%	63.91%	67.80%	65.78%
Decision Tree	62.18%	56.86%	58.88%	56.31%

Además se analizó la importancia de las palabras vacías en la clasificación de párrafos argumentativos. En la tabla 14 se observa en la cuarta fila (Unigramas stopwords) al utilizar únicamente la frecuencia de palabras vacías se consiguen resultados por encima del 62% de exactitud para los tres clasificadores analizados. Al incorporar todas las palabras y signos de puntuación como se observa en la primer fila (Unigramas (stopwords+ p+)) los clasificadores SVM y NB obtienen una mejor exactitud, sin embargo el clasificador DT (árbol de decisión) pierde exactitud debido al incremento en la cantidad de características ya que al parecer los árboles de decisión trabajan mejor con problemas de clasificación de menor dimensionalidad. Con esto se puede constatar la importancia de las palabras vacías en la identificación de argumentos.

Tabla 14. Exactitud y desviación estándar en comparativa de representaciones ignorando palabras vacías y puntuación

	SVM	NB	DT (J48)
Unigramas (p+ stopword+)	70.99 (9.31)	66.68 (10.65)	56.48 (15.18)
Unigramas (stopwords+ p-)	70.04 (9.6)	67.20 (11.4)	56.4 (14.8)
Unigramas (stopwords- p-)	68.73 (8.87)	65.25 (9.97)	61.73 (11.77)
Unigramas stopwords	66.08 (8.53)	64.30 (10.3)	62.2 (10.8)

Como se observó la información de las palabras vacías, puntuación y secuencias de palabras de longitud 1 a 3, aportan información valiosa para la clasificación de argumentos. Al realizar un filtrado adicional al seleccionar los bigramas y trigramas con frecuencia mayor a 5, se observa una mejoría en la exactitud de los clasificadores como se muestra en la tabla 15. El clasificador SVM obtiene un 75.54% de exactitud con una desviación estándar de 8.87. En tareas similares investigadores como Moens et al. (2007) obtuvieron un 73.75% de exactitud al clasificar oraciones como argumentativas o no.

Tabla 15. Exactitud y desviación estándar en clasificadores con unigramas, bigramas y trigramas con signos de puntuación y palabras vacías.

Unigrams + Bigrams (Freq. ≥ 5) + Trigrams (Freq. ≥ 5)				
	Precisión	Recuerdo	Medida F1	Exactitud
Support Vector Machine	85.65%	70.19%	75.76%	75.55%
Naïve Bayes	74.69%	66.28%	69.26%	67.08%
Decision Tree	61.08%	55.32%	57.47%	55.45%

7.3.4 Muestra con tres anotadores

A continuación se analizó la muestra anotada por tres expertos con conocimiento en la construcción y elaboración de argumentos. La proporción de párrafos argumentativos es de 62.4% y para los párrafos no argumentativos del restante 37.6% para 226 párrafos.

Para la siguiente tabla se consideró la siguiente notación: verbos principales (VP), pares de palabras (Comb_P), estadísticos del texto (Est_tex), marcadores argumentativos (MArg), Categorías de argumentos (CArg), Frecuencia relativa de tiempo y modo

(FRel_tm), Frecuencia relativa de la combinación de modo/tiempo (Comb_mt), Indicador binario de aparición de tiempo y modo (Bin_tm), Características de verbos más frecuentes (VM_Frec), P -Precisión, R - Recuerdo, F1 - Medida F1, E - Exactitud, SVM - Support Vector Machine, NB - Naive Bayes, DT - Decision Tree.

Como se observa en la Tabla 16 la representación con unigramas alcanza una exactitud de 75.2%, es decir la tradicional bolsa de palabras. Sin embargo, se observa como los pares de palabras alcanzan un nivel de recuerdo mayor con el clasificado NB donde se obtiene un 84.4%.

Tabla 16. Resultados de clasificación de párrafos argumentativo en Muestra

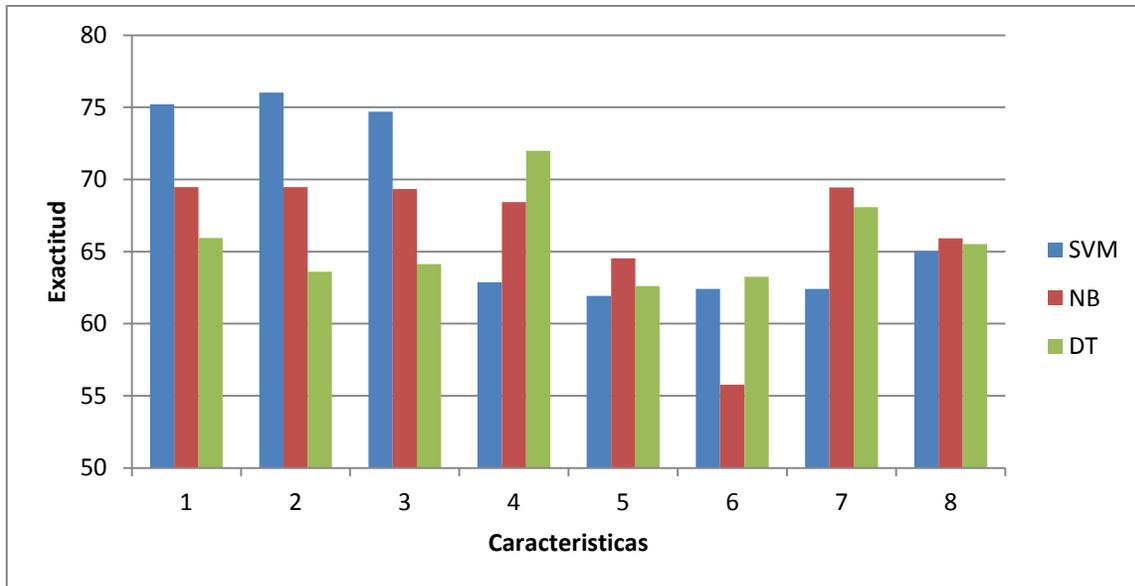
Características	SVM				NB				DT			
	P	R	F1	E	P	R	F1	E	P	R	F1	E
Unigramas	80.83	79.4	79.7	75.2	78.85	69.6	73.5	69.47	73.29	72.3	72.6	65.94
Unigramas + Bigramas freq \geq 5	82.96	78.7	80	76.02	77.6	71.7	74.2	69.47	72.46	68.8	69.9	63.61
Comb_p	90.43	67.3	76.3	74.7	71.9	84.4	77.4	69.34	73.74	66.7	69.8	64.14
VP	71.71	63.1	66.4	61.92	72.97	68.1	70.2	64.53	72.79	66.7	68.5	62.62
Est_tex	62.42	100	76.9	62.42	64.87	66	65.1	55.77	64.61	91.4	75.5	63.25
Marg	62.42	100	76.9	62.42	80.51	69.5	73.7	69.44	80.19	67.4	72	68.09
Carg	63.03	98.6	76.8	62.87	85.24	61.8	71	68.43	79.09	76.6	77.3	71.99
FRel_tm + Comb_mt + Bin_tm + VM_Frec	66.59	88.6	75.9	65.03	69.23	83	75.1	65.91	71.18	75.1	73	65.52

En la tabla 17 se presentan las mejores representaciones reportadas por dos investigadores del área y dos propuestas para ser comparadas:

Tabla 17. Comparativa de clasificación de párrafos en Muestra.

Características	SVM				NB				DT			
	P	R	F1	E	P	R	F1	E	P	R	F1	E
Florou et al. 2013	67.79	83	74.5	64.61	72.94	79.4	75.5	68.15	75.87	75.9	75.6	69.46
Moens et al. 2007	91.2	64.4	74.6	73.37	72.86	84.4	77.9	70.23	69.68	69.5	69.2	61.9
Comb_p + Carg	90.43	67.3	76.3	74.7	71.9	84.4	77.4	69.34	74.24	66	69.6	64.21
Comb_p + Carg + unigramas + bigramas freq \geq 5	91.83	70.1	78.9	77.26	71.9	84.4	77.4	69.34	69.18	67.4	68.1	60.75

En la tabla 17 se puede observar una comparativa con la mejor representación reportada en Florou et al. (2013), la cual consiste en los siguientes conjuntos de atributos: FRel_tm, Comb_mt, Bin_tm, VM_Frec, Carg. Así como la representación de Moens et al. (2007), la cual contiene: VP, Est_tex, Comb_p. La representación propuesta (Comb_p + Carg + unigramas + bigramas freq \geq 5) alcanza una mejor exactitud con 77.26%, así como una mejor medida F1 de 78.9% utilizando el clasificador SVM.



#	Descripción
1	Unigramas
2	Unigramas+Bigramas \geq 5
3	Combinación de palabras
4	Categoría de argumentos
5	Verbos principales
6	Estadísticos de texto
7	Marcadores
8	Frecuencia de modo - tiempo y combinación

Fig. 4. Comparativa de exactitud en clasificadores y características

En la figura 4 se observa una comparativa, las características que reportan una exactitud superior al 70% son las cuatro primeras (1-4), las cuales corresponden a: unigramas, unigramas + bigramas \geq 5, combinación de palabras y categoría de argumentos. Por tal motivo se seleccionaron dichas características para conformar nuestra representación propuesta.

Al analizar la muestra por oraciones se tienen 403 oraciones de las cuales el 69% tienen argumento. En la tabla 18 se presentan resultados de la tarea de clasificación realizada.

Como se observa nuevamente en la tabla 18, utilizando unigramas obtienen valores de exactitud de 79.39% con el clasificador Naive Bayes.

Tabla 18. Resultados de clasificación de oraciones en Muestra.

Características	SVM				NB				DT			
	P	R	F1	E	P	R	F1	E	P	R	F1	E
Unigramas	80.35	90.7	85.1	78.18	85.31	84.9	85	79.39	81.82	81.7	81.5	74.25
Unigramas + Bigramas freq \geq 5	80.7	88.2	84.2	77.17	84.26	83.1	83.6	77.68	80.02	79.5	79.5	71.69
Comb_p	81.89	84.5	83	76.17	78.83	78.4	78.5	70.46	80.47	77.3	78.6	71.21
VP	73.06	82.4	77.4	66.77	79.48	67.9	73.1	65.67	74.04	93.2	82.5	72.7
Est_tex	68.99	100	81.6	68.99	72.06	90.6	80.2	69.21	68.99	100	81.6	68.99
Marg	68.99	100	81.6	68.99	70.12	94.6	79.3	67.49	68.99	100	81.6	68.99
Carg	68.99	100	81.6	68.99	73.21	73.8	73.3	63.28	68.77	97.5	80.6	67.76
FRel_tm + Comb_mt + Bin_tm + VM_Frec	72.49	97.5	83.1	72.71	74.9	81.3	77.9	68.25	74.73	91.1	82	72.5

Tabla 19. Comparativa de clasificación de oraciones en Muestra.

Características	SVM				NB				DT			
	P	R	F1	E	P	R	F1	E	P	R	F1	E
Florou et al. 2013	72.12	97.5	82.9	72.22	75.08	82.4	78.5	69.01	75.02	85.6	79.9	70.41
Moens et al. 2007	83.49	84.2	83.7	77.4	78.38	77.7	77.9	69.71	79.68	80.9	80	72.43
Comb_p + Carg	82.14	86.7	84.2	77.67	78.94	78.8	78.8	70.7	80.62	79.1	79.5	72.19
Comb_p + Carg + unigramas + bigramas freq \geq 5	79.44	89.6	84.1	76.68	79.09	79.5	79.2	71.19	78.43	76.2	77	68.7

Como se observa en la tabla 19, la representación con (Comb_p + Carg) alcanza un 77.67% de exactitud y una medida F de 84.2%. En este caso al tratarse de oraciones, se observa cómo los unigramas superan a las representaciones analizadas.

Cabe hacer notar que el corpus muestra de párrafos y oraciones tiene un desbalance entre las clases. Para solucionar dicha situación, se tomaron párrafo u oraciones sin argumentación del corpus anotado de la sección “Planteamiento del problema”. Con esto, al corpus muestra de párrafos se le agregan 56 párrafos sin argumentación para obtener un total de 282 párrafos con 141 párrafo con argumento lo cual corresponde al 50%.

Tabla 20. Resultados de clasificación de párrafos en Muestra Balanceada

Características	SVM				NB				DT			
	P	R	F1	E	P	R	F1	E	P	R	F1	E
Unigramas	76.19	61	67.4	70.55	70.47	60.9	64.7	67.74	62.3	65	63.2	63.38
Unigramas + Bigramas freq \geq 5	91.83	70.1	78.9	77.26	71.9	84.4	77.4	69.34	69.18	67.4	68.1	60.75
Comb_p	79.65	52.5	62.4	69.48	59.14	74.5	65.8	61.36	59.58	63.8	61.4	60.57
VP	57.35	48.2	51.7	55.74	62.95	53.1	57	59.98	60.82	49	53.8	58.55
Est_tex	70.97	41.1	51.6	61.64	57.78	43.8	48.9	55.93	59.27	48.7	43.8	51.43
Marg	72.24	36.8	48.3	61.31	71.07	29.2	40.1	57.86	64.95	65.8	65.2	64.83
Carg	67.23	45.3	53.6	61.64	65.35	37.5	47.2	58.81	63.42	63.1	62.8	62.4
FRel_tm + Comb_mt + Bin_tm + VM_Frec	52.14	65.1	57.5	52.07	54.83	77.2	64	56.64	59.37	59.5	58.2	58.48

En la tabla 20 se observan como la representación de Unigramas + Bigramas alcanza una exactitud de 77.26% al utilizar SVM y se obtiene un recuerdo del 84.4% al utilizar NB.

Tabla 21. Comparativa de clasificación de párrafos en Muestra Balanceada

Características	SVM				NB				DT			
	P	R	F1	E	P	R	F1	E	P	R	F1	E
Florou et al. 2013	66.88	51.8	57.8	62.05	57.32	71.5	63.5	58.74	64.41	63.1	63.4	63.45
Moens et al. 2007	77.83	51.8	61.4	68.45	59.25	73.8	65.5	61.33	57.54	59.5	58.2	58.12
Comb_p + Carg	79.65	52.5	62.4	69.48	59.14	74.5	65.8	61.36	59.03	61.6	60.1	59.86
Comb_p + Carg + unigramas + bigramas freq \geq 5	85.33	57.5	67.4	73.43	59.14	74.5	65.8	61.36	64	59	61	62.48

Como se observa en la tabla 21 la representación con mayor exactitud es (Comb_p + Carg + unigramas + bigramas freq \geq 5) con 73.43% y una medida F de 67.4 %. Nuevamente se observa que el clasificador NB logra un mejor recuerdo con 74.5%.

Por otro lado al balancear el corpus muestra segmentado en oraciones, es necesario agregar 153 oraciones sin argumentos del corpus de “Planteamiento del problema” para lograr un total de 556 oraciones, de las cuales 278 tienen argumentos y representan el 50%.

Tabla 22. Resultados de clasificación de oraciones en Muestra Balanceada

Características	SVM				NB				DT			
	P	R	F1	E	P	R	F1	E	P	R	F1	E
Unigramas	76.54	74.5	75.3	75.57	74.19	74.1	73.9	73.93	69.86	66.9	68	68.71
Unigramas + Bigramas freq \geq 5	77.43	71.3	74	75.06	72.13	75.2	73.4	72.83	68.71	65.1	66.5	67.25
Comb_p	72.37	56.5	63.2	67.67	64.78	72.6	68.3	66.38	63.65	65.5	64.4	63.87
VP	64.78	55.7	59.7	62.76	73.85	55	62.6	67.82	80.86	49.2	60.4	68.32
Est_tex	57.79	65	60.6	58.78	56.32	71.5	62.8	57.92	68.04	38.4	41.8	56.3
Marg	59.6	42	45.4	55.39	64.97	50.3	56.4	61.35	64.97	50.3	56.4	61.35
Carg	67.76	43.8	53.1	61.49	65.53	40.7	49.8	59.7	60.73	45.7	51.5	57.74
FRel_tm + Comb_mt + Bin_tm + VM_Frec	57.16	74.5	64.5	59.35	56.84	82	67.1	59.89	58.68	70.1	63.7	60.23

En la tabla 22 se observa la representación de unigramas con una exactitud de 75.57% con SVM y la representación (FRel_tm + Comb_mt + Bin_tm + VM_Frec) con el mejor recuerdo de 82% con NB.

En la tabla 23 se observa la comparativa entre las representaciones de diferente autores en la cual se observa cómo la representación propuesta (Comb_p + Carg + unigramas + bigramas freq \geq 5) obtiene 72.33% de exactitud.

Tabla 23. Comparativa de clasificación de oraciones en Muestra Balanceada

Características	SVM				NB				DT			
	P	R	F1	E	P	R	F1	E	P	R	F1	E
Florou et al. 2013	58.1	66.2	61.7	59.17	57.61	82	67.6	60.79	55.6	58.6	56.8	55.73
Moens et al. 2007	72.1	59.4	64.9	68.01	64.89	73	68.5	66.56	67.39	70.1	68.6	67.96
Comb_p + Carg	71.59	57.3	63.5	67.49	64.78	72.6	68.3	66.38	64.38	64.4	64.2	64.2
Comb_p + Carg + unigramas + bigramas freq \geq 5	76.48	65.2	70	72.33	64.78	72.6	68.3	66.38	67.07	69	67.9	67.61

6.3.5 Corpus de sección “Planteamiento del Problema”

El corpus de la sección de “Planteamiento de Problema”, con un total de 100 secciones, 397 párrafos y 905 oraciones, se anotó por un experto en formulación y evaluación de argumentos. Al realizar la tarea de clasificación se observaron los siguientes resultados:

Tabla 24. Resultados de clasificación de párrafo en corpus Planteamiento del problema

Características	SVM				NB				DT			
	P	R	F1	E	P	R	F1	E	P	R	F1	E
Unigramas	67.48	60.4	63.1	64.46	61.26	60.8	60.7	59.71	65.31	61.8	63.3	63.5
Unigramas + Bigramas freq \geq 5	72.43	58.3	63.6	66.69	60.53	64.8	62.2	59.71	66.95	65.2	65.4	65.23
Comb_p	69.85	58.9	63.4	65.76	61.75	58.4	59.4	60.24	64.93	62.9	63.1	62.89
VP	62.84	53.5	57.2	59.15	62.56	49.5	54.6	58.45	55.2	49.6	51.7	52.2
Est_tex	51.07	97.5	67	50.86	60.85	19.6	29	53.12	57.68	65.4	56.6	52.97
Marg	69.99	47	55.7	62.4	70.41	48.9	57.3	63.14	63.56	63.6	62.9	61.92
Carg	70.01	55.4	61	64.44	72.04	43.6	53.5	62.2	64.23	58.4	60.7	61.73
FRel_tm + Comb_mt + Bin_tm + VM_Frec	54.3	80.4	64.5	54.86	54.17	69.1	60.3	54.12	54.57	59.9	56.8	53.97

Al analizar los párrafos del corpus de “Planteamiento del problema”, los unigramas + bigramas alcanzan una exactitud del 66.69% como se observa en la tabla 24.

Tabla 25. Comparativa de clasificación de párrafo en corpus Planteamiento del problema

Características	SVM				NB				DT			
	P	R	F1	E	P	R	F1	E	P	R	F1	E
Florou et al. 2013	70.63	54.9	61.2	64.68	61.77	61.3	60.3	60.2	52.51	46.7	48.9	50.62
Moens et al. 2007	71.53	58.4	63.8	66.53	62.17	59.9	60.4	60.72	66.53	64.8	64.8	64.39
Comb_p + Carg	71.31	60.8	65.1	67.02	61.75	58.4	59.4	60.24	64.43	63.9	63.5	62.92
Comb_p + Carg + unigramas + bigramas freq \geq 5	71.25	62.8	66.3	67.74	62	58.9	59.8	60.5	72.16	68.7	69.7	70

Como se observa en la tabla 25 la mejor representación es (Comb_p + Carg + unigramas + bigramas freq \geq 5) con 70% de exactitud utilizando el clasificador de Árbol de Decisión (DT).

Al analizar las oraciones del corpus de “Planteamiento del problema”, los unigramas + bigramas nuevamente obtienen la mejor exactitud al alcanzar un 60.89% como se observa en la tabla 26.

Tabla 26. Resultados de clasificación de oraciones en corpus Planteamiento del problema

Características	SVM				NB				DT			
	P	R	F1	E	P	R	F1	E	P	R	F1	E
Unigramas	64.24	70	66.9	59.89	64.2	51.6	57	55.16	62.18	64.3	63	56.46
Unigramas + Bigramas freq \geq 5	65.06	70.8	67.7	60.89	66.76	58.1	61.8	58.59	62.3	61.5	61.6	55.69
Comb_p	63.59	74	68.3	60.11	61.84	50.7	55.5	52.81	59.99	59.8	59.8	53.38
VP	60.57	72.3	65.8	56.46	57.32	51.6	54.2	49.39	58.18	81	67.6	54.91
Est_tex	58.23	100	73.6	58.23	56.65	46.9	48.9	47.98	57.93	98.8	73	57.57
Marg	58.23	100	73.6	58.23	65.18	44	52.4	53.82	58.23	100	73.6	58.23
Carg	58.23	100	73.6	58.23	65.56	40.3	49.6	52.94	60.74	74	66.2	56.59
FRel_tm + Comb_mt + Bin_tm + VM_Frec	57.83	97.9	72.7	57.23	57.73	77.5	65.8	54.26	57.92	78.9	66.8	54.37

En la tabla 27 podemos observar que la mejor representación es (Comb_p + Carg + unigramas + bigramas freq \geq 5) con 62.99% de exactitud utilizando el clasificador SVM.

Tabla 27. Comparativa de clasificación de oraciones en corpus Planteamiento del problema

Características	SVM				NB				DT			
	P	R	F1	E	P	R	F1	E	P	R	F1	E
Florou et al. 2013	57.87	97.5	72.6	57.24	59.39	65.7	61.8	54.03	61.62	65.5	63.4	56.25
Moens et al. 2007	65.47	72.5	68.7	61.63	62.16	50.7	55.7	53.03	58.85	58.1	58.4	51.95
Comb_p + Carg	63.96	73.6	68.3	60.33	61.84	50.7	55.5	52.81	58.99	58.6	58.7	52.14
Comb_p + Carg + unigramas + bigramas freq \geq 5	66.49	73.6	69.8	62.99	61.84	50.7	55.5	52.81	63.54	63.2	63.3	57.46

6.3.6 Corpus de sección “Justificación”

El corpus de la sección de “Justificación”, con un total de 100 secciones, 307 párrafos y 602 oraciones, se anotó por un experto en formulación y evaluación de argumentos. Al realizar la tarea de clasificación para la segmentación en párrafos se observaron los siguientes resultados:

En la tabla 28 se observa como las Comb_p (combinación de pares de palabras) alcanza un 71.01% de exactitud utilizando el clasificador de Árbol de Decisión (DT).

Tabla 28. Resultados de clasificación de párrafos en corpus justificación

Características	SVM				NB				DT			
	P	R	F1	E	P	R	F1	E	P	R	F1	E
Unigramas	68.01	60.6	63.7	61.66	63.98	53.4	57.3	57.02	73.26	75	74	70.66
Unigramas + Bigramas freq \geq 5	76.2	64	69.3	68.48	67.17	63.9	65	62.23	73.36	75.5	73.9	70.34
Comb_p	71.69	56.5	62.8	62.66	73.18	62.2	66.1	64.76	76.03	73.3	74.1	71.01
VP	63.67	51	55.4	56.06	60.24	54	56.3	53.73	57.77	62.8	59.8	53.82
Est_tex	56.03	100	71.8	56.03	65.93	40.1	49	54.64	68.1	54.8	58.2	57.6
Marg	67.86	66.2	65.6	63.18	68.1	51	57.5	58.9	65.02	72.6	68.5	62.53
Carg	71.23	66.8	68.8	66.1	74.86	46	56.3	61	66.25	70.4	67.8	63.17
FRel_tm + Comb_mt + Bin_tm + VM_Frec	70.56	80.2	74.9	70.04	66.79	69.2	67.8	63.5	63.3	61.7	61.8	59.76

En la tabla 29 se observa como la representación (Comb_p + Carg + unigramas + bigramas freq \geq 5) alcanza un 73.01% de exactitud utilizando el clasificador de Árbol de Decisión (DT) nuevamente.

Tabla 29. Comparativa de clasificación de párrafos en corpus justificación

Características	SVM				NB				DT			
	P	R	F1	E	P	R	F1	E	P	R	F1	E
Florou et al. 2013	72.98	80.8	76.5	72.26	69.92	65.1	67.2	64.85	65.47	68.6	66.3	60.99
Moens et al. 2007	71.3	56	62.2	62.35	70.77	61	64.9	63.14	71.3	72.1	71.3	67.39
Comb_p + Carg	70.99	55.9	62.2	62.02	73.18	62.2	66.1	64.76	75.87	72.7	73.7	70.69
Comb_p + Carg + unigramas + bigramas freq \geq 5	73.94	62.3	67.2	66.54	73.18	62.2	66.1	64.76	74.67	79.2	76.5	73.01

Por otro lado al analizar las oraciones en el corpus de “Justificación” se observa en la tabla 30 a los unigramas + bigramas freq \geq 5 con una exactitud del 67.96% con SVM.

Tabla 30. Resultados de clasificación de oraciones en corpus justificación

Características	SVM				NB				DT			
	P	R	F1	E	P	R	F1	E	P	R	F1	E
Unigramas	69.9	80.9	74.9	66.45	68.23	61.8	64.7	58.79	67.37	71.3	69.1	60.98
Unigramas + Bigramas freq \geq 5	71.49	80.6	75.7	67.96	70.76	62.9	66.3	61.12	68.23	69.6	68.7	60.96
Comb_p	67.04	78.7	72.3	62.79	65.22	57.8	61	54.99	65.65	69.9	67.5	58.98
VP	62.4	77.7	69.1	57.29	61.32	62.1	61.6	51.98	61.36	97	75.2	60.46
Est_tex	61.79	100	76.4	61.79	61.4	72.3	65.7	55.45	61.79	100	76.4	61.79
Marg	61.79	100	76.4	61.79	61.79	100	76.4	61.79	61.79	100	76.4	61.79
Carg	61.79	100	76.4	61.79	68.75	51.4	58.2	55.84	61.64	89.5	72.6	59.13
FRel_tm + Comb_mt + Bin_tm + VM_Frec	63.47	96.2	76.5	63.45	67.73	60.2	63.5	57.48	64.48	83.1	72.5	61.14

Al comparar con otras representaciones se observa que (Comb_p + Carg + unigramas + bigramas freq \geq 5) se alcanza un 65.78% de exactitud con SVM en la tabla 31.

Tabla 31. Comparativa de clasificación de oraciones en corpus justificación

características	SVM				NB				DT			
	P	R	F1	E	P	R	F1	E	P	R	F1	E
Florou et al. 2013	63.47	96.2	76.5	63.45	70.1	60.2	64.6	59.47	62.21	69.3	65.5	54.96
Moens et al. 2007	68.98	74.4	71.5	63.45	64.51	56.2	59.7	53.99	65.32	65.9	65.4	57.32
Comb_p + Carg	67.4	79.5	72.9	63.45	65.22	57.8	61	54.99	65.9	69.1	67.3	58.97
Comb_p + Carg + unigramas + bigramas freq \geq 5	68.58	82.8	74.9	65.78	65.22	57.8	61	54.99	67.56	67.2	67.1	59.8

6.3.7 Conclusión de experimentos

La tarea analizada en la presente sección corresponde a la identificación de argumentos en párrafos y oraciones. Esta tarea se considera necesaria para realizar un análisis descendente (top-down) del texto del estudiante, es decir como primer paso identificar los párrafos con argumentos y posteriormente realizar un análisis a detalle de su estructura argumentativa. Los experimentos se presentan en orden cronológico a medida que se realizaba la tarea de anotación. El primer conjunto de párrafos anotados corresponde a una “muestra” de 226 párrafos tomados de las secciones analizadas (planteamiento del problema, justificación y conclusiones), a continuación se anotó un conjunto de 397 párrafos de la sección de “planteamiento del problema”, y finalmente se anotaron 307 párrafos de la sección de “justificación”. La colección de documentos utilizada para dichas anotaciones fue Coltypi.

En la tabla 32 se observan las representaciones con mejor exactitud para la identificación de párrafos argumentativos para cada conjunto de datos anotado. Se observa cómo la representación propuesta obtiene en la mayoría de los conjuntos de datos la mejor exactitud. Dicha representación propuesta está compuesta por combinación de palabras (Comb_p), categorías de argumentos (Carg), unigramas y bigramas con una frecuencia mayor a 5, la cual consigue una exactitud por encima de 70% utilizando el clasificador SVM y DT. Se observa cómo las representaciones de (Florou et al. 2013) y (Moens et al. 2007) no se incluyeron en la tabla 32 debido a que se obtuvo una exactitud inferior a la representa-

ción propuesta en los conjuntos de datos anotados, lo cual nos indica que el tipo de textos (tesis y propuestas de investigación) utilizados influye en el tipo de características necesarias para resolver el problema de clasificación.

Tabla 32. Representaciones con mejor exactitud para identificación de párrafos con argumentos y sin argumentos

Conjunto de Datos	Representación	Clasificador	Exactitud
Muestra	Comb_p + Carg + unigramas + bigramas freq \geq 5	SVM	77.26
Muestra (balanceada)	Unigramas + Bigramas	SVM	77.26
Planteamiento del problema	Comb_p + Carg + unigramas + bigramas freq \geq 5	DT	70
Justificación	Comb_p + Carg + unigramas + bigramas freq \geq 5	DT	73.01

Tabla 33. Representaciones con mejor exactitud para identificación de oraciones argumentativas

Conjunto de Datos	Representación	Clasificador	Exactitud
Muestra	Unigramas	NB	79.39
Muestra (balanceada)	Unigramas	SVM	75.57
Planteamiento del problema	Comb_p + Carg + unigramas + bigramas freq \geq 5	SVM	62.99
Justificación	Unigramas + bigramas freq \geq 5	SVM	67.96

Finalmente en la tabla 33 se observan las representaciones para la tarea de identificación de oraciones argumentativas, donde en las filas 1 y 2 se observa cómo utilizando unigramas se obtienen niveles de exactitud superiores a 75% con clasificadores como NB y SVM. En las filas 3 se observa cómo la representación propuesta alcanza una exactitud superior al 62% utilizando el clasificador SVM, es decir consigue la mejor exactitud para la sección de planteamiento del problema. Los resultados presentados brindan sustento para utilizar la representación propuesta para realizar la tarea de identificación de párrafos y oraciones argumentativas en tesis y propuestas de investigación escritos en español, lo cual sería de apoyo en la redacción de dichos textos. Para mejorar la tarea de clasificación de oraciones argumentativas se analizaran otros aspectos como el contexto de la oración (p. ej. oraciones vecinas) y su posición en el párrafo.

7. Conclusiones

En este documento se presenta el trabajo realizado en el periodo Agosto 2015 – Junio 2016 y se describe parte de las actividades a llevar a cabo durante el programa doctoral. Como se indicó, existe la necesidad de apoyar a los estudiantes en el proceso de edición y revisión de argumentos, con el objetivo de mejorar la calidad de los escritos académicos en México, principalmente en el área de la escritura de textos de investigación como son las tesis y propuestas de investigación. Como se observó en los experimentos realizados, se tiene suficiente argumentación en los trabajos académicos, con el análisis del corpus realizado se puede estimar que más de la mitad de los párrafos escritos por alumnos incluyen argumentación. Por estas razones, la presente investigación doctoral se enfocará en hacer clasificación y extracción de los componentes argumentativos así como de sus relaciones. La finalidad de este esfuerzo es aprovechar dicha información argumentativa para mejorar los métodos de evaluación automática de textos académicos, para de esta forma apoyar a estudiantes y profesores. Como primer paso se observa en los resultados de los experimentos realizados, se logran alcanzar niveles de exactitud para identificar párrafos y oraciones argumentativas superiores a los obtenidos con representaciones propuestas por otros autores del área. Los niveles de exactitud observados utilizando los corpus anotados con nuestras representaciones propuestas son mayores con respecto a las representaciones de otros autores de un 0.75 a 4.98 puntos mayores para la identificación de párrafos argumentativos y de un 0.27 a 4.32 puntos mayores para la identificación de oraciones argumentativas. Los resultados indican que es factible apoyar en el análisis de la argumentación en el tiempo del programa doctoral. Se tienen programados experimentos para analizar características estructurales, contextuales y sintácticas para mejorar dicha tarea de clasificación, así como continuar con los pasos indicados en la metodología.

8. Publicaciones

Algunos de los resultados preliminares contenidos en la presente propuesta están por publicarse en:

- Garcia G., Jesus M. y López-López, Aurelio. 2016. Argumentation Identification for Academic Support in Undergraduate Writings. Proceedings of Eleventh European Conference On Technology Enhanced Learning. K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 98–109, 2016. DOI: 10.1007/978-3-319-45153-4_8 Springer International Publishing Switzerland 2016.

Referencias

- Argundi, Yolanda. (2005). Aprender a pensar escribiendo bien: desarrollo de habilidades para escribir. México, Trillas.
- Avendaño, C. S. (2005). Los conectores discursivos: su empleo en redacciones de estudiantes universitarios costarricenses. *Revista de Filología y Lingüística de la Universidad de Costa Rica*, 31(2).
- Bañales Faz, G., Vega López, N. A., Araujo Alvinada, N., Valladares, A., & Rodríguez Zamarripa, B. (2015). La Enseñanza de la Argumentación Escrita en la Universidad. *Revista Mexicana de Investigación Educativa*, 20(66).
- Briz, A., Pons, S., & Portolés, J. (2008). Diccionario de partículas discursivas del español. In *El diccionario como puente entre las lenguas y culturas del mundo. Actas del II Congreso Internacional de Lexicografía Hispánica*. Alicante, Biblioteca Virtual Cervantes (pp. 217-227).
- Burstein, J., Chodorow, M., & Leacock, C. (2003). CriterionSM Online Essay Evaluation: An Application for Automated Evaluation of Student Essays. In *IAAI* (pp. 3-10).
- Cabrio, E., & Villata, S. (2012). Combining textual entailment and argumentation theory for supporting online debates interactions. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2* (pp. 208-212). Association for Computational Linguistics.
- Capaldi, N. (2000). *Cómo ganar una discusión*. Gedisa.
- Carstens, L., & Toni, F. (2015). Towards relation based argumentation mining. In *Proceedings of the 2nd Workshop on Argumentation Mining*, (pp. 29-34). Association for Computational Linguistics, Denver, CO, USA.
- Chen, Y. Y., Liu, C. L., Lee, C. H., & Chang, T. H. (2010). An unsupervised automated essay scoring system. *IEEE Intelligent Systems*, 25(5), 61-67.
- Cho, K., & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers & Education*, 48(3), 409-426.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20, 37-46
- Eckle-Kohler, J., Kluge, R., & Gurevych, I. (2015). On the Role of Discourse Markers for Discriminating Claims and Premises in Argumentative Discourse. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 2249 -2255). Association for Computational Linguistics, Lisbon, Portugal.
- Feng, V. W., & Hirst, G. (2011). Classifying arguments by scheme. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1* (pp. 987-996). Association for Computational Linguistics.

- Florou, E., S. Konstantopoulos, A. Koukourikos, & P. Karampiperis. (2013). Argument extraction for supporting public policy formulation. In Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (pp. 49-54).
- Freeman, J. B. (2011). *Argument Structure: Representation and Theory* (Vol. 18). Springer Science & Business Media.
- González-López, s., & López-López A. (2015). Colección de Tesis y Propuesta de Investigación en TICs: un recurso para su análisis y estudio. XIII Congreso Nacional de Investigación Educativa (pp. 15).
- Goudas, T., Louizos, C., Petasis, G., & Karkaletsis, V. (2014). Argument extraction from news, blogs, and social media. In Hellenic Conference on Artificial Intelligence (pp. 287-299). Springer International Publishing.
- Green, N. L. (2015). Identifying Argumentation Schemes in Genetics Research Articles. In Proceedings of the 2nd Workshop on Argumentation Mining, (pp. 12-21). Association for Computational Linguistics, Denver, CO, USA.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA Data Mining Software: An Update. *ACM SIGKDD explorations newsletter*, 11(1), 10-18.
- Katzav, J., Reed, C., & Rowe, G. W. (2004). Argument research corpus. M.-P. Huget (ed.), *Communication in Multiagent Systems*, Lecture Notes in Computer Science, Springer Verlag, Berlin, Germany, 269-283.
- Kirschner, C., Eckle-Kohler, J., & Gurevych, I. (2015). Linking the thoughts: Analysis of argumentation structures in scientific publications. In Proceedings of the 2nd Workshop on Argumentation Mining (pp. 1-11). Association for Computational Linguistics, Denver, CO, USA.
- Kouylekov, M., & Negri, M. (2010). An open-source package for recognizing textual entailment. In Proceedings of the ACL 2010 System Demonstrations (pp. 42-47). Association for Computational Linguistics.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159-174.
- Lawrence, J., & Reed, C. (2014). Combining Argument Mining Techniques. In Proceedings of the 2nd Workshop on Argumentation Mining, (pp. 127-136). Association for Computational Linguistics, Denver, CO, USA.
- Lindsay, David. (2011). *Scientific writing = thinking in words*. CSIRO PUBLISHING, Collingwood, Australia.
- López Ferrero, C. (2003). La argumentación en los géneros académicos. In Actas del Congreso Internacional "La Argumentación" (pp. 1121-1129). Buenos Aires: Universidad de Buenos Aires.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

- Mochales, Palau & Moens, M. F. (2009). Argumentation mining: the detection, classification and structure of arguments in text. In Proceedings of the 12th international conference on artificial intelligence and law (pp. 98-107). ACM.
- Mochales, R., & Moens, M. F. (2008). Study on the structure of argumentation in case law. *Frontiers in Artificial Intelligence and Applications*, 189(1), 11–20.
- Mochales, R., & Moens, M. F. (2011). Argumentation mining. *Artificial Intelligence and Law*, 19(1), 1-22.
- Moens, M. F., Boiy, E., Palau, R. M., & Reed, C. (2007). Automatic detection of arguments in legal texts. In Proceedings of the 11th international conference on Artificial intelligence and law (pp. 225-230). ACM.
- Nguyen, H. V., & Litman, D. J. (2015). Extracting argument and domain words for identifying argument components in texts. In Proceedings of the 2nd Workshop on Argumentation Mining (pp. 22-28). Association for Computational Linguistics, Denver, CO, USA.
- Ong, N., Litman, D., & Brusilovsky, A. (2014). Ontology-based argument mining and automatic essay scoring. In Proceedings of the First Workshop on Argumentation Mining, (pp. 24–28). Association for Computational Linguistics, USA.
- Padró, L., & Stanilovsky, E. (2012). Freeling 3.0: Towards wider multilinguality. In LREC2012 (pp. 2473-2479).
- Park, J., Blake, C., & Cardie, C. (2015). Toward machine-assisted participation in eRule-making: an argumentation model of evaluability. In Proceedings of the 15th International Conference on Artificial Intelligence and Law (pp. 206-210). ACM.
- Peldszus, A., & Stede, M. (2013). From Argument Diagrams to Argumentation Mining in Texts: *International Journal of Cognitive Informatics and Natural Intelligence*, 7(1), 1–31.
- Peldszus, A., & Stede, M. (2015). Joint prediction in MST-style discourse parsing for argumentation mining. In Proc. of the Conference on Empirical Methods in Natural Language Processing (pp. 938-948).
- Persing, I., & Ng, V. (2015). Modeling argument strength in student essays. In Proceedings of the 7th International Joint Conference on Natural Language Processing (pp. 543–552). Association for Computational Linguistics, Beijing, China.
- Persing, I., & Ng, V. (2016). End-to-End Argumentation Mining in Student Essays. In Proceedings of NAACL-HLT (pp. 1384-1394).
- Persing, I., Davis, A., & Ng, V. (2010). Modeling organization in student essays. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (pp. 229-239). Association for Computational Linguistics.
- Petasis, G., & Karkaletsis, V. (2016). In Proceedings of the 3rd Workshop on Argument Mining, (pp. 94–102). Association for Computational Linguistics, Berlin, Germany.
- Rodríguez, C. F. (2009). *Diccionario de conectores y operadores del español*. Arco/Libros.

- Rooney, N., Wang, H., & Browne, F. (2012). Applying Kernel Methods to Argumentation Mining. In Proceedings of the 25th FLAIRS Conference, (pp. 272-275).
- Roscoe, R. D., Allen, L. K., Weston, J. L., Crossley, S. A., & McNamara, D. S. (2014). The Writing Pal intelligent tutoring system: Usability testing and development. *Computers and Composition*, 34, 39-59
- Sardianos, C., Katakis, I. M., Petasis, G., & Karkaletsis, V. (2015). Argument Extraction from News. In Proceedings of the 2nd Workshop on Argumentation Mining, (pp. 56–66). Association for Computational Linguistics, Denver, Colorado, USA.
- Song, Y., & Deane, K. P. (2014). Applying Argumentation Schemes for Essay Scoring. In Proceedings of the First Workshop on Argumentation Mining, (pp. 69–78). Association for Computational Linguistics, Baltimore, Maryland, USA.
- Stab, C., & Gurevych, I. (2014). Identifying Argumentative Discourse Structures in Persuasive Essays. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), (pp. 46–56).
- Stab, C., & Gurevych, I. (2016). Parsing Argumentation Structures in Persuasive Essays. arXiv preprint arXiv:1604.07370.
- Stab, C., Kirschner, C., Eckle-Kohler, J., & Gurevych, I. (2014). Argumentation mining in persuasive essays and scientific articles from the discourse structure perspective. *Frontiers and Connections between Argumentation Theory and Natural Language Processing*, (pp. 21-25). Bertinoro, Italy.
- Teufel, S., & Moens, M. (2002). Summarizing Scientific Articles: Experiments with Relevance and Rhetorical Status. *Computational Linguistics*, 28(4), 409–445.
- Teufel, Simone. (1999). *Argumentative zoning: Information extraction from scientific text.* (Doctoral dissertation, University of Edinburgh).
- Toulmin, S. E. (1958). *The uses of argument.* England: Cambridge University Press.
- Venegas V., René. (2003). Análisis Semántico Latente: una panorámica de su desarrollo. *Revista Signos*, 36(53), 121-138.
- Villalba, M. P. G., & Saint-Dizier, P. (2012). Some Facets of Argument Mining for Opinion Analysis. *COMMA*, 245, 23-34.
- Wachsmuth, H., Kiesel, J., & Stein, B. (2015). Sentiment Flow—A General Model of Web Review Argumentation. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (pp. 601-611).
- Walton D., Reed C., & Macagno F. (2008). *Argumentation Schemes.* Cambridge University Press.
- Walton, D. (2005). *Fundamentals of critical argumentation.* Cambridge University Press.
- Weston, Anthony. (2005). *Las claves de la argumentación.* Ed. Ariel, Barcelona, España.
- Wyner, A. Z., & Bench-Capon, T. J. (2007). Towards an extensible argumentation system. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty* (pp. 283-294). Springer Berlin Heidelberg.
- Wyner, A., Mochales-Palau, R., Moens, M. F., & Milward, D. (2010). Approaches to text mining arguments from legal cases (pp. 60-79). Springer Berlin Heidelberg.

Anexos

Anexo A

Definición de medida Kappa

Medida estadística que ajusta el efecto del azar en la proporción de la concordancia observada. Mide el acuerdo entre dos observadores en sus correspondientes clasificaciones de N elementos en C categorías mutuamente excluyentes.

$$k = \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)}$$

Donde Pr(a) es el acuerdo observado relativo entre los observadores, y Pr(e) es la probabilidad hipotética de acuerdo por azar, utilizando los datos observados para calcular las probabilidades de que cada observador clasifique aleatoriamente cada categoría. Si los evaluadores están completamente de acuerdo, entonces $\kappa = 1$. Si no hay acuerdo entre los calificadores distinto al que cabría esperar por azar (según lo definido por Pr (e)), $\kappa \leq 0$.

Cohen, Jacob (1960). "A coefficient of agreement for nominal scales". *Educational and Psychological Measurement* 20 (1): 37–46.

Anexo B

Definición de medida Fleiss Kappa

Fleiss Kappa es una medida estadística para evaluar la fiabilidad de acuerdo entre un número fijo de evaluadores al asignar clasificaciones categóricas a una serie de elementos. Esto contrasta con Cohen Kappa, que sólo funcionan cuando se evalúa el acuerdo entre dos anotadores. La medida calcula el grado de acuerdo en la clasificación considerando el efecto del azar.

$$K = \frac{\bar{P} - \bar{P}_e}{1 - P_e}$$

Donde el factor $(1 - P_e)$ indica el grado de acuerdo alcanzable por encima de lo que se podría predecir por azar. Y el factor $(\bar{P} - \bar{P}_e)$ indica el grado de acuerdo realmente alcanzado por encima del azar. Si los evaluadores están en completo acuerdo entonces Kappa = 1. Si no hay acuerdo entre los evaluadores (aparte de lo que se esperaría por casualidad) entonces Kappa = 0.

Fleiss, J. L. (1971) "Measuring nominal scale agreement among many raters." *Psychological Bulletin*, Vol. 76, No. 5 pp. 378–382

Anexo C

Definición de medida F1 (Han, Pei & Kamber, 2011, p. 405)

Para definir la medida F es necesario definir primero las medidas de “precision” y “recall”, las cuales son ampliamente utilizadas en clasificación.

“precision” puede ser pensado como una medida de la exactitud (es decir, qué porcentaje de tuplas etiquetado como positivos son en realidad tales).

“recall” es una medida de la completitud (qué porcentaje de tuplas positivas están etiquetados como tales).

La medida F1 es la media armónica de “precision” y “recall”. Se da la misma importancia a la “precision” y “recall”. Su fórmula es la siguiente:

$$F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. Elsevier.