



INAOE

La Web como Recurso Lingüístico para la Desambiguación Semántica

Por

AARÓN PANCARDO RODRÍGUEZ

*Tesis sometida como requisito parcial para obtener el grado de **Maestro en Ciencias** en la especialidad de **Ciencias Computacionales** en el Instituto Nacional de Astrofísica, Óptica y Electrónica.*

Supervisada por:

Dr. Manuel Montes y Gómez

Dr. Luis Villaseñor Pineda

Coordinación de Ciencias Computacionales

INAOE

Tonantzintla, Puebla, México

Febrero 2006



Resumen

La selección del sentido más apropiado de una palabra ambigua en un contexto determinado es uno de los problemas centrales del Procesamiento del Lenguaje Natural. Generalmente, para esta tarea es necesario contar con una fuente de información semántica, es decir, recursos lingüísticos como diccionarios, tesauros, etc. El presente trabajo explora el uso de la Web como recurso lingüístico en esta tarea. El reto de este trabajo es la transformación de la información léxica presente en la Web a información semántica útil para la tarea de desambiguación. Para ello, se propusieron y probaron diferentes métodos para determinar la cercanía semántica de dos términos. Los resultados alcanzados demostraron que la Web puede apoyar primordialmente para descartar los sentidos más improbables de una palabra polisémica. De esta manera, se propuso un método híbrido bajo el cual se integra el conocimiento extraído de la Web con métodos tradicionales basados en un enfoque supervisado. En los experimentos realizados con este método se logró demostrar que la información extraída de la Web puede apoyar la tarea de desambiguación. Gracias a la información aportada por la Web se logró un incremento de 10.4% de precisión en promedio.

Abstract

Selecting the most appropriate sense of an ambiguous word in a particular context is one of the central problems of Natural Language Processing. Usually, for this task it is necessary to have semantic information, that is, some linguistic resources such as dictionaries, thesaurus, etc. The present work explores the use of the Web as a linguistic resource for this task. Its challenge is to transform the lexical information from the Web to semantic information useful for the disambiguation task. In order to do that, we proposed and evaluated different methods that allow determining the semantic proximity of two given terms. The achieved results showed that the Web can be mainly used to discard the most improbable senses of a polysemic word. In this way, we proposed a hybrid method that integrates the knowledge extracted from the Web into a supervised approach. The experiments using this method indicated that the information extracted from the Web may support the disambiguation task. Basically, it caused an increment of 10.4 percent on the average precision.

Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico otorgado a través de la beca para estudios de maestría con el número de referencia 181454.

Al Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE) por todas las facilidades administrativas que me otorgaron. A la Coordinación de Ciencias Computaciones por haberme aceptado para formar parte de la comunidad científica y compartir nuevas experiencias en este campo del saber.

De manera especial, al Dr. Manuel Montes y Gómez y al Dr. Luis Villaseñor Pineda quienes con su ejemplo y dedicación al trabajo, contribuyeron de manera decidida en mi formación científica en el área del Procesamiento del Lenguaje Natural.

Al comité formado por el Dr. Saúl E. Pomares H., Dr. Aurelio López L. y el Dr. J. Francisco Martínez T. quienes me ayudaron con su profesionalismo y paciencia en la presentación de éste trabajo.

Dedicatorias

A Dios quien me ha dado fortaleza y entendimiento.

A mis padres Ernesto y Myrna, mi hermano Isaías, mis tíos y primos, que me apoyaron en todo momento sin importar la distancia y, en especial, a mi abuela Alba Yolanda, por sus cuidados y sabios consejos.

A Martha, mi novia, por su cariño e inspiración.

Contenido

1 Introducción	9
1.1 Descripción del Problema	10
1.2 Organización de la Tesis	13
2 Desambiguación del Sentido de las Palabras: conceptos básicos.....	15
2.1 La importancia de WSD.....	15
2.2 Metodología básica de WSD.....	17
2.2.1 Métodos basados en conocimiento	18
2.2.2 Métodos basados en corpus.....	23
2.2.2.1 Métodos supervisados – corpus etiquetado.....	23
2.2.2.2 Métodos no supervisados – corpus no etiquetado.....	26
2.2.3 Métodos híbridos y Bootstrapping.....	27
2.3 Recursos Lingüísticos	28
2.3.1 WordNet.....	28
2.3.2 SENSEVAL	30
2.4 Algoritmos de Aprendizaje Automático	34
2.4.1 Clasificador Naïve Bayes.....	35
2.4.2 C4.5	36
2.4.3 k-Vecinos Más Cercanos	38
2.4.4 Regresión Localmente Ponderada (LWR)	39
2.4.5 Máquinas de Vectores de Soporte.....	40
2.4.6 Naïve Bayes Localmente Ponderado (LWNB).....	42
2.5 Consideraciones finales.....	44
3 Método basado en la Distancia Normalizada de Google	46
3.1 Motivación	46
3.2 Distancia Normalizada de Google	49
3.3 Metodología propuesta.....	52

3.4 Experimentos.....	54
3.5 Discusión de los resultados	60
4 Combinando NGD con un Enfoque Supervisado.....	61
4.1 Un ensamble para WSD.....	61
4.2 Integrando el clasificador NGD	63
4.2.1 Primer esquema: NGD en la fase de clasificación.....	64
4.2.2 Segundo esquema: NGD en la fase de votación	65
4.3 Discusiones	67
5 Conclusiones y Trabajo Futuro	68
5.1 Trabajo futuro.....	69
Apéndice 1:	
Comparación de clasificadores supervisados	71
Apéndice 2:	
Resultados de los ensambles usando el primer enfoque del capítulo 4.	73
Apéndice 3:	
Resultados de los ensambles usando el segundo enfoque del capítulo 4.	79
Lista de figuras.....	90
Lista de tablas.....	91
Referencias.....	94
Publicaciones	100

Capítulo 1

Introducción

En la actualidad, el recurso más valioso para el ser humano es *la información*. Con la tecnología informática, ha sido posible manipular una gran cantidad de información y almacenarla en forma electrónica generando documentos de diversos géneros e idiomas. Además, la mayor parte de esta información se encuentra disponible en la Web, donde se estima que el 80% se representa textualmente [1]. Debido a que esta masa de información crece exponencialmente, surge la necesidad de procesarla automáticamente para facilitar muchas tareas. Por tal motivo se ha recurrido a una de las ramas de la Inteligencia Artificial: el *Procesamiento del Lenguaje Natural (PLN)*. El enfoque principal de esta área consiste en crear métodos, técnicas y herramientas computacionales que permitan realizar análisis de información escrita u oral y que faciliten la búsqueda y organización de dicha información. Una de las tareas más difíciles en el procesamiento automático del lenguaje humano es la resolución de la ambigüedad de la palabra. Fundamentalmente esta resolución es necesaria dado que una palabra puede ser interpretada de diferentes formas, es decir, posee más de un significado o sentido. Este fenómeno lingüístico se le conoce como *polisemia*. Determinar el sentido correcto de una palabra en un texto o en una conversación es una tarea constante en la comunicación humana que, sin embargo, rara vez causa problemas. Por el contrario, para una computadora se convierte en una enorme dificultad, aún cuando las palabras sean obvias para un humano. Esto se debe a que las máquinas interpretan los textos como un conjunto de palabras sin ningún significado y/o valor. Para lograr la comprensión e interpretación adecuada de la información, es necesario analizar a fondo cada una de las palabras y así obtener la interpretación o sentido más apropiado. Por esta razón, surge la necesidad de buscar

métodos que ayuden a determinar el significado adecuado de las palabras que hay dentro de un texto con la ayuda del conocimiento más allá del texto mismo. Encontrar o crear estos métodos y alternativas para solucionar este tipo de problemas es precisamente la motivación del presente trabajo.

La elección del sentido más apropiado para una palabra polisémica se ha convertido en una de las tareas más importantes y difíciles dentro del área de PLN llamada Desambiguación del Sentido de las Palabras o, en inglés, Word Sense Disambiguation (en este documento se utilizarán las siglas WSD para referirse a esta tarea). Esta tarea es necesaria para lograr el desarrollo de la mayoría de las aplicaciones y otras tareas del PLN. Hasta ahora, los métodos propuestos han logrado resultados interesantes, desafortunadamente estos métodos están limitados al número de palabras a desambiguar dados los recursos lingüísticos utilizados. Así, la falta de información sobre el uso de la mayoría de las palabras es una de las causas principales de estas limitantes. De ahí, que la recopilación de recursos lingüísticos suficientes para WSD es de extrema importancia. No obstante, actualmente tenemos a nuestra disposición un enorme recurso lingüístico: la Web. Dentro de ella tenemos suficiente información para determinar el uso de una palabra y en consecuencia su sentido. Por este motivo, la Web resulta ser un recurso atractivo para crear métodos de WSD aprovechando la información que nos brinda. De lo anterior, surgen las preguntas de investigación de esta tesis y que serán contestadas a lo largo de este documento: ¿Cómo debe ser la metodología que utiliza la información de la Web para convertirlo en un recurso para desambiguar directamente? y ¿Qué tan eficiente puede ser esta metodología al combinarla con otros métodos?

1.1 Descripción del Problema

La tarea de WSD surgió como un problema en el área de la Traducción Automática en 1949 (ver [9] y [10]) al tratar de traducir automáticamente una palabra polisémica de un idioma a otro. Para lograrlo era esencial saber a qué sentido se refería una palabra en el idioma original para elegir la expresión apropiada en el idioma destino.

En 1960, Bar Hillel se encontró con la siguiente frase en inglés: “*Little John was looking for his toy box. Finally, he found it. The box was in the pen. John was very happy.*” [11]. La palabra en inglés “*pen*” se podía referir a una pluma para escribir ó a un corral para niños. ¿Cuál es el sentido más adecuado para la palabra “*pen*”? Este problema no sólo afecta en la Traducción Automática, sino también en las áreas donde la precisión de la búsqueda y el análisis de la información son esenciales.

Otro ejemplo para demostrar la dificultad de esta tarea es la palabra polisémica “*gato*”, donde dos de los sentidos más populares para esta palabra son: (1) animal mamífero felino y (2) máquina con engranes que sirve para levantar objetos pesados. La ambigüedad se presenta cuando no se tiene idea a qué sentido se está refiriendo la palabra. Para tratar de resolver este problema, se recurre a analizar el contexto, es decir, las palabras que rodean a la palabra que se quiere desambiguar. Por ejemplo, en la oración “*El gato necesita alimentos para sobrevivir*” se puede encontrar las palabras que hay en contexto que son claves para determinar el sentido de la palabra “*gato*”, en este caso, las palabras “*alimento*” y “*sobrevivir*” ayudan a determinar el sentido de la palabra el cual se refiere al animal mamífero felino. La desambiguación de una palabra no es un problema para un humano dado que gracias a la interpretación del enunciado él puede determinar la relación semántica precisa existente entre las palabras del contexto y la palabra que se quiere desambiguar. Desafortunadamente, la interpretación automática de un enunciado involucra la correcta desambiguación de la palabra. Así la tarea de desambiguación es una tarea previa a la interpretación considerando únicamente información léxica orientada a establecer estadísticamente su concurrencia con palabras presentes en su contexto inmediato.

En los años de 1970 y 1980, se presentaron los primeros sistemas de WSD. Estos sistemas fueron basados en reglas. Para ello se usaron fuentes de conocimiento (corpus y reglas) que se desarrollaban manualmente enfocándose específicamente en la desambiguación de algunas palabras en específico. Más tarde, en los años 1990 surgen métodos basados en técnicas de Aprendizaje Automático (AA) (refiérase a [3] para un estudio). El propósito del AA es utilizar métodos y técnicas que permitan a la

máquina adquirir conocimientos por medio de ejemplos, en este caso, sobre el uso de los sentidos de una palabra. La información de los ejemplos se encuentra representada de manera que los métodos y técnicas de AA puedan interpretarlas. Con la ayuda de estos algoritmos se crearon sistemas de WSD basados en Métodos Supervisados, es decir, sistemas que dependen de conocimiento previamente organizado para realizar la tarea de desambiguación. Este conocimiento se concentraba en forma de textos y/o documentos de varias fuentes (noticias, periódicos, publicaciones, etc.) donde las palabras a estudiar han sido marcadas puntualizando su sentido (véase [10] y [12]). Así estos corpus reúnen ejemplos de uso de esas palabras especificando explícitamente el sentido en cada caso. Las técnicas de AA toman esta información como conjuntos de entrenamiento resultando en sistemas supervisados de WSD [2]. Por supuesto, los métodos supervisados dependen del uso de estos conjuntos de datos haciendo que el conocimiento dado al programa de computadora sea limitado, es decir, la cobertura de nuestro sistema sólo puede enfocarse en aquellas palabras etiquetadas en el corpus. Lo que es más, la creación de estos conjuntos de datos es muy costosa tanto en términos de recursos humanos como en tiempo.

Ante la necesidad de evitar esta dependencia, surgieron los sistemas basados en métodos que utilizan otros recursos: los métodos no supervisados y los métodos mínimamente supervisados. Éstos no requieren de un conjunto de datos etiquetados para el entrenamiento. En su lugar, utilizan otros recursos, tales como diccionarios, modelos matemáticos e incluso conjuntos de datos no etiquetados. Aunque no son tan eficaces como los métodos supervisados, se cuenta con la ventaja de poder tener una cobertura más amplia en las palabras gracias a que ya no se depende de datos etiquetados.

Un enfoque reciente entre los sistemas no supervisados y mínimamente supervisados, es el uso de la Web como un recurso [5] para desambiguar, en contraste con los otros métodos supervisados, estos sistemas generalmente no utilizan técnicas de Aprendizaje Automático. Sin embargo, a pesar de tener a la Web como un recurso léxico, los resultados están lejos de ser exactos (véase [6] como un ejemplo de este caso). Muchos de estos sistemas utilizan la Web como una fuente para obtener más

ejemplos y crear un subconjunto de datos no etiquetados para después etiquetarlos automáticamente [13], pero aún así no es suficiente.

A diferencia de los métodos mencionados, el enfoque que se propone en esta tesis plantea un punto de vista distinto en cuanto al uso de la Web: utilizarla directamente como conocimiento desambiguador. La hipótesis que se pretende responder a lo largo de este documento es que al existir muchos documentos (mucha información) en la Web, se puede decir que existen evidencias sobre el uso adecuado de las palabras polisémicas. De esta manera, la información contenida en la Web se puede utilizar para apoyar directamente en la desambiguación de las palabras polisémicas y facilitar el proceso de la elección del sentido más adecuado. Por esta razón, es necesario la creación de una metodología que haga explícito el conocimiento desambiguador contenido implícitamente en la Web. La búsqueda de esta metodología es el objetivo principal de esta tesis y con esto buscamos resolver a las preguntas de investigación mencionadas: ¿Cómo debe ser la metodología que utiliza la información de la Web para convertirlo en un recurso para desambiguar directamente?, en otras palabras; ¿Cómo usar la Web para ayudar en el proceso de la elección del sentido de la palabra? Además, ¿Cómo combinar el conocimiento extraído de la Web con los métodos mencionados anteriormente? Estas preguntas serán respondidas a lo largo de esta tesis, los experimentos basados en el enfoque propuesto demuestran la importancia de aprovechar la Web como un recurso para apoyar a la tarea de WSD.

1.2 Organización de la Tesis

El contenido de este documento está organizado de la siguiente manera. En el capítulo 2 se presentan los conceptos básicos de WSD y se analizan los diferentes enfoques existentes de WSD, así como los conjuntos de datos utilizados. Además se presenta una breve explicación sobre los algoritmos de Aprendizaje Automáticos utilizados para la experimentación en esta tesis. En el tercer capítulo se describe la metodología propuesta, la cual está basada en el uso de la Web como conocimiento

desambiguador en el proceso de elección del sentido de una palabra. Además, se muestran y discuten los resultados de esta metodología. El capítulo 4 muestra un esquema experimental donde se combinan los métodos supervisados y el método basado en la Web. Por último, el capítulo 5 presenta las conclusiones y discusiones sobre esta investigación, además de posibles direcciones para un trabajo futuro.

Capítulo 2

Desambiguación del sentido de las palabras: conceptos básicos

En este capítulo se describen los conceptos básicos en la tarea de desambiguación del sentido de las palabras (WSD, por sus siglas en inglés). En las primeras secciones se ejemplifica el uso de WSD en dos tareas del procesamiento del lenguaje natural (PLN). Posteriormente se presentan los principales enfoques utilizados en WSD, además de describirse los recursos lingüísticos utilizados para esta tarea. Finalmente, en la última parte de este capítulo, se describen brevemente los algoritmos de aprendizaje automático más comúnmente utilizados en WSD.

2.1 La importancia de WSD

La necesidad de que una computadora pueda interpretar la información correctamente y de esta manera realizar su trabajo en forma eficaz hace que WSD tome un papel relevante dentro del PLN, pues ésta es una tarea fundamental de la cual dependen muchas otras tareas. La Traducción Automática y la Recuperación de Información son dos tareas dentro del PLN que ejemplifican la necesidad de la resolución de la ambigüedad de los sentidos de una palabra para cumplir sus objetivos.

La Traducción Automática es una de las tareas donde la WSD juega un rol muy importante. Esta tarea pretende realizar automáticamente una traducción de un texto, frase u oración que está en un idioma (idioma origen) al idioma a elegir (idioma destino); por supuesto, esto no sólo consiste en intercambiar las palabras del idioma origen por la del idioma destino. Básicamente el proceso de traducción automática requiere de dos fases: (1) el entendimiento del lenguaje original y (2) la

generación de la sentencia traducida del texto en el idioma objetivo. En ambas fases es necesario resolver el problema de la ambigüedad del sentido de la palabra. La primera fase, el problema puede ocurrir cuando hay una palabra polisémica en el idioma original generando posibles traducciones distintas entre sí. Un ejemplo es la palabra polisémica en francés “*grille*”, que puede ser traducida al inglés como “*railings*”(carriles), “*bar*” (barras), “*gate*” (puente), “*grid*” (rejilla), “*scale*” (escala) y “*schedule*” (agenda) dependiendo de su contexto. De la misma manera ocurre en la segunda fase, el problema surge cuando existe más de una traducción posible para una palabra que no es ambigua en el idioma original. Por ejemplo, la palabra en inglés “*valley*” se puede referir en el idioma gaélico como “*strath*” (un valle con un río muy ancho) o “*glen*” (un valle localizado entre colinas escarpadas). Debido a estos problemas, el trabajo en WSD incide directamente en los resultados de esta área (Bar-Hillel [11] y Wilks et. al. [15]). El beneficio consiste en la elección de las palabras más adecuadas del idioma destino, basándose en el contexto que las rodea en el idioma origen, permitiendo que la traducción sea más apropiada.

Otra de las aplicaciones más importantes dentro del PLN, es la Recuperación de Información. El propósito de esta área es encontrar los documentos de interés que se encuentra dentro de una colección de documentos. Sin embargo, uno de los problemas que se enfrenta en esta área es la precisión en la búsqueda. Es decir, los documentos resultantes de la búsqueda pueden contener documentos no deseados, así como, dejar fuera documentos pertinentes. Esto es debido a la forma de realizar la búsqueda. Básicamente, se utilizan las palabras de la consulta recopilando aquellos documentos que contengan dichas palabras. Como es de imaginarse, existe la posibilidad de que estas palabras sean polisémicas. Para ejemplificar el problema, supóngase que se desea buscar sobre “*java*” en Google y se han encontrado 4320000 documentos en español con este término; entre los documentos resultantes se tendrán más de un tema, debido que “*java*” es una palabra polisémica. Entre los posibles sentidos de “*java*” tenemos: el nombre de una isla localizada en el archipiélago indonesio; un tipo de café; o un lenguaje de programación. Sin embargo, si se añade algunas palabras que estén relacionadas con el sentido del término deseado, el

número de documentos será menor y con más posibilidades de que el tema del documento encontrado sea el sentido objetivo. En la tabla 2.1 se ejemplifica los resultados de la búsqueda en Google utilizando los términos mencionados, la segunda columna indica la cantidad de documentos en donde los términos están presentes. Por esta razón, la desambiguación del sentido de una palabra mejorará los resultados en la búsqueda de documentos, al analizar y verificar el sentido adecuado de las palabras utilizadas para la búsqueda.

Términos	# de documentos
Java	4320000
java programación	2640000
java café	215000
java isla	192000

Tabla 2.1: Ejemplo de resultados de la búsqueda en Google.

2.2 Metodología básica de WSD

WSD se puede definir como el proceso de elección del sentido más adecuado de una palabra polisémica ayudándose del contexto que la rodea [8]. El contexto son las palabras que se encuentran en el lado derecho e izquierdo de la palabra a desambiguar dentro del texto, frase u oración. En muchas ocasiones, el tamaño del contexto debe ser definido, es decir, cuántas palabras se incluirán para el proceso de desambiguación. Por ejemplo, para desambiguar la palabra “gato” en la oración “*una necesidad del gato es la comida*”, usando un tamaño de contexto de 3 palabras en ambos lados, el contexto del lado izquierdo está compuesto de “*una necesidad del*” y el lado derecho de “*es la comida*”. Las palabras del contexto ayudan a los métodos de WSD para encontrar relaciones o patrones que caractericen el sentido de la palabra a desambiguar, en este caso las palabras “*comida*” y “*necesidad*” determinarán que el sentido adecuado para la palabra “gato” es el de un animal. A partir de esta idea, hoy en día existen muchos sistemas WSD con distintos enfoques. Retomando el trabajo de Mark[8], y de Rada Mihalcea y Ted Pedersen [12] se presenta en las secciones subsecuentes una clasificación de estos enfoques, los cuales se pueden reunir en tres grandes grupos:

1. Métodos basados en conocimiento
2. Métodos basados en corpus
 - a. Métodos supervisados – corpus etiquetado
 - b. Métodos no supervisados – corpus no etiquetado
3. Métodos híbridos y Bootstrapping

Antes de continuar con la descripción de los enfoques, es importante aclarar los siguientes términos: precisión, recuerdo y cobertura. Estos términos son medidas para determinar la eficacia de los métodos. Para evaluar este tipo de trabajos comúnmente se utiliza un corpus de referencia, es decir, un conjunto de palabras polisémicas en contextos específicos donde el sentido de la palabra ha sido marcado previamente por un juez. La precisión es el porcentaje de palabras correctamente desambiguadas por el sistema de WSD dado un corpus de referencia (número de instancias correctamente desambiguadas por el sistema WSD / número de instancias a las cuales el sistema de WSD propuso una respuesta). El recuerdo es el porcentaje de palabras que han sido correctamente desambiguadas dentro del conjunto de todas las palabras de prueba (número de instancias correctamente desambiguadas / número total de instancias en el conjunto de prueba). La cobertura es el porcentaje de palabras a las que el sistema WSD ha dado respuesta (número de instancias a las cuales el sistema de WSD propuso una respuesta / número total de instancias en el conjunto de prueba). Por ejemplo, suponiendo que se está manejando un conjunto de prueba con 100 palabras y el sistema WSD sólo es capaz de proponer respuestas para 75 palabras; y de estas 75 sólo 50 palabras han sido desambiguadas correctamente. La precisión en este ejemplo es de $50 / 75 = 0.66$, el recuerdo es $50 / 100 = 0.50$ y por último, la cobertura es de $75 / 100 = 0.75$.

2.2.1 Métodos basados en conocimiento

La idea básica de estos métodos consiste en utilizar recursos externos para desambiguar las palabras, tales como diccionarios, tesauros (diccionarios que

muestran las palabras relacionadas con el significado y sentido de una palabra, como los sinónimos), textos sin ningún tipo de etiquetado e incluso la Web. El propósito de estos recursos dentro de WSD es proveer una lista de significados, definiciones o ejemplos típicos sobre el uso de las palabras. Los diccionarios más populares utilizados por estos métodos son conocidos como MRD (Machine Readable Dictionaries), porque la información que está contenida en estos diccionarios puede ser leída por una computadora. Algunos diccionarios MRD son: Longman Dictionary of Contemporary English (LDOCE), Collins English Dictionary (CED) y WordNet (ver sección 3.3.1). En la tabla 2.2 se muestra un ejemplo de las definiciones de cada sentido de la palabra en inglés “*plant*” obtenidas desde WordNet.

-
1. Buildings for carrying on industrial labor; "they built a large plant to manufacture automobiles".
 2. A living organism lacking the power of locomotion.
 3. Something planted secretly for discovery by another; "the police used a plant to trick the thieves"; "he claimed that the evidence against him was a plant".
 4. An actor situated in the audience whose acting is rehearsed but seems spontaneous to the audience.
-

Tabla 2.2: Definiciones de los sentidos de la palabra en inglés “*plant*” obtenidas desde WordNet 2.0.

Uno de los primeros trabajos en el uso del MRD es el de Lesk [33], quien destacó por sus resultados (50% - 70% de precisión en desambiguar los sentidos correctamente) usando un conjunto de ejemplos pequeños manualmente etiquetados y las definiciones del diccionario “*Oxford Advanced Learner’s Dictionary*” para tratar de identificar el sentido más adecuado. En la tabla 2.3 se describe el algoritmo propuesto por Lesk. El funcionamiento se basa en encontrar la cantidad de coincidencias entre las palabras de las definiciones de dos palabras que se desean desambiguar.

-
1. Adquirir desde un diccionario MRD todas las definiciones de los sentidos de las palabras a desambiguar.
 2. Determinar las coincidencias de las palabras en las definiciones para todas las posibles combinaciones de los sentidos.
 3. Escoger los sentidos en donde las palabras coinciden más.
-

Tabla 2.3: Algoritmo original de Lesk.

En la figura 2.1 muestra un ejemplo de la desambiguación de las palabras en “PINE CONE” usando el algoritmo de Lesk. El criterio para elegir el sentido más adecuado es seleccionar el sentido donde exista el mayor número de coincidencias entre las palabras de las definiciones de ambas palabras. En el caso del ejemplo, se han elegido los sentidos 1 para “PINE” y 3 para “CONE”.

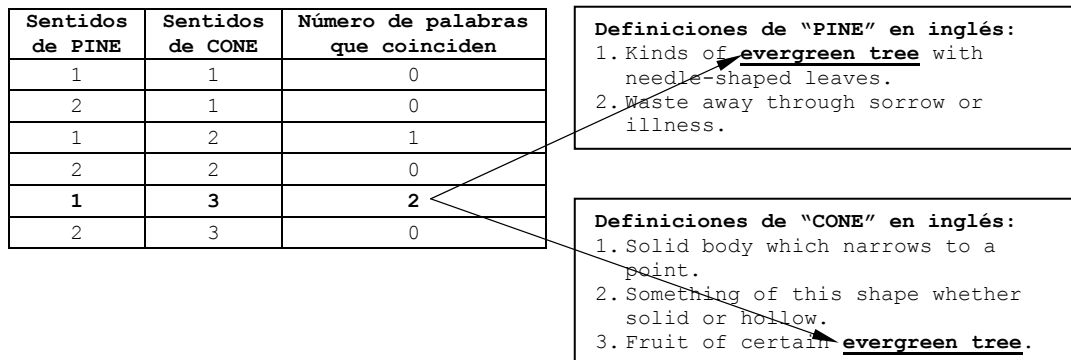


Figura 2.1: Ejemplo del algoritmo original de Lesk:

El problema del algoritmo de Lesk consiste en encontrar una combinación adecuada de los sentidos cuando se trata de desambiguar más de 2 palabras, o bien, con un tamaño de contexto de más de una palabra en uno o ambos lados de la oración o frase. En la figura 2.2 ejemplifica este problema, donde al tratar de desambiguar las palabras polisémicas del texto mostrado en el ejemplo (9 en total), se producen más de 1.9 millones de combinaciones de sentidos. Encontrar una combinación adecuada puede crear mucho tiempo y confusiones.

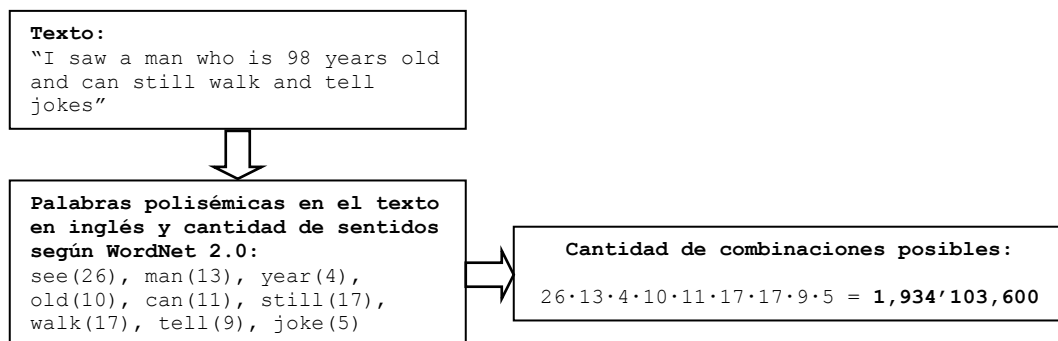


Figura 2.2: El problema de Lesk: la cantidad de combinaciones posibles.

Más tarde Kilgariff y Rosenzweig realizaron un algoritmo simplificado de Lesk [47] para evitar el problema del algoritmo original, modificando el segundo

paso del algoritmo original (ver tabla 2.4): en lugar de buscar coincidencias de palabras que ocurren dentro de las definiciones de los sentidos de las palabras, sólo se realiza la búsqueda entre las palabras de la definición y las palabras que se encuentran en el contexto del uso de la palabra a desambiguar. En la figura 2.3 muestra un ejemplo, donde la palabra objetivo es “*pine*”; utilizando el algoritmo se determina que el sentido es uno porque la palabra “*tree*” esta presente en el texto y en la definición.

-
1. Adquirir desde un diccionario MRD todas las definiciones de los sentidos de la palabra a desambiguar.
 2. Determinar las coincidencias entre las palabras que se encuentran en las definiciones de los sentidos y las palabras usadas en el contexto que rodea a la palabra a desambiguar.
 3. Escoger el sentido en donde las palabras coincidan más.
-

Tabla 2.4: Algoritmo simplificado de Lesk.

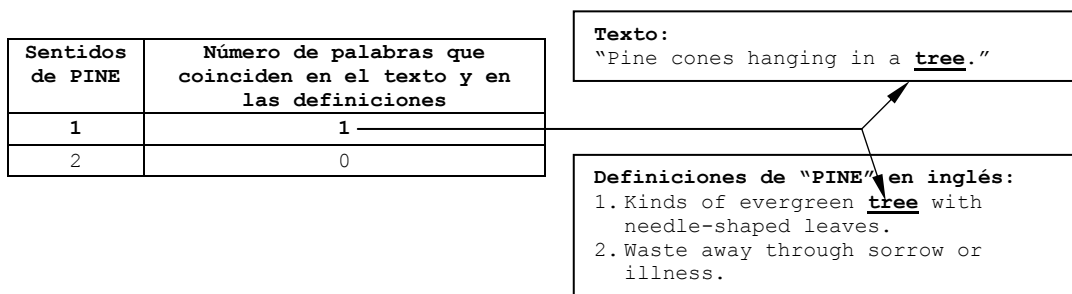


Figura 2.3: Ejemplo del algoritmo simplificado de Lesk: Desambiguación del sentido de la palabra “PINE”.

En los experimentos realizados por Vasilescu, et. al. [48], se muestra una comparación entre el algoritmo original de Lesk y el algoritmo simplificado usando el conjunto de datos de Senseval-2 All-Words (ver sección 2.3.2). Los resultados obtenidos en ese trabajo demuestra que el algoritmo simplificado es superior al original: 58% de desambiguación correcta contra el 42% de precisión generado por el algoritmo original.

Otro recurso potencial para este enfoque es la Web (véase el trabajo de Gonzalo et. al. [37] para más detalles) y entre los primeros investigadores en usar la Web como un recurso léxico para desambiguar palabras se encuentra Agirre et. al. [36]. El trabajo básicamente consistió en crear colecciones de documentos similares

extraídas desde la Web agrupadas por las palabras que están relacionadas con los sentidos de la palabra a desambiguar (por ejemplo, los sinónimos obtenidos por WordNet), de esta manera, cada colección representa un sentido. Después se usaron estas colecciones como fuente de conocimiento y compararlas con el contexto de la palabra objetivo. Para realizar la comparación se utilizaron las colecciones en lugar de las definiciones del diccionario. Usando un conjunto de 20 palabras distintas y 2444 palabras a desambiguar para prueba, el resultado obtenido en ese trabajo fue de 41% de precisión en la desambiguación correcta de los sentidos. Por otro lado, Rosso et. al. [38] utilizó la Web para encontrar directamente co-ocurrencias de las palabras del contexto y de los sinónimos de la palabra objetivo obtenido en WordNet. Usando el corpus de Senseval-3 (ver sección 2.3.2), el resultado de este método fue de 77% de precisión., sin embargo el recuerdo que se obtuvo fue de 33.7%. A pesar de tener estos resultados en este último trabajo de este enfoque, es novedoso y prometedor al usar la Web como un conocimiento externo.

Los métodos de este enfoque poseen la gran desventaja de depender de información ofrecida por un MRD. Construir estos recursos es muy laborioso y muy costoso de construir. Además se pueden mencionar entre sus limitantes las siguientes:

- La información que poseen los diccionarios es inconsistente (ver [49]), es decir, la cantidad de sentidos y las definiciones que posee una palabra puede variar. Además, la mayoría de los MRD son comerciales a excepción de WordNet.
- En el algoritmo de Lesk y similares, la ausencia o la presencia de las palabras que hay en la definición del significado de un sentido puede afectar en los resultados, es decir, la calidad de las definiciones es de gran importancia.

2.2.2 Métodos basados en corpus

Un corpus es un conjunto de textos recopilados, ya sea de un mismo tema o varios. El propósito de un corpus es convertirse en un conjunto de datos para proveer ejemplos de oraciones y ejemplos de uso de varias palabras para ser utilizados en algoritmos de aprendizaje automático. Senseval-3 es un ejemplo de este tipo de corpus (ver sección 2.3.2 para la descripción detallada de este corpus). Dependiendo de la naturaleza de los algoritmos y de la tarea, las palabras que hay en un corpus puede estar previamente desambiguadas o no. Los métodos basados en corpus puede subdividirse en dos, al usar: (1) métodos supervisados – corpus etiquetado; y (2) métodos no supervisados – corpus no etiquetado.

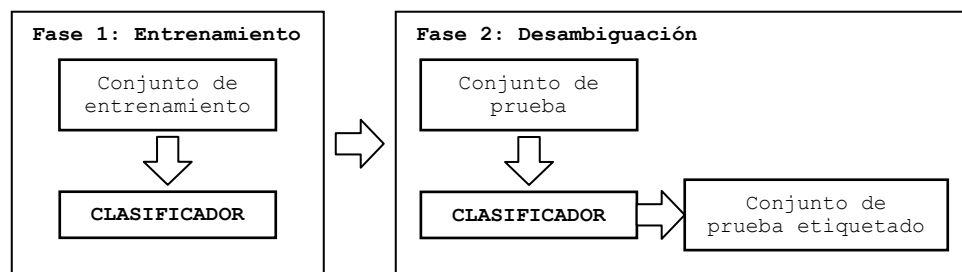


Figura 2.4: Esquema general de los métodos supervisados.

2.2.2.1 Métodos supervisados – corpus etiquetado

En este enfoque, WSD se reduce a un problema de clasificación (ver sección 2.4) donde se asigna a una palabra objetivo el sentido más apropiado dado un conjunto de posibles combinaciones de las palabras del contexto [12]. En otras palabras, estos métodos utilizan clasificadores (o algoritmos de aprendizaje, ver sección 2.4) para llevar a cabo la desambiguación y generalmente se divide en dos fases (ver figura 2.4): el entrenamiento y la desambiguación de los sentidos (o clasificación en términos de aprendizaje automático). En la fase del entrenamiento es requerido un conjunto de datos etiquetados para el entrenamiento del clasificador, es decir, un conjunto de palabras y el sentido para cada una de esas palabras bajo contextos de uso dados. (Véase la sección 2.3.2 para un ejemplo de este tipo de corpus *Senseval-3*

English Lexical Samples). Por esta razón, estos métodos son conocidos como métodos supervisados. El propósito del entrenamiento del clasificador es permitir que el algoritmo encuentre automáticamente relaciones o patrones entre las palabras del conjunto de entrenamiento para poder realizar la desambiguación de los sentidos de esas palabras. Una vez entrenado el clasificador, se procede a seguir con la segunda fase, la desambiguación. El clasificador intentará desambiguar las palabras de un conjunto de datos etiquetados que están destinados para la evaluación. A este conjunto se le llama conjunto de prueba, en donde la etiqueta será ignorada para propósitos de clasificación. Al final de la clasificación se evalúan los resultados al comparar los sentidos propuestos por el clasificador y los sentidos correctos.

Los conjuntos de datos (para entrenamiento y prueba) para los métodos supervisados suelen ser representados por atributos. Los atributos son características que aportan información sobre el contexto para desambiguar el sentido de la palabra. Algunos atributos comunes (ver [10] y [12]) son:

- **Co-ocurrencias de las palabras.** Estos atributos son palabras que están dentro del contexto y ayudan a definir el sentido de la palabra objetivo.
- **N-gramas.** Un n-grama es una secuencia de n palabras que provienen de una oración o texto, en el caso de WSD, del contexto. El orden de estas palabras es el mismo que se encuentra en el contexto. Por ejemplo, de la oración: “*hola gran mundo*”; “*hola gran*” y “*gran mundo*” son bigramas, pues están formados por dos palabras consecutivas.
- **Colocaciones.** Las colocaciones son parecidas a las n-gramas, son secuencias significativas de palabras que juntas crean un significado distinto. Estas colocaciones pueden ser similares cuando se trata del mismo sentido de una palabra.
- **Etiquetas de parte de la oración.** En lugar de usar palabras que representen el contexto, se identifica la categoría gramatical de la palabra en el fragmento (p. e. verbo, sustantivo, adverbio, etc). La categoría de una palabra puede variar según el sentido de la misma. Por ejemplo, en la frase

“la ayuda recibida” la palabra ayuda es un sustantivo; en contraste, en la frase *“mi hermano ayuda a su novia”* se trata de un verbo.

Entre los trabajos de este tipo de métodos se destaca el de Weiss [18] en 1973, quien ha demostrado que es posible adquirir reglas de desambiguación por medio de corpus etiquetados. Weiss trabajó con un conjunto pequeño de 5 palabras y 20 oraciones para el entrenamiento y 30 oraciones para prueba. A pesar del tamaño de este conjunto, los resultados son alentadores (aproximadamente 90% de la desambiguación fue correcta). Más tarde, en 1998, Pedersen [20] realizó en su tesis una comparación entre varios algoritmos para el aprendizaje supervisado usando un conjunto de datos para tratar de desambiguar 13 palabras. La conclusión de este trabajo mostró que los algoritmos probabilísticos (Naïve Bayes (ver sección 2.4.1) y variantes) han tenido más éxito que los otros algoritmos de aprendizaje supervisado con un resultado aproximado de 84% de precisión. Por otra parte, G. Paliouras [4] realizó otra comparación de algoritmos supervisados en la que los algoritmos basados en árboles de decisión (por ejemplo C4.5, ver sección 2.4.2) fueron mejores (82.6% de precisión y 77.4% de recuerdo) para este último trabajo se usó un corpus con 3516 ejemplos de 355 palabras, con un promedio de 4.67 sentidos por palabra. Otros ejemplos de métodos supervisados se pueden encontrar en [23], [24], [25], [26] y [27]. Por último cabe mencionar, que en trabajos recientes [2] se ha mostrado que este método ha llegado a ser muy eficaz usando el corpus de SENSEVAL-3 (ver sección 2.2.1) el cual posee ejemplos sobre 57 palabras (7860 ejemplos para el entrenamiento y 3944 ejemplos para prueba). Entre los mejores resultados usando este conjunto, fue el sistema diseñado por C. Grozea de la Universidad de Bucarest usando el algoritmo de Naïve Bayes. Este sistema generó un resultado de 72.9% en precisión y recuerdo.

A pesar de los buenos resultados que obtienen estos métodos, su principal desventaja radica en el uso de un corpus previamente etiquetado. Por un lado, la creación de este tipo de conjuntos de datos es de un alto costo en términos de tiempo y de recursos humanos. Por otro lado, la cobertura del sistema siempre estará limitada

al conjunto de palabras del corpus de entrenamiento así como a los sentidos etiquetados.

2.2.2.2 Métodos no supervisados – corpus no etiquetado

Al contrario de los métodos supervisados y los basados en conocimientos, los métodos no supervisados identifican patrones en los conjuntos de datos sin el beneficio de los datos etiquetados o de otros recursos como son los MRD (ver [8] y [12]). Estos patrones son utilizados para dividir los datos en grupos, donde cada uno de los miembros de un grupo posee varias características en común con el resto de los miembros del mismo grupo. Este enfoque se basa en la hipótesis de que las palabras con significados similares tienden a tener contextos similares. En otras palabras, este enfoque utiliza métodos que agrupan palabras basándose en la similitud del contexto y cada grupo representa un sentido. Al no contar con información sobre los sentidos de una palabra, la tarea se convierte en *Discriminación de los Sentidos de la Palabra* y sólo poseen un conjunto de datos no etiquetados como único recurso.

La metodología general de este enfoque típicamente se basa en la selección de aquellas palabras que se desean discriminar con su respectivo contexto. A partir de estas palabras seleccionadas se realizan las agrupaciones basándose en el contexto de las palabras seleccionadas. La figura 2.5 ilustra este proceso. Suponiendo que las figuras son las palabras que se desean agrupar, se elige una figura que represente a las figuras que se desea para la agrupación (en este caso, un círculo, un rombo y un cuadrado), donde las figuras que son iguales representan a aquellas palabras que tienen un contexto similar. Una vez elegido las palabras, el método agrupa todas las palabras que sean parecidos a las palabras que se eligieron (en el caso de la figura 2.5 son 3 grupos).

En [21] se muestra uno de los trabajos de Discriminación de los Sentidos de la Palabra de T. Pedersen, quien utilizó 13 palabras para crear 3 grupos (3 sentidos) obteniendo un resultado promedio de 65% de precisión en donde las palabras correctas se encuentran en las agrupaciones apropiadas. Otros trabajos bajo este enfoque se pueden apreciar en [30], [31] y [32].

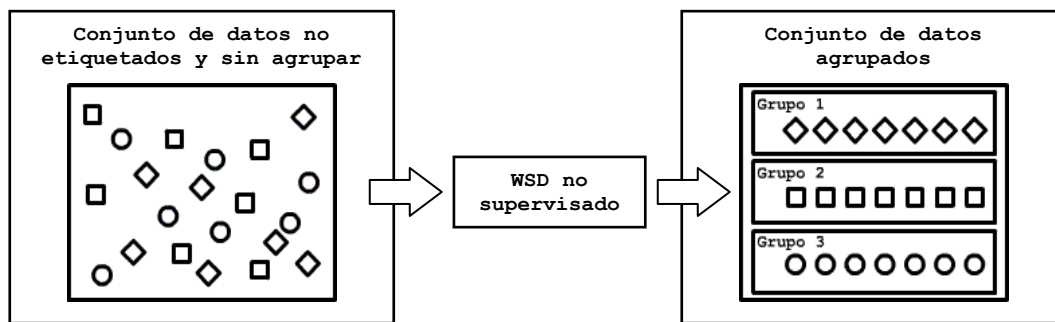


Figura 2.5: Esquema general de los métodos no supervisados.

Las ventajas de este enfoque son la calidad de agrupamiento a pesar de la falta de información de los sentidos de las palabras en otros recursos y la independencia hacia los datos etiquetados. Sin embargo, estas agrupaciones no representan los verdaderos sentidos y pueden variar también la cantidad de sentidos. Por tanto, este enfoque no ayuda del todo a desambiguar el sentido de las palabras.

2.2.3 Métodos híbridos y Bootstrapping

Como pudo observarse los métodos antes mencionados tienen ventajas y desventajas, una manera para reducir las desventajas es la combinación de estos métodos resultando en métodos híbridos. Un ejemplo de este tipo de métodos es el trabajo de Luk [35], quien creó un sistema que utiliza información de las definiciones de diccionarios (basados en conocimiento) e información estadística que provienen de un corpus no etiquetado (métodos no supervisados). El resultado de este sistema fue de 77% de precisión y 100% de cobertura al tratar de desambiguar 616 ejemplos sobre 12 palabras distintas. Otro ejemplo es el trabajo de Mihalcea y Moldovan [13] en donde la Web es utilizada como un conjunto de datos no etiquetados y al mismo tiempo como una fuente de conocimiento. Con la ayuda de WordNet, en ese trabajo se extrajeron ejemplos desde la Web y fueron etiquetados automáticamente los sentidos basados en WordNet. Bajo el juicio humano, el resultado reportado en este último trabajo fue de un 92% de precisión de 658 ejemplos revisados manualmente.

Bootstrapping es un enfoque que consiste en utilizar dos conjuntos de datos: uno etiquetado y otro no etiquetado. El objetivo de este enfoque es etiquetar los

sentidos de las palabras que se encuentran en el conjunto de datos no etiquetados a partir de un pequeño conjunto inicial de datos etiquetados. El algoritmo propuesto por Yarowsky en 1995 es un ejemplo muy claro de este tipo de enfoque [22]: usando 12 palabras con 2 sentidos cada uno y con una cantidad promedio de 3936 ejemplos, en el mejor de los casos logró obtener el 96.5% de precisión.

2.3 Recursos Lingüísticos

En esta sección se describen los recursos lingüísticos más utilizados en la tarea de WSD: WordNet y Senseval-3 ELS. Ambos recursos aquí expuestos fueron utilizados en los experimentos del presente trabajo.

2.3.1 WordNet

WordNet es un diccionario MRD para el idioma inglés (ver [50], [51], y [52]) convirtiéndose en uno de los recursos más valiosos para el PLN. El desarrollo de WordNet inició en 1985 y fue creado en el Laboratorio de Ciencias Cognitivas de las Universidad de Princeton bajo la dirección del Profesor de Psicología George A. Miller. Este recurso posee una base de datos que agrupa las palabras en conjuntos de sinónimos llamados *synsets* y provee definiciones, comentarios y ejemplos de uso de estas palabras y sentidos de las mismas. De esta manera, combina los elementos de un diccionario (definiciones y algunos ejemplos) y los de un tesoro (sinónimos), y crea un apoyo muy importante para el análisis automático de textos y palabras.

Actualmente, la base de datos de WordNet contiene alrededor de 150,000 palabras organizadas en más de 115,000 *synsets* formando un total de más 203,000 definiciones y sentidos de las palabras. WordNet maneja 4 categorías léxicas (o tipos de partes de la oración) en sus *synsets*: sustantivos, verbos, adjetivos y adverbios. Cada uno de los *synsets* contiene un grupo de sinónimos o colocaciones. En la tabla 2.5 se muestra un ejemplo típico donde se muestran los *synsets* de la palabra “*plant*” como sustantivo junto con sus definiciones de cada uno de los sentidos de esta palabra.

The noun plant has 4 senses (first 3 from tagged texts)

1. (338) **plant**, works, industrial plant -- (buildings for carrying on industrial labor; "they built a large plant to manufacture automobiles")
 2. (207) **plant**, flora, plant life -- (a living organism lacking the power of locomotion)
 3. (2) **plant** -- (something planted secretly for discovery by another; "the police used a plant to trick the thieves"; "he claimed that the evidence against him was a plant")
 4. **plant** -- (an actor situated in the audience whose acting is rehearsed but seems spontaneous to the audience)
-

Tabla 2.5: Ejemplos de synsets y definiciones de la palabra “*plant*”.

Además, la mayor parte de los synsets están conectados a otros synsets bajo una red de relaciones semánticas. Entre estas relaciones se tienen las siguientes:

- **Sinónimos.** Palabras con significados idénticos o similares.
- **Antónimos.** Palabras con significados opuestos.
- **Hipéronimos.** Palabras que son más generales que otras en su significado. Por ejemplos, “*vehículo*” es el hipéronimo de las palabras “*tren*”, “*automóvil*”, “*aeroplano*” y “*motocicleta*”.
- **Hipónimos.** Palabras que son más específicas que otras en su significado. Por ejemplo, “*tulipán*”, “*rosa*” y “*girasol*” son hipónimos de la palabra “*flor*”.
- **Holónimos.** Palabras que denotan ser una parte o miembro de un todo (denotado por otra palabra). Por ejemplo “*auto*” es el holónimo de las palabras “*llantas*”, “*motor*” y “*volante*”.
- **Merónimos.** Palabras que denotan ser un todo de las palabras que denotan ser una parte o miembro de ese todo. Por ejemplo “*llantas*”, “*motor*” y “*volante*” son merónimos de “*auto*”.

A modo de ejemplo, en la tabla 2.6 se muestra los hipéronimos de la palabra “*plant*” para su segundo sentido (planta en su sentido de flora) desde WordNet, como se puede apreciar en la tabla, el hipéronimo más general es “*entity*” (entidad).

Sense 2

plant, flora, plant life -- (a living organism lacking the power of locomotion)

=> **organism, being** -- (a living thing that has (or can develop) the ability to act or function independently)

=> **living thing, animate thing** -- (a living (or once living) entity)

=> **object, physical object** -- (a tangible and visible entity; an entity that can cast a shadow; "it was full of rackets, balls and other objects")

=> **entity, physical thing** -- (that which is perceived or known or inferred to have its own physical existence (living or nonliving))

Tabla 2.6: Ejemplos de synsets hiperonímicos y sus definiciones sobre el segundo sentido de la palabra "*plant*".

A pesar de algunas limitaciones de WordNet (por ejemplo: las inconsistencias en las definiciones y falta de relaciones semánticas, ver [52] para más detalles), se ha convertido en un recurso indispensable para WSD, sobre todo para los enfoques basados en conocimientos e híbridos, ya que cuenta con información sobre los sentidos de las palabras. Los trabajos de Rosso y Mihalcea mencionados en las secciones 2.2.1 y 2.2.3 respectivamente, son ejemplos del uso de este recurso.

2.3.2 SENSEVAL

Como se observó en las secciones previas los resultados de los diversos enfoques son difíciles de comparar entre sí. La evaluación de estos métodos no sólo depende del enfoque mismo (v. g. se tienen o no ejemplos de entrenamiento y/o ejemplos de evaluación) también depende del número de palabras a desambiguar, del número de sentidos considerados para cada palabra, del tamaño del corpus, del tipo de etiquetado, de la categoría gramatical de la palabra a desambiguar, del algoritmo de aprendizaje utilizado, entre otros. Así la comparación objetiva de los resultados alcanzados por cada método propuesto es una tarea difícil. De ahí que se hayan propuesto foros de evaluación, que bajo un mismo esquema de evaluación comprueba el desempeño de los diferentes métodos propuestos.

En 1998, Adam Kilgarriff propuso un esquema de evaluación para los sistemas WSD: SENSEVAL (ver [53]). SENSEVAL surge como una organización dedicada a la evaluación de sistemas WSD para identificar claramente las ventajas y desventajas de los sistemas participantes. Hasta ahora, sólo se han realizado 3 eventos: SENSEVAL-1 (Inglaterra, 1998, ver [54]), SENSEVAL-2 (Francia, 2001, ver [55]) y SENSEVAL-3 (España, 2004). En este último evento, se definieron 2 tipos de tareas para WSD con su respectivo corpus de referencia diseñado específicamente para estas tareas: All-Words Task y English Lexical Sample. A continuación se describe cada una de estas tareas.

1. **All-Words Task.** La meta de esta tarea es desambiguar todas las palabras polisémicas posibles que hay dentro de un texto (ver [56]). El conjunto de datos para esta tarea posee todas las palabras etiquetas que se desean desambiguar para propósitos de entrenamiento y evaluación de los sistemas. En la tabla 2.6 se muestra una frase de este corpus “*His shout had been involuntary*”, donde las palabras polisémicas marcadas en negritas son las palabras que se desean desambiguar. Las etiquetas <head id=...> se utilizan para identificar las palabras a desambiguar. Es importante remarcar que este corpus dentro de la modalidad en el idioma inglés (English All-words) posee 2212 palabras polisémicas etiquetadas con el sentido apropiado.

```

<?xml version="1.0"?>
<!DOCTYPE corpus SYSTEM "all-words.dtd">
<corpus lang="en">

<text id="d000">
.
.
His
<head id="d000.s014.t001">shout</head>
had
<head id="d000.s014.t003">been</head>
<head id="d000.s014.t004">involuntary</head>
.
.
</text>
.
.

```

Tabla 2.7: Extracto del corpus de la tarea All-Words en inglés.

2. Lexical Sample Task. El objetivo de esta tarea es desambiguar una sola palabra polisémica (ver [57]). Esta palabra es la única unidad léxica que está etiquetada con su sentido correspondiente (ver tabla 2.8). Para el idioma inglés, el corpus es conocido como SENSEVAL-3 English Lexical Sample (ELS). Este conjunto contiene más de 11,000 ejemplos sobre 57 palabras específicas (7860 ejemplos para entrenamiento y 3944 para prueba). En la tabla 2.9 se muestra el contenido de este corpus. La palabra objetivo se muestra en negritas y marcada por la etiqueta <head>. La etiqueta <answer> contiene el sentido correcto de la palabra a desambiguar, donde el parámetro *instante* es una clave para identificar el texto y *senseid* indica el sentido. Los valores de *senseid* están codificados y SENSEVAL proporciona los respectivos sentidos de WordNet equivalentes. Los ejemplos de este corpus fueron extraídos del corpus British Nacional Corpus (BNC), el cual posee más de 100 millones de palabras. Los ejemplos fueron etiquetados manualmente (ver [57] y [58]).

```

<lexelt item="image.n">
.
.
<instance id="image.n.bnc.00003870" docsrc="BNC">
<answer instance="image.n.bnc.00003870" senseid="image%1:06:00::"/>
<context>
After all , it is an original work in itself . Nevertheless , the
differences between painting and print may be instructive , and help
the interpretation of both . By other forms of reproduction an
<head>image</head> may be more or less degraded , so that nothing can be
learnt from them . There remains the courtesy paid by one art to another
, as in the poem which Baudelaire suggested might be a high form of
critical writing . Or it may be music , as in Mussorgsky 's
composition which he entitled Pictures from an Exhibition .
</context>
</instance>
.
.
</lexelt>

```

Tabla 2.8: Extracto del corpus de la tarea English Lexical Samples en inglés.

Dentro de la tarea de “*Lexical Sample*”, participaron 47 sistemas, donde el mejor sistema (ver [57]) entre los sistemas basados en métodos supervisados se obtuvo un 72.9% de precisión y recuerdo; mientras que el peor sistema obtuvo en esta categoría fue de 78.2% en precisión y 31% en recuerdo. Por otra parte, en la categoría

de los sistemas basados en métodos no supervisados o en conocimiento el mejor fue de 66.1% en precisión y 65% en recuerdo; mientras que el peor sistema sólo logró alcanzar el 19.7% en precisión y 11.7% en recuerdo.

La tabla 2.9 muestra el inventario de las palabras utilizadas para la tarea de “*Lexical Sample*”. Para cada palabra se indica su categoría gramatical o parte de la oración (P. O.); el número de sentidos (# sent.); el número de ejemplos de entrenamiento (# ej. ent) y el número de ejemplos de prueba para su evaluación (#ej. prueba). El inventario contiene 57 palabras, donde 20 son sustantivos, 32 son verbos y 5 son adjetivos; la cantidad promedio de los sentidos es de 6.47 sentidos por palabra. Los conjuntos para cada palabra se dividen en dos tipos: para el entrenamiento, y para prueba. El primero es necesario para los métodos supervisados.

Palabra	P.O.	# sent.	# ej. ent.	# ej. prueba	Palabra	P.O.	# sent.	# ej. ent.	# ej. prueba
activate	V	5	228	114	Miss	V	8	58	30
add	V	6	263	132	Note	V	3	132	67
appear	V	3	265	133	Operate	V	5	35	18
argument	S	5	221	111	organization	S	7	112	56
arm	S	6	266	133	paper	S	7	232	117
ask	V	6	261	131	party	S	5	230	116
atmosphere	S	6	161	81	performance	S	5	172	87
audience	S	4	200	100	plan	S	3	166	84
bank	S	10	262	132	play	V	12	104	52
begin	V	4	181	79	produce	V	6	186	94
climb	V	5	133	67	provide	V	6	136	69
decide	V	4	122	62	receive	V	9	52	27
degree	S	7	256	128	remain	V	3	139	70
difference	S	5	226	114	rule	V	4	59	30
different	A	5	98	50	shelter	S	5	196	98
difficulty	S	4	46	23	simple	A	7	36	18
disc	S	4	200	100	smell	V	7	108	55
eat	V	7	181	87	solid	A	14	58	29
encounter	V	4	130	65	sort	S	4	190	96
expect	V	3	156	78	source	S	7	64	32
express	V	4	110	55	suspend	V	7	128	64
hear	V	7	63	32	talk	V	9	146	73
hot	A	22	86	43	treat	V	9	112	57
image	S	7	146	74	use	V	5	26	14
important	A	5	36	19	wash	V	12	66	34
interest	S	7	185	93	watch	V	7	100	51
judgment	S	7	62	32	win	V	7	78	39
lose	V	9	71	36	write	V	8	44	23
mean	V	7	80	40					
TOTAL DE EJEMPLOS								7860	3944

Tabla 2.9: Inventario de palabras que contiene SENSEVAL-3 ELS.

2.4 Algoritmos de Aprendizaje Automático

El Aprendizaje Automático es la disciplina que estudia cómo construir sistemas computacionales que mejoren automáticamente mediante la experiencia. En otras palabras, se dice que un programa ha adquirido conocimiento (o “aprendido”) para realizar una tarea específica T si después de proporcionarle la experiencia E mediante un conjunto de ejemplos de T el sistema es capaz de desempeñarse cuando se presentan nuevas situaciones de la tarea. El desempeño es medido usando una métrica de calidad P (precisión y recuerdo). Por lo tanto, un problema de aprendizaje bien definido requiere que T , E y P estén bien especificados (referirse a [3] para un estudio más detallado). En esta disciplina, se han desarrollado varios algoritmos para realizar el proceso de aprendizaje, por esta razón reciben el nombre de *algoritmos de aprendizaje*.

En general, estos algoritmos son utilizados para el proceso de *clasificación*. Este proceso puede ser formalizado como la tarea de aproximar una *función objetivo* desconocida $\Phi: I \times C \rightarrow \{V, F\}$ (que describe cómo las instancias del problema deben ser clasificadas de acuerdo a un experto en el dominio) por medio de una función $\Theta: I \times C \rightarrow \{V, F\}$ llamada el *clasificador*, donde $C = \{c_1, \dots, c_{|c|}\}$ es un conjunto predefinido de categorías o clases, I es un conjunto de instancias del problema, y $\{V, F\}$ son los valores de *Verdadero* y *Falso* respectivamente. Comúnmente cada instancia $i_j \in I$ es un ejemplo que está representado como una lista $A = \langle a_1, a_2, \dots, a_{|A|} \rangle$ de valores característicos, conocidos como *atributos*, por ejemplo $i_j = \langle a_{1j}, a_{2j}, \dots, a_{|A|j} \rangle$. Si $\Phi: i_j \times c_i \rightarrow V$, entonces i_j es llamado un ejemplo positivo de c_i , mientras si $\Phi: i_j \times c_i \rightarrow F$ éste es llamado un ejemplo negativo de c_i .

Para generar automáticamente el clasificador de c_i es necesario realizar un proceso inductivo, llamado el *aprendiz*, el cual por observar los atributos de un conjunto de instancias preclasificadas bajo c_i o \bar{c}_i , adquiere los atributos que una instancia desconocida debe tener para pertenecer a la categoría. Por tal motivo, en la

construcción del clasificador se requiere la disponibilidad inicial de una colección Ω de ejemplos tales que el valor de $\Phi(i_j, c_i)$ es conocido para cada $\langle i_j, c_i \rangle \in \Omega \times C$. A la colección usualmente se le llama *conjunto de entrenamiento* (Tr). En resumen, al proceso anterior se le identifica como *aprendizaje supervisado* debido a la dependencia de Tr .

Es por esta razón, que los métodos de WSD que usen el aprendizaje supervisado se le conocen como *métodos WSD supervisados*. A continuación se detallan brevemente los algoritmos utilizados en el presente trabajo, dichos algoritmos fueron seleccionados a partir de un estudio de diferentes clasificadores utilizados en la tarea de WSD (el estudio se resume en el apéndice 1).

2.4.1 Clasificador Naïve Bayes

El clasificador *Naïve Bayes* (NB) se considera como parte de los clasificadores probabilísticos, los cuales se basan en la suposición que las cantidades de interés se rigen por distribuciones de probabilidad, y que la decisión óptima puede tomarse por medio de razonar acerca de esas probabilidades junto con los datos observados [3]. Dentro de WSD, este algoritmo se encuentra entre los más utilizados (ver [12]). En [42] se presenta una guía básica de las diferentes direcciones que han tomado las investigaciones sobre Naïve Bayes, las cuales se caracterizan por las modificaciones realizadas al algoritmo. En este trabajo empleamos el Naïve Bayes tradicional, el cual se describe a continuación.

En este esquema el clasificador es construido usando el conjunto de entrenamiento Tr para estimar la probabilidad de cada clase. Entonces, cuando una nueva instancia i_j es presentada, el clasificador le asigna la categoría $c \in C$ más probable aplicando la regla:

$$c = \arg \max_{c_i \in C} P(c_i | i_j), \quad (2.1)$$

utilizando el teorema de Bayes para estimar la probabilidad tenemos

$$c = \arg \max_{c_i \in C} \frac{P(i_j | c_i) P(c_i)}{P(i_j)}, \quad (2.2)$$

el denominador en la ecuación anterior no difiere entre categorías y puede omitirse

$$c = \arg \max_{c_i \in C} P(i_j | c_i) P(c_i), \quad (2.3)$$

tomando en cuenta que el esquema es llamado “naïve” debido al supuesto de independencia entre atributos, por ejemplo, se asume que las características son condicionalmente independientes dadas las clases. Esto simplifica los cálculos produciendo

$$c = \arg \max_{c_i \in C} P(c_i) \prod_{k=1}^n P(a_{kj} | c_i), \quad (2.4)$$

donde $P(c_i)$ es la fracción de ejemplos en Tr que pertenecen a la clase c_i , y $P(a_{kj} | c_i)$ se calcula de acuerdo al teorema de Bayes.

En resumen, la tarea de aprendizaje en el clasificador Naïve Bayes consiste en construir una hipótesis por medio de estimar las diferentes probabilidades $P(C_i)$ y $P(a_{kj} | c_i)$ en términos de sus frecuencias sobre Tr . En [20] y [39] se presentan una descripción detallada de los cálculos.

2.4.2 C4.5

El esquema *C4.5* fue diseñado como una extensión del algoritmo *ID3* [43], este último forma parte de los clasificadores conocidos como *árboles de decisión*, los cuales son árboles donde sus nodos internos son etiquetados como atributos, las ramas salientes de cada nodo representan pruebas para los valores del atributo, y las hojas del árbol identifican a las categorías. Estos algoritmos proporcionan un método práctico para aproximar conceptos y funciones con valores discretos. En [4], [27] son ejemplos en donde el algoritmo es usado. A continuación se presenta la descripción del algoritmo *ID3* con el objetivo de facilitar la posterior descripción de *C4.5* (para más detalles refiérase a [44]).

Para construir el árbol, *ID3* usa una aproximación descendente que da preferencia a los árboles pequeños sobre los grandes. El nodo raíz es seleccionado por poseer el atributo más valioso en el conjunto de entrenamiento, es decir, aquél atributo con el que mejor clasifica las instancias; la búsqueda se realiza por medio de

una prueba estadística que mide qué tan bien un atributo separa el conjunto de entrenamiento de acuerdo a las clases. Una vez que la raíz es seleccionada, se agrega una rama desde la raíz para cada posible valor del atributo correspondiente, y el conjunto de entrenamiento es ordenado en los nodos apropiados, por ejemplo cada nodo contiene los ejemplos que cumplen la restricción de la rama anterior. Para seleccionar el atributo más valioso en cada punto del árbol, se repite el proceso entero usando el conjunto de entrenamiento asociado con el nodo. De manera que cuando una nueva instancia necesita ser clasificada, los atributos especificados por los nodos son evaluados iniciando por el nodo raíz, entonces de manera descendente se recorren las ramas del árbol que corresponden a los valores de los atributos en la instancia dada, el proceso se repite hasta que una hoja es alcanzada, y es en este punto donde la etiqueta asociada a la hoja es asignada a la nueva instancia como su categoría.

ID3(Tr , c_i , A)

1. Crear un nodo **raíz** para el árbol
 2. Si todos los ejemplos en Tr son positivos, regresar el árbol con el único nodo **raíz** etiquetado como c_i
 3. Si todos los ejemplos en Tr son negativos, regresar el árbol con el único nodo **raíz** etiquetado como \bar{c}_i
 4. Si la lista A está vacía, regresar el árbol con el único nodo **raíz** etiquetado como el valor de c_i más frecuente en Tr
 5. En otro caso comenzar
 6. Sea a el atributo en A que mejor clasifica a Tr
 7. Etiquetar a la **raíz** como a
 8. Para cada posible valor v_i de a (por ejemplo, $v_i \in \text{Valores}(a)$)
 - a. Agregar una rama bajo el nodo raíz correspondiente a la prueba $a = v_i$
 - b. Sea Tr_{v_i} el subconjunto de ejemplos para los que $a = v_i$
 - c. Si Tr_{v_i} está vacío
 - i. Agregar debajo de la rama un nodo con el valor de c_i más frecuente en Tr como etiqueta
 - d. De lo contrario
 - i. ID3(Tr_{v_i} , c_i , $A - \{a\}$)
 9. Terminar
 10. Regresar raíz
-

Tabla 2.10: Algoritmo de ID3.

En la tabla 2.10 se resume el algoritmo ID3, donde la medida tradicional para encontrar el atributo más valioso es la *ganancia en la información* (GI), que mide qué tan bien un atributo dado separa el conjunto de entrenamiento conforme a las clases.

En la ecuación 2.5 se presenta la forma de calcular la ganancia en la información del atributo a respecto a Tr .

$$GI(Tr, a) \equiv Entropía(Tr) - \sum_{v_i \in \text{Valores}(a)} \frac{|Tr_{v_i}|}{|Tr|} Entropía(Tr_{v_i}), \quad (2.5)$$

donde $Entropía(S)$ se calcula de la siguiente manera:

$$Entropía(S) \equiv \sum_{i=1}^{|C|} -P(c_i) \log_2 P(c_i) \quad (2.6)$$

Finalmente, una vez introducido ID3 los pasos a seguir en C4.5 son :

1. Separar Ω en conjunto de entrenamiento y conjunto de validación.
2. Construir el árbol de decisión para el *conjunto de entrenamiento* (aplicar ID3).
3. Convertir el árbol en un conjunto de reglas equivalente, donde el número de reglas es igual al número de posibles rutas desde la raíz a los nodos hoja.
4. Podar cada regla eliminando precondiciones que resulten en mejorar la exactitud en el *conjunto de validación*.
5. Ordenar las reglas descendientemente de acuerdo a su exactitud, y usarlas en ese orden para clasificar futuros ejemplos.

2.4.3 k-Vecinos Más Cercanos

k-Vecinos más cercanos (k -NN, por sus siglas en inglés) es uno de los métodos de aprendizaje basados en instancias más básicos, pero con resultados aceptables en tareas que involucran en la tarea de WSD (ver [4] y [64]). En resumen, este algoritmo no tiene una fase de entrenamiento fuera de línea, por lo tanto, el principal cálculo se da en línea cuando se localizan los k vecinos más cercanos. La idea en el algoritmo es almacenar el conjunto de entrenamiento, de modo tal que para clasificar una nueva instancia, se busca en los ejemplos almacenados casos similares y se asigna la clase más probable en éstos.

En la tabla 2.11 se resume el algoritmo, aquí una manera común de encontrar los k ejemplos más cercanos a la instancia i_q es por medio de la *distancia Euclidiana*, donde la distancia entre las instancias i_j e i_q es definida por la siguiente ecuación:

$$d(i_j, i_q) \equiv \sqrt{\sum_{k=1}^{|A|} (a_{kj} - a_{kq})^2} \quad (2.7)$$

2.4.4 Regresión Localmente Ponderada (LWR)

La *Regresión Localmente Ponderada* (o en sus siglas en inglés *LWR*) al igual que *k-NN*, pertenece a la familiar de los algoritmos basados en instancias (ver en [3] para un estudio). A diferencia de *k-NN*, *LWR* utilizar las cercanías o las distancias ya pesadas de los ejemplos de entrenamiento para formar una aproximación local a la función objetivo f . Para aproximar f en el vecindario que rodea a la instancia a clasificar x_q es posible utilizar cualquier función tal como las funciones lineales, funciones cuadráticas, etc. En el presente trabajo se utiliza una función lineal para realizar la aproximación.

Entrenamiento:

1. Para cada ejemplo en Tr , agregar el ejemplo a la lista `ejemplos_entrenamiento`

Clasificación:

1. Dada una instancia de prueba i_q a ser clasificada,
2. Sean i_1, \dots, i_k los k ejemplos de la `lista_entrenamiento` que son más cercanos a i_q
3. Regresar

$$c = \arg \max_{c_j \in C} \sum_{j=1}^k \delta(c_i, c_j)$$

donde $\delta(a,b)=1$ si $a=b$ y $\delta(a,b)=0$ en otro caso.

Tabla 2.11: Algoritmo de k vecinos más cercanos.

Para realizar la clasificación dado un punto objetivo x_q en base a los parámetros y_q , se asigna a cada ejemplo del conjunto de entrenamiento un peso dado por la inversa de la distancia entre cada uno de los punto de entrenamiento (representación de los ejemplos dados para el entrenamiento) y el punto objetivo (que representa la instancia a clasificar) de la siguiente manera:

$$w_i = \frac{1}{|x_q - x_i|} \quad (2.8)$$

Considerando las siguientes matrices: la matriz de pesos W , una matriz diagonal con las entradas w_1, \dots, w_n ; la matriz X cuyas filas son los vectores x_1, \dots, x_n , los parámetros de los ejemplos en el conjunto de entrenamiento, con la adición de un “1” en la última columna; y Y es una matriz cuyas filas son los vectores y_1, \dots, y_n , los parámetros de salida de los ejemplos en el conjunto de entrenamiento. Los datos de los pesos de entrenamiento son dados por $Z = WX$ y la función objetivo a pesar es $V = WY$. Por último se utiliza un estimador para la función objetivo, donde Z^* es el pseudoinverso de Z :

$$y_q = x_q^T Z^* V \quad (2.9)$$

2.4.5 Máquinas de Vectores de Soporte

Las *máquinas de vectores de soporte* (SVM, por sus siglas en inglés) han mostrado conseguir buen desempeño de generalización sobre una amplia variedad de problemas de clasificación, destacando recientemente en problemas de clasificación de textos (ver [40, 60, 61]), donde se aprecia que SVM tiende a minimizar el error de generalización, por ejemplo. los errores del clasificador sobre nuevas instancias. En términos geométricos, SVM puede ser visto como el intento de encontrar una superficie (σ_i) que separe a los ejemplos positivos de los negativos por el margen más amplio posible. (ver [62] para una descripción detallada del algoritmo).

La búsqueda de σ_i que cumple que la distancia mínima entre él y un ejemplo de entrenamiento sea máxima, se realiza a través de todas las superficies $\sigma_1, \sigma_2, \dots$ en el espacio $|A|$ -dimensional que separan a los ejemplos positivos de los negativos en el conjunto de entrenamiento (conocidas como superficies de decisión). Para entender mejor la idea detrás del algoritmo SVM tomaremos el caso en el que los ejemplos positivos y negativos son linealmente separables, por lo tanto las superficies de decisión son $(|A| - 1)$ -hiperplanos. Por ejemplo, en el caso de dos dimensiones varias

líneas pueden ser tomadas como superficies de decisión (ver figura 2.6), entonces el método SVM elige el elemento medio del conjunto más ancho de líneas paralelas, por ejemplo, desde el conjunto en el que la distancia máxima entre dos de sus elementos es la mayor. Cabe resaltar que la mejor superficie de decisión es determinada únicamente por un conjunto pequeño de ejemplos de entrenamiento, llamados vectores de soporte (en la figura 2.6 los cuadros distinguen los vectores de soporte).

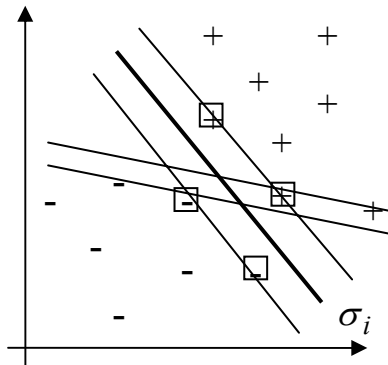


Figura 2.6: Problema de clasificación linealmente separable

Una ventaja importante de SVM es que permite construir clasificadores no lineales, por ejemplo, el algoritmo representa datos de entrenamiento no lineales en un espacio de alta dimensionalidad (llamado el espacio de características), y construye el hiperplano que tiene el margen máximo (Ver figura 2.7). Además, debido al uso de una función kernel para realizar el mapeo, es posible calcular el hiperplano sin representar explícitamente el espacio de características.

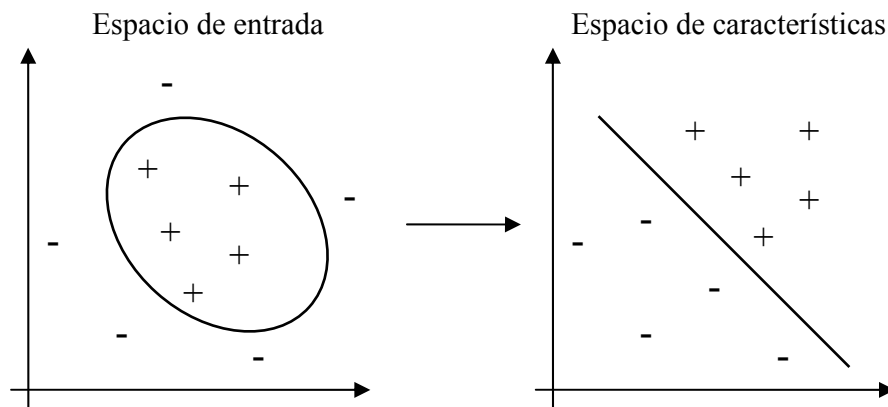


Figura 2.7: Mapeo de datos no lineales a un espacio de mayor dimensionalidad.

En el presente trabajo se utiliza el método de optimización mínima secuencial (SMO, por sus siglas en inglés) para entrenar el algoritmo SVM. En general, SMO divide la gran cantidad de problemas de programación cuadrática (QP, por sus siglas en inglés) que necesitan ser resueltos en el algoritmo SVM por una serie de problemas QP más pequeños. El lector puede consultar [65] para una descripción detallada del método, la cual se omitió porque no representa el objetivo de la tesis.

2.4.6 Naïve Bayes Localmente Ponderado (LWNB)

Naïve Bayes Localmente Ponderado (Locally Weighted Naïve Bayes, LWNB en sus siglas en inglés) es una variante de *LWR* y *k-NN*, donde un modelo de Naïve Bayes (*NB*) local es ajustado a un subconjunto de datos que están en el vecindario de la instancia a clasificar (ver [66]). El algoritmo en general es como sigue:

1. Se utiliza *k-NN* para encontrar los k vecinos más cercanos a la instancia a clasificar. Donde las instancias de entrenamiento se le asigna un peso y aquellos que están lejos de las instancias tendrán un valor bajo (cercano o igual a 0).
2. Las k instancias seleccionadas y pesadas serán utilizadas para crear el modelo de clasificación con *NB*.
3. Se realiza la clasificación de la instancia usando el modelo.

Sea d_i la distancia Euclidiana al i -ésimo vecino más cercano x_i . Se asumen que todos los atributos están normalizados con valores entre 0 y 1 antes de que las distancias sean calculadas, y los atributos nominales son binarizados. Sea f la función de peso con $f(y) = 0$ para todo $y \geq 1$. Con esta función se asignan los pesos w_i de cada instancia x_i de la siguiente manera:

$$w_i = f(d_i / d_k) \quad (2.10)$$

Esto significa que la instancia x_k (la instancia más lejana) y todas las instancias que están lejanas de la instancia a clasificar reciben un peso de cero, y todas las instancias idénticas poseen un peso de 1.

Cualquier función de disminución monotónica con la propiedad mencionada es una función candidata para pesar. En este caso, la función de pesos f_{lineal} se define de la siguiente manera:

$$f_{lineal}(y) = 1 - y \quad \text{para } y \in [0,1], \quad (2.11)$$

de esta forma, los pesos se disminuyen linealmente con la distancia.

Muchas veces ocurre el problema de la frecuencia-cero, donde la mayoría de los pesos asignados a los vecinos es 0 debido a la función 2.10. Para evitar este problema, se ha implementado en el algoritmo el estimador de Laplace para estimar la probabilidad condicional para los atributos nominales e interactúe con el esquema de los pesos. Por esta razón, se escala los pesos en base al total de los pesos de las instancias usadas para generar el modelo de NB y la suma de los pesos es aproximado a k . Asumiendo que hay r instancias de entrenamiento x_i con $d_i \leq d_k$. Entonces los pesos re-escalados w'_i es calculado de la siguiente manera, considerando que n es el número total de instancias de entrenamiento:

$$w'_i = \frac{w_i \times r}{\sum_{q=0}^n w_q}, \quad (2.12)$$

NB calcula la probabilidad posterior de la clase c_l (en este caso, la clase c es el sentido y l es el número del sentido) para una instancia que se desea clasificar con los valores de los atributos a_1, a_2, \dots, a_m como sigue:

$$p(c_l | a_1, a_2, \dots, a_m) = \frac{p(c_l) \prod_{j=1}^m p(a_j | c_l)}{\sum_{q=1}^o [p(c_q) \prod_{j=1}^m p(a_j | c_q)]}, \quad (2.13)$$

donde o es la cantidad de clases.

La probabilidad individual en el lado izquierdo de la ecuación 2.11 son estimadas basadas en el peso de los datos. La probabilidad para la clase c_l se vuelve

$$p(c_l) = \frac{1 + \sum_{i=0}^n I(c_i = c_l) w'_i}{o + \sum_{i=0}^n w'_i}, \quad (2.14)$$

donde c_i es el valor de la clase de la instancia de entrenamiento con el índice i , y la función indicador $I(x = y)$ es 1 si $x = y$ y 0 en otro caso.

Asumiendo que el atributo j es nominal, la probabilidad condicional de a_j (el valor de este atributo en la instancia a clasificar) es dada por

$$p(a_j|c_i) = \frac{1 + \sum_{i=0}^n I(a_j = a_{ij})I(c_i = c_i)w'_i}{n_j + \sum_{i=0}^n I(a_j = a_{ij})w'_i}, \quad (2.15)$$

donde n_j es el número de valores para el atributo j , y a_{ij} es el valor del atributo j en la instancia i .

2.5 Consideraciones finales

En el artículo de W. Gale [59] se establece que en general un porcentaje de 65% a 70% es un rendimiento aceptable para la desambiguación automática. Sin embargo, en el evento de SENSEVAL-3, se demostraron que los mejores sistemas WSD participantes fueron aquellos que utilizaron los métodos supervisados donde han superado el 70% de precisión, mientras que los métodos no supervisados están por debajo del 70% (ver [57] para más detalles sobre estos límites). Aunque los métodos supervisados han destacado en sus resultados, éstos todavía dependen de los datos etiquetados y de su adquisición, haciéndolos limitados en el vocabulario, es decir, sólo cubren algunas palabras que existen en el idioma (en este caso, en Inglés). Por el contrario, a pesar de tener poca precisión los métodos basados en conocimiento son los que más palabras cubren en el vocabulario gracias a la información que poseen los recursos que lo apoyan.

Por otro lado, los métodos que en su totalidad no son supervisados sólo están enfocados en la discriminación de los sentidos de la palabra (ver sección 2.2.2.2). Estos métodos agrupan las palabras objetivo por su contexto, donde cada grupo es un sentido, sin embargo, la cantidad de grupos pueden variar y no siempre coinciden en la cantidad ni en definición de los sentidos definidos en un diccionario. Por esta razón, la tarea de estos métodos no es del todo de utilidad para WSD.

Cabe mencionar que hasta ahora el uso de la Web se ha orientado a la obtención de ejemplos de uso de las palabras a desambiguar. Estos ejemplos de uso son usados posteriormente bajo un esquema supervisado. La idea del presente trabajo

va más allá al intentar utilizar la información de la Web directamente en la desambiguación. Así se propone un método basado en conocimiento utilizando la información que proveen WordNet y la Web para la desambiguación del sentido de las palabras. Para evaluar la efectividad de este método se utilizaron los 20 sustantivos del corpus SENSEVAL-3 ELS. Los beneficios que provee el método se analizan usando un esquema híbrido experimental que combina métodos supervisados y la información que proporciona el método basado en la Web. Los próximos capítulos describen ambos métodos y discuten sus resultados.

Capítulo 3

Método basado en la Distancia Normalizada de Google

En este capítulo se describe un nuevo método basado en conocimiento, donde la Web en conjunto con WordNet son utilizados como fuente de información para desambiguar directamente el sentido de una palabra. Durante este capítulo, se presenta la motivación de este método y se describe la idea detrás de este método: la distancia normalizada de Google (NGD). La parte final del capítulo describe los experimentos realizados y los resultados alcanzados con este método.

3.1 Motivación

En los últimos años, la Web ha servido como un recurso para WSD en varios trabajos, tal como se muestra en el trabajo de Gonzalo [37], en donde cada página web o documento puede ser visto como ejemplos de uso del vocabulario de un idioma. Al combinar estas páginas se obtiene un corpus masivo no etiquetado. A diferencia de trabajos previos, cuyo enfoque es usar estos ejemplos de uso como corpus de entrenamiento, nuestro objetivo es comprobar si la información de la Web puede ser usada directamente en la tarea de desambiguación al determinar si existe o no una relación semántica entre palabras. Para ello, nos basaremos en la frecuencia de co-ocurrencia de las palabras en la Web. Por ejemplo, al realizar una búsqueda con las palabras “*head*” y “*hat*”, el número de co-ocurrencias es mayor que usando las palabras “*head*” y “*banana*”. Lo cual indica una posible relación semántica entre las dos primeras palabras.

En un primer intento de utilizar la Web como una fuente de conocimiento desambiguador, se ha diseñado un método previo que consiste en la selección del

sentido más adecuado de un sustantivo considerando exclusivamente su frecuencia de co-ocurrencia con otras palabras en la Web. Este método utiliza WordNet para recuperar los sinónimos y los hipéronimos de la palabra objetivo y mide la co-ocurrencia de éstos con los sustantivos del contexto. Aquel sinónimo cuya co-ocurrencia con los sustantivos del contexto sea más alta será considerado el sustantivo más adecuado, con ello se determina el sentido más apropiado. Formalmente, el método se describe así: dado un sustantivo w , con $|w|$ sentidos, inmerso en el contexto C formado por los sustantivos presentes en su contexto inmediato (en la misma oración), la función $\Gamma(w_k, C)$ indica la cohesión temática entre el sentido k del sustantivo w y su contexto C .

La estimación de $\Gamma(w_k, C)$ en la Web considera los n sinónimos de w_k definidos en Wordnet 2.0 $\{s_{ik}, 0 < i \leq n\}$, así como sus m hiperónimos $\{h_{jk}, 0 < j \leq m\}$. La co-ocurrencia de estos elementos con el contexto se calcula mediante la función $f_s(x, y)$. Esta función retorna el número de páginas en la Web que contienen el patrón x AND y de acuerdo con el motor de búsqueda S . Asimismo, $f_s(x)$ es una función que regresa el número de páginas web que contienen la cadena x usando el buscador S .

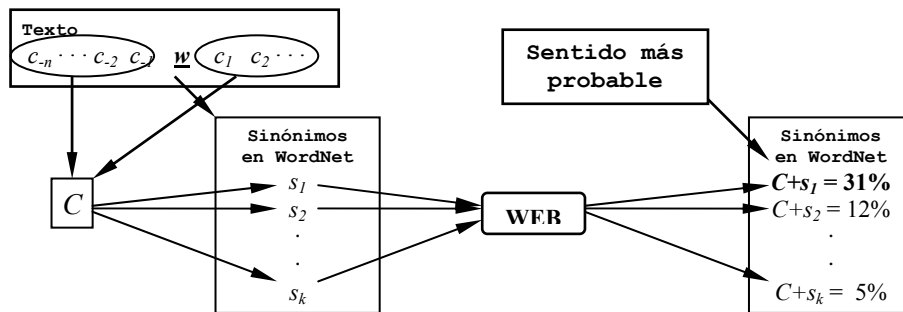


Figura 3.1: Esquema general del método previo.

Si se asume que $\Gamma(w_k, C) \approx P_{web}(w_k|C)$, es decir, que la cohesión temática entre el sentido w_k del sustantivo en cuestión y su contexto es aproximadamente proporcional a la probabilidad de, dada una página web que contiene los sustantivos pertenecientes a C , encontrar el sustantivo w con el sentido k , entonces la relación temática entre w_k y C puede calcularse de las siguientes dos maneras:

$$\Gamma(w_k, C) = \frac{1}{n+m} \left(\sum_{i=1}^n P(s_{ik}|C) + \sum_{j=1}^m P(h_{jk}|C) \right) \quad (3.1)$$

$$\Gamma(w_k, C) = \arg \max_{\substack{0 < i \leq n \\ 0 < j \leq m}} (P(s_{ik}|C), P(h_{jk}|C)) \quad (3.2)$$

donde, $P(s_{ik}|C) = f_s(C, s_{ik}) / f_s(C)$, y $P(h_{jk}|C) = f_s(C, h_{jk}) / f_s(C)$.

Asimismo, si suponemos que $\Gamma(w_k, C) \approx P_{web}(C|w_k)$, es decir, que la cohesión temática entre el sentido w_k del sustantivo en cuestión y su contexto es aproximadamente proporcional a la probabilidad de, dada una página web que contiene el sustantivo w con su sentido k , encontrar todos y cada uno de los conceptos de su contexto, entonces la relación temática entre w_k y C puede calcularse de las siguientes dos maneras:

$$\Gamma(w_k, C) = \frac{1}{n+m} \left(\sum_{i=1}^n P(C|s_{ik}) + \sum_{j=1}^m P(C|h_{jk}) \right) \quad (3.3)$$

$$\Gamma(w_k, C) = \arg \max_{\substack{0 < i \leq n \\ 0 < j \leq m}} (P(C|s_{ik}), P(C|h_{jk})) \quad (3.4)$$

donde, $P(C|s_{ik}) = f_s(C, s_{ik}) / f_s(s_{ik})$, y $P(C|h_{jk}) = f_s(C, h_{jk}) / f_s(h_{jk})$.

Las fórmulas 1 y 3 se basan en el promedio de las probabilidades. Ellas suponen que todos los sinónimos e hiperónimos del sentido w_k deben tener una relación con el contexto C para lograr distinguir una cohesión temática entre ellos. Por el contrario, las fórmulas 2 y 4 se basan en el máximo de las probabilidades, indicando que es suficiente que alguno de los sinónimos o hiperónimos de w_k tenga relación con C para establecer su cohesión temática.

El algoritmo para desambiguar un sustantivo considerando la cohesión temática con su contexto se detalla a continuación:

1. Seleccionar el conjunto C de sustantivos alrededor de w usando la oración como límite del contexto.
2. Para cada sentido w_k de w , y para cada sinónimo s_{ik} e hiperónimo h_{jk} de w_k , calcular $f_s(C, s_{ik})$, $f_s(C, h_{jk})$, $f_s(s_{ik})$ y $f_s(h_{jk})$.
3. Asignar a cada sentido w_k un peso dependiendo de la función $\Gamma(w_k, C)$, la cual combina los resultados obtenidos en el paso anterior.
4. Seleccionar el sentido (o sentidos) con los pesos mayores.

En la tabla 3.1 muestra los resultados usando las cuatro fórmulas mencionadas (precisión y recuerdo) y un conjunto de prueba con 215 sustantivos seleccionados del corpus de SENSEVAL-3 All-Words. Se ha evaluado de 3 formas: utilizando el sentido más probable (el de más peso), los dos sentidos con más probabilidades y la mitad de los sentidos más probables de una palabra (por ejemplo, cuando hay 8 sentidos para una palabra, se seleccionan los 4 más probables).

Evaluación	Fórmula	Google	Altavista	MSN
1 sentido	1	22.7%	18.6%	18.1%
	2	27.9%	18.6%	19.0%
	3	23.7%	21.5%	20.9%
	4	25.1%	21.5%	21.8%
2 sentidos	1	34.8%	35.3%	36.2%
	2	39.5%	33.9%	35.8%
	3	43.2%	38.6%	40.0%
	4	45.1%	42.7%	46.9%
⌈n/2⌉ sentidos	1	59.5%	53.4%	54.8%
	2	60.0%	57.6%	60.9%
	3	64.1%	61.3%	63.7%
	4	64.6%	62.3%	67.9%

Tabla 3.1: Porcentajes de precisión/recuerdo del experimento previo.

Los resultados en la selección de un sentido (27.9% en precisión/recuerdo) son muy bajos en comparación con el promedio obtenido en la evaluación de SENSEVAL-3 All-Words (52.2% en precisión/recuerdo). Con ello, podemos concluir que a pesar de poder extraer cierta información de la Web, ésta aún no es suficiente. De ahí se desprenden dos puntos a explorar: (i) proponer una nueva función capaz de capturar de una mejor manera la información semántica de la Web; y (ii) utilizar un nuevo esquema para relacionar el contexto de la palabra con sus diferentes sentidos. Las siguientes secciones abordan estos dos puntos.

3.2 Distancia Normalizada de Google

En esta sección se presenta una nueva función para extraer la información semántica al sustituir la frecuencia de co-ocurrencia por una medida más cercana a una relación semántica: la *Distancia Normalizada de Google* (NGD, de las siglas en inglés de *Normalized Google Distance*). Esta medida fue desarrollada por R. Cilibrasi y P. Vitanyi con el propósito de extraer información sobre la cercanía que existe entre

dos palabras (véase [16] y [17] para un estudio más detallado). La ventaja de esta medida es su capacidad de ofrecer información a nivel semántico a partir de información a nivel léxico, es decir, usando exclusivamente las palabras sin tener conocimiento sobre el significado de las mismas. Cabe mencionar que es la primera vez que se propone el uso de esta medida en la tarea de WSD. A continuación se presenta una descripción detallada de esta medida.

NGD representa la distancia relativa entre el significado de dos palabras. Entre menor sea la distancia más fuerte es la relación entre las palabras. La ecuación para obtener la distancia normalizada es como sigue:

$$NGD(x, y) = \frac{G(xy) - \text{ArgMin}\{G(x), G(y)\}}{\text{ArgMax}\{G(x), G(y)\}} \quad (3.1)$$

$$G(q) = \log \frac{M}{\text{Google}(q)}, \quad (3.2)$$

donde $q = \{x, y, xy\}$ (el conjunto de términos de búsqueda), $M = |\text{Google}|$ (el tamaño de las páginas indexadas en Google) y $\text{Google}(q)$ es el número de páginas usando los términos de búsqueda q .

En la tabla 3.2 se muestra los diferentes casos de NGD. La columna de *figuras* muestra una interpretación gráfica de cada caso. En la tercera columna se establecen las condiciones para que NGD devuelva el valor que se indica en la última columna.

Los conjuntos de x e y en esta tabla suponen que $x \subseteq M$ e $y \subseteq M$ para todos los casos. En el caso 1, cuando se realiza la búsqueda usando ambos términos xy y no se descubren co-ocurrencias se puede decir que x e y no están relacionadas y por tanto, $NGD(x,y)$ tiende a infinito. Por el contrario, en los casos 2 y 3 la distancia posee un valor mayor a 0 y menor que 1, indicando que las palabras x e y están relacionadas. El caso 4 muestra cuando x e y son totalmente iguales y $NGD(x,y) = 0$. Los casos 5 y 6, son casos que en la práctica se presentan raramente y siempre generan una distancia mayor o igual a 1. Es importante notar que en teoría, se debe considerar que $|x|, |y| \leq M$ para que la medida sea útil en la práctica. Sin embargo, hay que recordar que la Web no es un corpus completo en donde estén representadas

todas las palabras y todos los sentidos de las palabras. Así, en ocasiones se presentan irregularidades. La causa de estas irregularidades es debido a que las frecuencias de co-ocurrencias de los términos de búsqueda x , y y xy en ocasiones se aproximan demasiado a M , provocando un comportamiento irregular de la medida NGD.

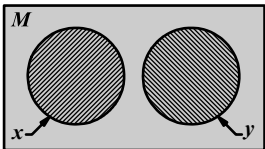
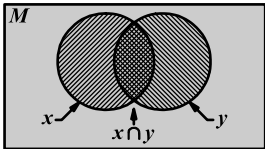
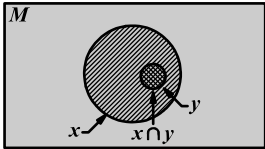
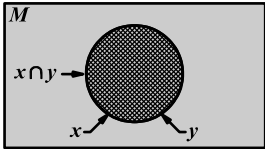
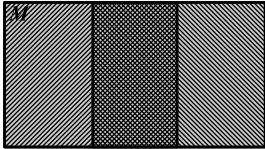
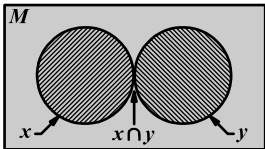
Caso	Figura	Condiciones	NGD(x,y)
1		$(x \cap y) = \emptyset$	∞
2		$(x \cap y) \neq \emptyset$	$0 < NGD(x,y) < 1$
3		$y \subset x$ $(x \cap y) \neq \emptyset$	$0 < NGD(x,y) < 1$
4		$x = y = x \cap y$	0
5		$(x + y) > M $ $(x \cap y) \neq \emptyset$	$NGD(x,y) > 1$
6		$(x \cap y) \neq \emptyset$ $G(xy) \geq 2(\max\{G(x), G(y)\})$	$NGD(x,y) \geq 1$

Tabla 3.2: Valores que devuelve NGD en ciertas condiciones.

La función NGD ha sido utilizada hasta el momento para determinar el grado de cercanía entre los significados de dos palabras. En [16] se muestran resultados de varios experimentos al tratar de distinguir el significado de varias palabras. Por ejemplo, se eligió un conjunto de palabras sobre números y colores, aplicando NGD varias veces entre las palabras se logra obtener un mapa con los valores de las

distancias de estas palabras, de esta manera el método es capaz de agrupar estas palabras por sus distancias semánticas. Al comparar los resultados obtenidos con los propuestos por expertos se llegó a un acuerdo de 87.5%.

3.3 Metodología propuesta

El método basado en la medida de *NGD* de aquí en adelante lo referiremos como el *clasificador NGD*. La función *NGD* posee dos entradas para las palabras que se desean medir y la salida es la distancia de similitud entre estas palabras. Si se repite la función de *NGD* con varias palabras con respecto a una, es posible construir un mapa (o un árbol) con los valores de las distancias. De esta manera, la idea general del método se basa en la existencia de una palabra dentro del contexto que posee una corta distancia con la palabra objetivo, y gracias a esta palabra se podrá determinar el sentido de la palabra objetivo. La manera en que se determina el sentido más adecuado es comparando las distancias entre los sinónimos de la palabra objetivo y las palabras dentro del contexto identificando aquél sinónimo cuya distancia es la más pequeña. Para determinar los sinónimos de la palabra objetivo se utilizó WordNet 2.0. En la figura 3.2 se ejemplifica esta idea usando la oración “*El gato necesita alimentos para vivir*”. Donde el conjunto de palabras del contexto contiene “*necesita*”, “*alimentos*” y “*vivir*”, y el conjunto de sinónimos contiene los sinónimos “*felino*” (1er sentido) y “*herramienta*” (2o sentido). La distancia entre “*gato*” y “*vivir*” es la más corta. Posteriormente, se extrae las distancias entre la palabra “*vivir*” y los sinónimos. De esta manera se determina que el sentido más adecuado es “*felino*” (1er sentido) dado que la distancia es más corta.

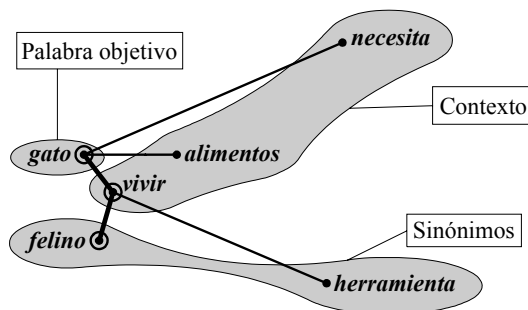


Figura 3.2: Ejemplo de desambiguación de la palabra “*gato*” con el método propuesto.

En general, el método propuesto se divide en 3 fases principales:

1. **Filtrado de las palabras de contexto.** El objetivo de esta fase es eliminar las palabras vacías y otras palabras que no son válidas (por ejemplo, números de serie, códigos o nombres raros). Las palabras vacías no son de interés para el método pues este tipo de palabras están presentes en la mayor parte de los documentos, lo que provoca que la distancia con este tipo de palabras sea muy pequeña provocando confusiones. Para esta fase se utilizó una lista existente con 180 palabras vacías (Ver apéndice 3).
2. **Selección de la palabra del contexto c_w .** El propósito de esta fase no es sólo obtener las distancias entre las palabras, también es ahorrar tiempo evitando el cálculo de las relaciones y la realización de numerosas consultas en Google cada que vez que se realiza la función de NGD para todas las combinaciones entre la palabra del contexto c_i con todos los sinónimos s_i durante la fase siguiente. Para cumplir el propósito, la función NGD es aplicada varias veces usando la palabra objetivo w y las palabras del contexto c_i , de esta manera se logra hacer un mapa con las distancias de estas palabras y se elige la palabra del contexto más cercana (denotada como c_w).
3. **Selección de los sentidos s_i más adecuados.** Esta fase consiste en agregar en una lista los sentidos más probables, es decir, todos aquellos sentidos cuyas distancias entre los sinónimos s_i (uno para cada sentido de la palabra objetivo w) y la palabra seleccionada del contexto c_w esté por debajo de un umbral. Este umbral es definido *a priori*. El propósito de este umbral es determinar la distancia máxima que puede poseer la relación entre s_i y c_w para que el sentido i sea agregado a la lista de sentidos propuestos. Es importante notar que esta lista puede llegar a estar vacía si se utiliza un umbral muy bajo (por ejemplo 0.1).

A continuación, en la tabla 3.3 se muestra el algoritmo propuesto a detalle:

-
1. Se obtienen las palabras del contexto de la palabra a desambiguar w :

$$c_{-n} \quad \dots \quad c_{-2} \quad c_{-1} \quad w \quad c_1 \quad c_2 \quad \dots \quad c_n$$
 2. Las palabras del contexto $\{c_{-n} \dots c_n\}$ pasan por un proceso de filtrado. El proceso consiste en la eliminación de las palabras en la siguiente orden:
 - a. Se eliminan las palabras vacías.
 - b. Se eliminan las palabras que son desconocidas por WordNet.
 3. Se calcula $NGD(c_i, w)$ para cada uno de las palabras del contexto c_i con la palabra a desambiguar w .
 4. Se elige la palabra del contexto c_k donde el valor de $NGD(c_i, w)$ es la distancia más corta de todas las distancias que hay entre las palabras del contexto c_i y w .
 5. Se extrae una lista de los sinónimos de cada sentido s_i con WordNet, y se calcula $NGD(c_w, s_i)$ para cada sinónimo s_i con la palabra del contexto c_w .
 6. Se eligen los sentidos s_i cuyo valor de NGD sea el menor que el umbral u previamente definido por el usuario, y son agregados a la lista de sentidos propuestos l .
-

Tabla 3.3: Algoritmo para el clasificador NGD.

A modo de ejemplo, en la figura 3.3 muestra cómo actúa el algoritmo a través de las fases usando un umbral de 0.6. Se puede apreciar que las palabras innecesarias se han eliminado en el cuadro *palabras filtradas*, una vez filtrado se extrae la distancia de cada una de las palabras del contexto filtradas c_i en relación con la palabra objetivo w . En la figura se muestra la tabla para la selección de c_w donde la palabra resultantes es “*photograph*”. De la misma manera, se procede a calcular las distancias entre el sinónimo s_i de cada sentido i y la palabra del contexto c_w . Los sinónimos son extraídos de WordNet 2.0 automáticamente. Nuevamente se crea una tabla para la selección de los sentidos, en esta tabla se muestra las distancias generadas por $NGD(c_w, s_i)$ y se eligen los sentidos 1 y 2 porque sus distancias son más cortas que el umbral definido (en este ejemplo, con un valor de 0.6).

3.4 Experimentos

Para evaluar el algoritmo se realizaron diferentes experimentos bajo las siguientes consideraciones:

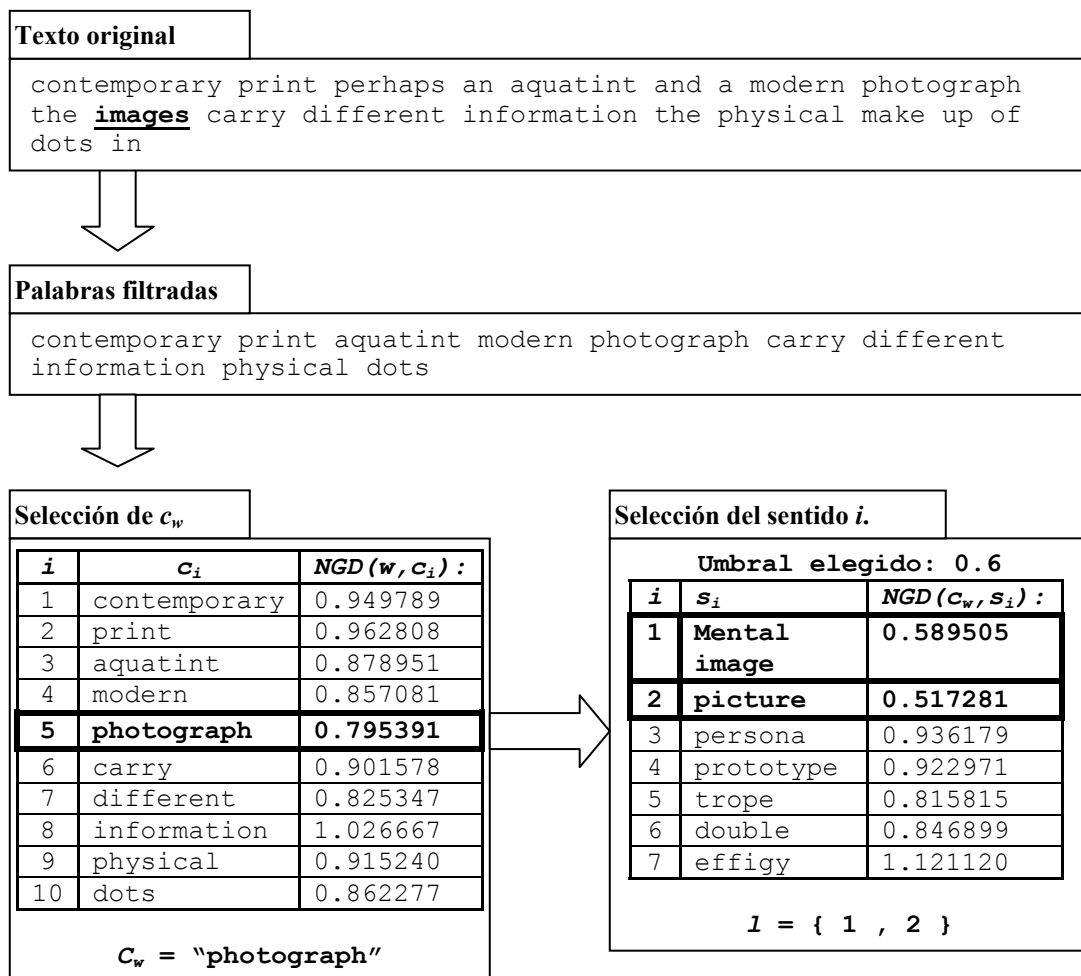


Figura 3.3: Ejemplo del algoritmo propuesto

- Los conjuntos de datos fueron construidos a partir del corpus de prueba de SENSEVAL-3 ELS. Se creó un conjunto para cada uno de los sustantivos (20 en total), con un promedio de 99 ejemplos y 5.85 sentidos por sustantivos (ver tabla 3.4, donde se muestran el inventario de las palabras) para obtener resultados individuales. Para los ejemplos del conjunto de datos, se extrajeron 20 palabras del contexto original para cada ejemplo (10 palabras de cada lado). Teóricamente es posible aumentar el tamaño del contexto, sin embargo, en la práctica el tiempo de respuesta se incrementa considerablemente dado el número de consultas a Google. Hay que recordar que cada vez que es utilizada la función de NGD se necesitan

realizar 3 consultas a la Web. En promedio cada consulta tardó 5 segundos en este experimento.

Palabra	Ejemplos	Sentidos
argument	129	5
arm	135	6
atmosphere	102	6
audience	107	4
bank	138	10
degree	140	7
difference	130	5
difficulty	23	4
disc	109	4
image	75	7
interest	95	7
judgment	32	7
organization	57	7
paper	136	7
party	125	5
performance	91	5
plan	99	3
shelter	115	5
sort	109	4
source	35	9
Total	1982	117
Promedio	99.1	5.85

Tabla 3.4: Palabras en inglés que fueron utilizadas para este experimento.

- En base a los resultados reportados en [16] y [17], los valores que se establecieron para el umbral en el presente experimento se encuentran en un rango entre 0.1 al 1, donde los valores representan la distancia máxima que se desea que haya entre dos palabras (en este caso, la distancia entre s_i y c_w). El propósito de estos diferentes valores es analizar el comportamiento resultante del algoritmo.
- Para comparar los resultados obtenidos, se tomó como referencia (*baseline*) los resultados del método conocido como *Sentido Más Frecuente (MFS)*, de sus siglas en inglés de *Most Frequent Sense*, ver [12] para más detalles). Este método es comúnmente usado en las evaluaciones de los sistemas de WSD. El MFS aprovecha un fenómeno propio del lenguaje: una palabra polisémica, a pesar de tener varias interpretaciones, siempre tendrá un sentido más comúnmente usado. Debido a que los

recursos lingüísticos como WordNet se construyen respetando este fenómeno siempre el primer sentido enlistado para una palabra polisémica será el más común. De esta manera, el MFS nos da un valor de referencia (*baseline*) muy adecuado. El valor promedio de MFS que se obtuvo fue de 49% en precisión y recuerdo usando los conjuntos mencionados sobre WordNet.

Para la evaluación, se utilizaron las siguientes métricas: *precisión*, *recuerdo* y *cobertura*. En la tabla 3.5 se describen estas tres métricas, donde T es la cantidad total de ejemplos del conjunto, V es el número de ejemplos correctamente clasificados y U es el número de ejemplos clasificados cuya lista de sentidos probables no está vacía (sin importar si están correctamente clasificados o no). Un ejemplo se considera correcto cuando al menos uno de los sentidos enlistados es correcto.

Métricas	Ecuación
Cobertura (C)	$C = \frac{U}{T}$
Recuerdo (R)	$R = \frac{V}{T}$
Precisión (P)	$P = \frac{V}{U}$

Tabla 3.5: Descripción de las métricas utilizadas para evaluar el algoritmo.

Al aplicar el método propuesto y considerando como respuesta un sólo sentido¹ se obtuvo un resultado del 23.5% en precisión y recuerdo, con una cobertura del 100%. Este resultado es pobre y aparentemente con ello se puede concluir que el uso de la información extraída de la Web no es adecuada para seleccionar el sentido más probable. Sin embargo, al modificar el esquema y considerar un umbral de cercanía se pudo analizar a detalle el comportamiento del método propuesto. De esta manera, cuando el método propone más de un sentido (los sentidos más cercanos), se observó que el método devuelve una lista con los sentidos más probables. A continuación, se muestran en las tablas 3.6 y 3.7 se muestran los porcentajes

¹ Esto se logra con una modificación sencilla al algoritmo: sólo se escoge aquél sinónimo s_i cuya distancia a la palabra del contexto seleccionada c_k sea la más cercana.

resultantes con las métricas mencionadas utilizando umbrales de 0.1 al 0.5 y de 0.6 al 1 respectivamente. El experimento duró aproximadamente 70 horas consecutivas.

Palabra	Umbral														
	0.1			0.2			0.3			0.4			0.5		
	P	R	C	P	R	C	P	R	C	P	R	C	P	R	C
arm	81.5	81.5	100.0	82.2	82.2	100.0	82.2	82.2	100.0	82.2	82.2	100.0	82.2	82.2	100.0
atmosphere	2.0	2.0	100.0	2.0	2.0	100.0	4.9	4.9	100.0	5.9	5.9	100.0	5.9	5.9	100.0
audience	93.5	93.5	100.0	93.5	93.5	100.0	93.5	93.5	100.0	93.5	93.5	100.0	93.5	93.5	100.0
bank	21.7	21.7	100.0	22.5	22.5	100.0	23.9	23.9	100.0	23.9	23.9	100.0	23.9	23.9	100.0
difference	57.7	57.7	100.0	57.7	57.7	100.0	57.7	57.7	100.0	57.7	57.7	100.0	57.7	57.7	100.0
image	100.0	2.7	2.7	40.0	2.7	6.7	100.0	6.7	6.7	100.0	8.0	8.0	100.0	9.3	9.3
judgment	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	66.7	6.3	9.4
organization	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.8	50.0	1.8	3.5	33.3	1.8	5.3
plan	50.0	1.0	2.0	28.6	2.0	7.1	87.5	7.1	8.1	100.0	8.1	8.1	100.0	8.1	8.1
source	11.4	11.4	100.0	14.3	14.3	100.0	14.3	14.3	100.0	20.0	20.0	100.0	34.3	34.3	100.0
degree	6.4	6.4	100.0	8.6	8.6	100.0	10.0	10.0	100.0	10.7	10.7	100.0	10.7	10.7	100.0
difficulty	69.6	69.6	100.0	69.6	69.6	100.0	69.6	69.6	100.0	69.6	69.6	100.0	78.3	78.3	100.0
disc	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	3.7	3.7
interest	3.2	3.2	100.0	3.2	3.2	100.0	3.2	3.2	100.0	3.2	3.2	100.0	9.5	9.5	100.0
paper	55.9	55.9	100.0	57.4	57.4	100.0	57.4	57.4	100.0	58.1	58.1	100.0	59.6	59.6	100.0
performance	30.8	30.8	100.0	31.9	31.9	100.0	31.9	31.9	100.0	31.9	31.9	100.0	33.0	33.0	100.0
shelter	68.7	68.7	100.0	68.7	68.7	100.0	68.7	68.7	100.0	68.7	68.7	100.0	68.7	68.7	100.0
sort	39.4	39.4	100.0	39.4	39.4	100.0	39.4	39.4	100.0	40.4	40.4	100.0	46.8	46.8	100.0
argument	4.7	4.7	100.0	7.8	7.8	100.0	10.1	10.1	100.0	10.1	10.1	100.0	10.9	10.9	100.0
party	31.2	31.2	100.0	31.2	31.2	100.0	32.0	32.0	100.0	32.0	32.0	100.0	35.2	35.2	100.0
Promedio	36.4	29.1	75.2	32.9	29.7	75.7	39.3	30.6	75.8	42.9	31.3	76.0	52.5	34.0	76.8

Tabla 3.6: Porcentajes de los resultados para cada palabra con umbrales de 0.1 al 0.5.

Palabra	Umbral														
	0.6			0.7			0.8			0.9			1		
	P	R	C	P	R	C	P	R	C	P	R	C	P	R	C
arm	88.1	88.1	100.0	95.6	95.6	100.0	97.8	97.8	100.0	98.5	98.5	100.0	98.5	98.5	100.0
atmosphere	10.8	10.8	100.0	34.3	34.3	100.0	64.7	64.7	100.0	85.3	85.3	100.0	95.1	95.1	100.0
audience	96.3	96.3	100.0	99.1	99.1	100.0	99.1	99.1	100.0	99.1	99.1	100.0	99.1	99.1	100.0
bank	23.9	23.9	100.0	28.3	28.3	100.0	52.2	52.2	100.0	92.8	92.8	100.0	99.3	99.3	100.0
difference	71.5	71.5	100.0	93.8	93.8	100.0	97.7	97.7	100.0	98.5	98.5	100.0	98.5	98.5	100.0
image	45.7	21.3	46.7	41.4	38.7	93.3	68.0	68.0	100.0	89.3	89.3	100.0	92.0	92.0	100.0
judgment	81.8	56.3	68.8	93.8	93.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
organization	17.6	5.3	29.8	40.0	24.6	61.4	74.5	71.9	96.5	98.2	98.2	100.0	100.0	100.0	100.0
plan	100.0	8.1	8.1	58.8	10.1	17.2	29.9	26.3	87.9	85.9	85.9	100.0	98.0	98.0	100.0
source	48.6	48.6	100.0	71.4	71.4	100.0	85.7	85.7	100.0	97.1	97.1	100.0	97.1	97.1	100.0
degree	27.1	27.1	100.0	75.0	75.0	100.0	80.0	80.0	100.0	91.4	91.4	100.0	99.3	99.3	100.0
difficulty	82.6	82.6	100.0	87.0	87.0	100.0	91.3	91.3	100.0	91.3	91.3	100.0	95.7	95.7	100.0
disc	81.8	8.3	10.1	68.1	29.4	43.1	80.2	59.6	74.3	97.2	95.4	98.2	98.1	96.3	98.2
interest	26.3	26.3	100.0	65.3	65.3	100.0	78.9	78.9	100.0	94.7	94.7	100.0	100.0	100.0	100.0
paper	61.0	61.0	100.0	71.3	71.3	100.0	91.2	91.2	100.0	96.3	96.3	100.0	96.3	96.3	100.0
performance	42.9	42.9	100.0	67.0	67.0	100.0	78.0	78.0	100.0	96.7	96.7	100.0	98.9	98.9	100.0
shelter	68.7	68.7	100.0	71.3	71.3	100.0	87.0	87.0	100.0	96.5	96.5	100.0	97.4	97.4	100.0
sort	69.7	69.7	100.0	90.8	90.8	100.0	94.5	94.5	100.0	98.2	98.2	100.0	98.2	98.2	100.0
argument	18.6	18.6	100.0	38.0	38.0	100.0	54.3	54.3	100.0	89.1	89.1	100.0	93.8	93.8	100.0
party	45.6	45.6	100.0	53.6	53.6	100.0	85.6	85.6	100.0	90.4	90.4	100.0	92.8	92.8	100.0
Promedio	55.4	44.0	83.2	67.2	61.9	90.8	79.5	78.2	97.9	94.3	94.2	99.9	97.4	97.3	99.9

Tabla 3.7: Porcentajes de los resultados para cada palabra con umbrales de 0.6 al 1.0.

En la figura 3.4 se resumen los resultados obtenidos mostrando los promedios para cada umbral. Cabe recordar que el promedio general de sentidos por palabra en el conjunto de prueba es de 5.85 sentidos. Como puede apreciarse con un umbral de 0.7 se obtienen mejores resultados que el método MFS de referencia. En la gráfica de la figura 3.5, se puede apreciar el crecimiento gradual del tamaño promedio de la lista de sentidos (TPL) al incrementar el umbral. Por supuesto, mientras más grande la lista de sentidos la precisión como el recuerdo aumenta. Se puede observar el crecimiento del TPL es casi directamente proporcional al recuerdo.

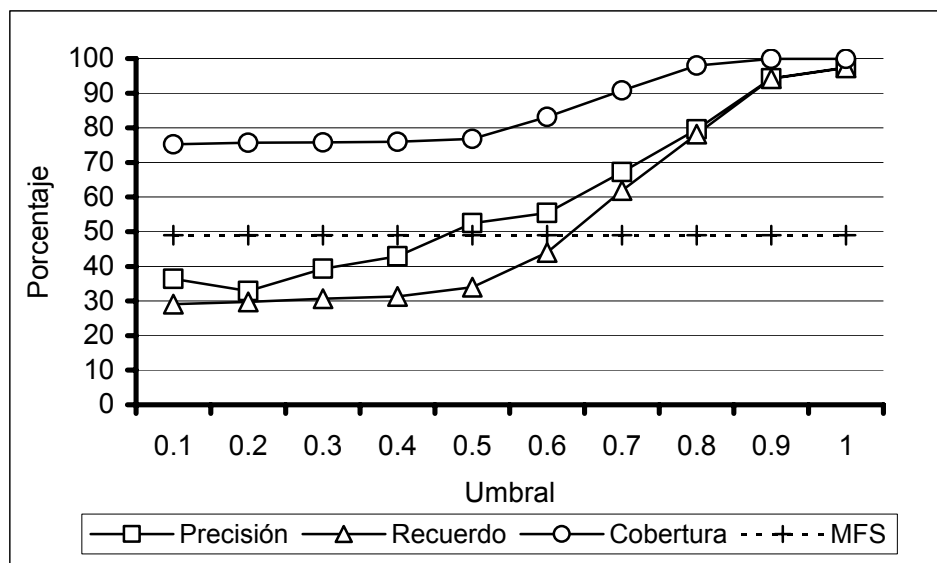


Figura 3.4: Gráfica de Precisión/Recuerdo/Cobertura vs Umbral.

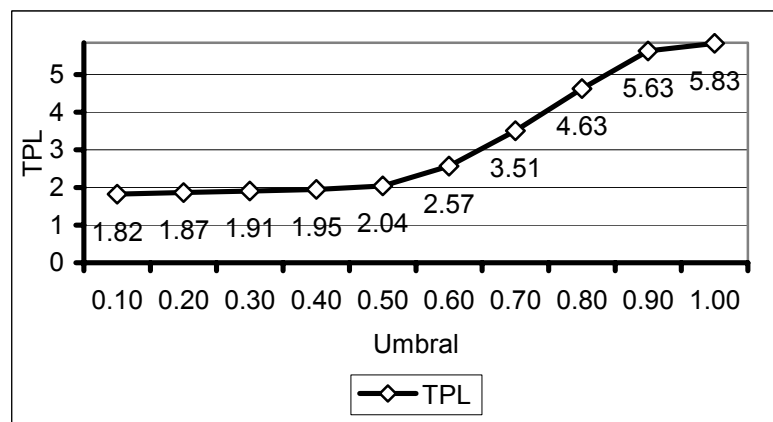


Figura 3.5: Gráfica de TPL vs Umbral.

3.5 Discusión de los resultados

El objetivo de este experimento consistió en comprobar la utilidad de la Web al usarla directamente en la tarea de WSD. Desafortunadamente, el primer experimento no fue exitoso. Este primer método, forzado a devolver un sólo sentido, obtuvo una precisión y recuerdo muy bajos (23.5%). Sin embargo, al realizar un análisis más detallado considerando un umbral de cercanía, se observó que es posible aprovechar la información de la Web. Los resultados de estos experimentos posteriores, demuestran que utilizando umbrales altos, se obtienen listas grandes incluyendo la gran mayoría de los sentidos posibles. A pesar de resolver el problema se manera parcial, se rechazan con gran acierto los sentidos más improbables. Así la Web puede ser usada para determinar los sentidos más probables, aquellos cuya distancia no sobrepasan el umbral, de aquellos sentidos más improbables, aquellos que sobrepasan el umbral.

Al verlo desde este punto de vista, el método ha demostrado ser mejor cuando se trata de rechazar los sentidos menos probables. Los resultados muestran que al utilizar umbrales muy altos se descartaban pocos sentidos. En el caso de un umbral = 0.8, el tamaño promedio de la lista fue de 4.63 sentidos, lo que quiere decir que se han rechazados 1.2 sentidos en promedio de 5.85. Bajo esta idea, surge la motivación de aprovechar esta ventaja combinándolo con otros clasificadores. El siguiente capítulo describe cómo se aprovecha la información que el método ha podido extraer de la Web bajo distintos enfoques propuestos y poder analizar el beneficio adquirido.

Capítulo 4

Combinando NGD con un enfoque supervisado

Tradicionalmente para solucionar el problema de WSD se utilizan esquemas híbridos combinando diferentes métodos y diferentes recursos. En el presente capítulo se presenta un esquema para integrar la información extraída de la Web con un enfoque supervisado. El propósito es demostrar que el conocimiento proveniente de la Web puede apoyar el proceso de clasificación (utilizando métodos supervisados) y mejorar sus resultados. La primera sección de este capítulo presenta un ensamble de clasificadores para la tarea de WSD. Los resultados de este primer ensamble serán utilizados como *baseline*. Posteriormente, se presentan dos esquemas integrando el clasificador NGD dentro de este enfoque supervisado. Finalmente se describen los experimentos y los resultados obtenidos.

4.1 Un ensamble para WSD

Un ensamble, es un conjunto de clasificadores que determinan la clase de una instancia por votación. En general, un ensamble posee dos fases: la fase de clasificación y la fase de votación. En la primera fase, cada clasificador propone un sentido para la palabra a desambiguar. En la siguiente fase, se contabilizan las coincidencias de las respuestas de cada clasificador y se elige el sentido cuya cantidad de votos es mayor. Así se tendrá un ensamble para cada palabra a desambiguar. Antes de integrar el clasificador NGD, se realizaron una serie de experimentos para tener resultados de referencia y así comprobar el impacto del clasificador NGD. A continuación se describen las consideraciones para estos experimentos:

- Se utilizaron los conjuntos de 19 sustantivos de SENSEVAL-3 para entrenamiento y prueba. Se extrajeron los ejemplos con sus respectivas palabras del contexto (10 palabras en ambos lados). Cada palabra posee un conjunto de entrenamiento y prueba. En promedio, para cada palabra el conjunto posee para entrenamiento 207.1 ejemplos y para prueba 102.47 ejemplos. En general, las palabras poseen en promedio 5.68 sentidos. En la tabla 4.1 se muestra el inventario utilizado para la realización de este experimento.

Palabra	Ejemplos (entrenamiento)	Ejemplos (prueba)	Sentidos
argument	278	129	5
arm	269	135	6
atmosphere	199	102	6
audience	229	107	4
bank	273	138	10
degree	287	140	7
difference	253	130	5
difficulty	50	23	4
disc	227	109	4
image	148	75	7
interest	186	95	7
judgment	62	32	7
organization	114	57	7
paper	284	136	7
party	247	125	5
performance	188	91	5
plan	185	99	3
shelter	226	115	5
sort	230	109	4
Total	3935	1947	108
Promedio	207.10	102.47	5.68

Tabla 4.1: Palabras en inglés que fueron utilizadas para este experimento.

- Los clasificadores que integran el ensamble son de distinto tipo: basados en Instancias (K-nn, LWR y LWNB), basados en probabilidades (Naïve Bayes y LWNB), árboles de decisiones (C4.5) y por último, las máquinas de vectores de soporte (o en sus siglas en inglés SVM). Los conjuntos de datos para el entrenamiento de los clasificadores fueron transformados a una representación binaria. Esta conversión consiste en crear un vocabulario de las palabras que están presentes en el conjunto de

entrenamiento donde serán representadas como atributos, y el valor sólo serán 0 y 1 según estén presentes o ausentes. El valor de la etiqueta de cada ejemplo para el entrenamiento es su sentido.

- Para analizar el desempeño de los enfoques, se crearon varios ensambles con diferentes combinaciones de clasificadores y cantidades sin utilizar el método propuesto. De esta manera se logra crear un *baseline* (punto de referencia). El criterio para la selección de los clasificadores se basa en un trabajo previo donde se analizó el desempeño de cada algoritmo. Un breve resumen y las observaciones de este trabajo se incluyen en el apéndice 1.

Los resultados obtenidos con estos ensambles se muestran en la tabla 4.2. Por razones de claridad sólo se muestran los promedios de los ensambles (hay que recordar que se tiene un ensamble por cada palabra). Como puede verse en la tabla se experimentó variando la cantidad de clasificadores por ensamble. Desde el caso más simple con 2 clasificadores hasta incluir los 6 clasificadores. Es importante mencionar que la cobertura es de 100%, ya que bajo este enfoque siempre se propone un sentido dada una palabra.

Num. Clas.	Clasificadores	Prec.
2	NB, LWNB	47.4768
3	NB, LWNB, SVM	47.5442
4	NB, LWNB, SVM, LWL	46.7406
5	NB, LWNB, SVM, LWL, K-nn	46.8578
6	NB, LWNB, SVM, LWL, K-nn, J48	45.1908

Tabla 4.2: Porcentajes promedios de la precisión y clasificadores usados.

4.2 Integrando el clasificador NGD

En esta sección se proponen dos enfoques distintos para integrar el clasificador NGD bajo el enfoque supervisado, bajo el esquema de ensambles antes descrito. En un primer esquema el clasificador NGD se integra como un clasificador más dentro del ensamble, es decir, su aporte será considerado durante la fase de clasificación del ensamble. En un segundo esquema, el clasificador NGD se integra como un filtro antes de la fase de votación en el ensamble, es decir, el clasificador NGD aprueba o

censura los votos de los clasificadores. En las siguientes secciones se detallan los enfoques mencionados y posteriormente se analizan los resultados obtenidos.

4.2.1 Primer esquema: NGD en la fase de clasificación

Como se ha mencionado anteriormente, un ensamble es un conjunto de clasificadores en donde cada uno propone un sentido para la palabra objetivo. En este enfoque se pretende utilizar el clasificador NGD propuesto interviniendo como un clasificador más dentro del ensamble y aprovechar el uso de la lista de sentidos propuestos dentro de la fase de clasificación. Cada uno de los sentidos propuestos en la lista se cuenta como un voto. La salida de este ensamble es una lista, esto se debe a que es posible que se presenten empates, así la lista de respuesta contiene los sentidos con la misma probabilidad. En la figura 4.1 se puede apreciar este primer esquema. Al igual que los experimentos realizados en el capítulo anterior, se utilizaron diferentes umbrales para el clasificador NGD. Estos valores están en un rango de entre 0.1 y 1. En la tabla 4.3 se aprecian los resultados del ensamble utilizando distintos valores para el umbral y para las diferentes configuraciones de los ensambles.

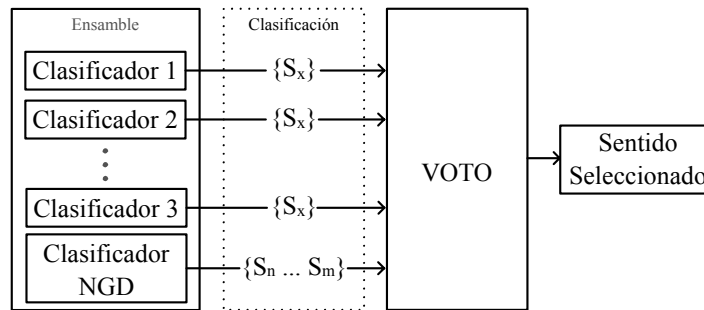


Figura 4.1: Estructura del ensamble del primer enfoque.

Como puede observarse aun con un umbral bajo, en todos los casos se supera el *baseline* (el ensamble antes de integrar el clasificador NGD). En la tabla 4.3, los mejores resultados están marcados en negritas. Además, se puede observar que el máximo TPL que se ha obtenido bajo este enfoque es de 1.21 sentidos. Esto quiere decir, que la mayor parte de las votaciones finales proponen un solo sentido. En general, se puede observar la integración del clasificador NGD ayudó a mejorar la

precisión. Para ver resultados más detallados, véase las tablas del apéndice 3. Es importante aclarar que en este enfoque, todas las instancias poseen por lo menos un sentido propuesto, por lo que la cobertura es del 100%.

Umbral	Número de clasificadores									
	2		3		4		5		6	
	P	TPL	P	TPL	P	TPL	P	TPL	P	TPL
0.1	51.30	1.21	48.93	1.07	50.47	1.12	49.21	1.07	48.63	1.08
0.2	51.30	1.21	48.93	1.07	50.47	1.12	49.16	1.07	48.63	1.08
0.3	51.30	1.21	48.93	1.07	50.51	1.12	49.16	1.07	48.63	1.08
0.4	51.25	1.21	48.93	1.07	50.51	1.12	49.16	1.07	48.63	1.08
0.5	51.51	1.20	49.05	1.07	50.51	1.12	49.17	1.07	48.63	1.08
0.6	51.16	1.19	48.98	1.07	50.72	1.13	48.63	1.06	48.71	1.09
0.7	50.88	1.15	49.39	1.08	51.19	1.13	48.84	1.07	48.88	1.09
0.8	50.82	1.12	49.12	1.07	50.93	1.13	48.93	1.07	48.67	1.09
0.9	51.72	1.14	48.85	1.05	51.43	1.14	49.05	1.08	49.04	1.10
1.0	51.80	1.14	48.59	1.05	51.51	1.14	48.94	1.07	49.07	1.10
Baseline	47.48	1.00	47.54	1.00	46.74	1.00	46.86	1.00	45.19	1.00

Tabla 4.3: Tabla de los porcentajes de precisión (P) y el tamaño promedio de la lista de sentidos propuestos (TPL) obtenidos usando el primer enfoque.

4.2.2 Segundo esquema: NGD en la fase de votación

A diferencia del enfoque anterior, el siguiente experimento muestra cómo se comporta el ensamble cuando el clasificador NGD interviene en la fase de la votación. En el ensamble normal, es común que los votos sean libres y un clasificador pueda votar por un sentido que no es probable dada una palabra, aún cuando los otros clasificadores votan por otros. La idea general de este enfoque puede ser vista como un proceso de intersección de 2 conjuntos: los votos propuestos por los clasificadores y la lista de los sentidos propuestos por NGD. La contabilidad sólo se realiza si los sentidos votados se encuentran en la lista de los sentidos propuestos. Como se mencionó en el capítulo anterior el clasificador NGD tiene un excelente desempeño eliminando sentidos improbables. Bajo este esquema se espera evitar votos por los sentidos menos probables y sólo permitir el conteo de los votos en donde el sentido sea probable. En la figura 4.2 se muestra el esquema en general de este enfoque. En este experimento, también se cuenta con la característica de devolver una lista de posibles sentidos cuando existe un empate en el voto. Además es posible que el ensamble no pueda proponer un sentido, es decir, devuelva una lista de sentidos

vacía. Esto se debe a que a veces los sentidos votados no se encuentran en la lista propuesta por el clasificador NGD. Por esta razón, en este enfoque las métricas de recuerdo y cobertura (no necesariamente es el 100%) difieren del esquema anterior.

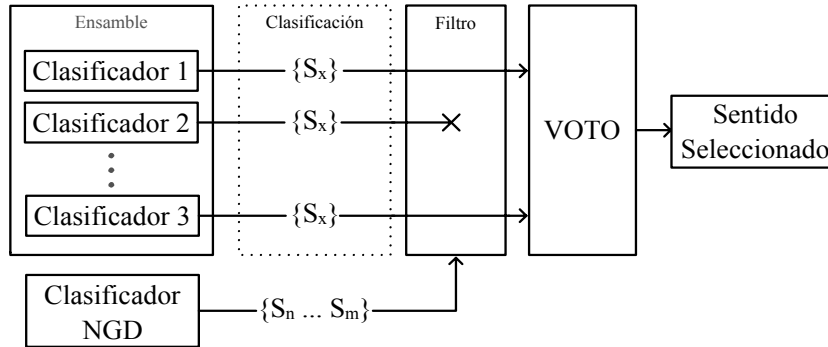


Figura 4.2: Esquema del ensamble usando el clasificador NGD para filtrar los votos.

En la tabla 4.4, se muestran los resultados obtenidos para este enfoque utilizando diferentes valores del umbral para el clasificador NGD. Se puede apreciar el crecimiento de los porcentajes de las métricas a medida que aumenta el umbral. En el apéndice 4 se muestran tablas que detallan estos resultados.

# Cl.	Métrica	Umbral										Base-line
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	
2	R	16.50	16.71	17.24	17.51	19.00	23.86	32.06	41.11	49.57	51.46	47.48
	P	40.04	38.75	43.29	47.60	53.28	51.52	51.53	52.18	53.81	53.81	47.48
	C	31.62	32.36	33.06	33.54	36.07	47.07	62.68	77.42	91.14	94.93	100.00
	TPL	0.81	0.81	0.84	0.93	1.10	1.07	1.08	1.09	1.11	1.12	1.00
3	R	16.22	16.43	16.96	17.23	18.63	22.94	30.88	38.96	46.84	48.20	47.54
	P	38.82	37.52	42.06	46.37	51.85	48.94	49.12	48.96	50.63	50.40	47.54
	C	33.84	34.58	35.28	35.76	38.46	48.98	64.55	78.93	91.86	95.25	100.00
	TPL	0.74	0.74	0.79	0.90	1.00	1.00	1.01	1.01	1.01	1.01	1.00
4	R	17.71	18.00	18.57	18.84	20.29	25.74	33.61	41.70	49.92	51.49	46.74
	P	38.95	39.05	44.25	48.56	53.35	52.55	51.10	51.63	53.84	53.71	46.74
	C	38.00	38.60	39.31	39.78	42.54	52.82	67.03	79.73	91.95	95.28	100.00
	TPL	0.81	0.84	0.90	1.01	1.06	1.08	1.08	1.10	1.12	1.12	1.00
5	R	17.18	17.37	17.94	18.25	19.56	24.41	31.89	39.71	48.08	49.39	46.86
	P	27.54	27.35	32.76	38.55	41.15	44.41	44.69	48.07	50.46	50.03	46.86
	C	46.04	46.46	47.17	47.63	50.24	58.55	71.30	81.76	94.83	98.43	100.00
	TPL	0.85	0.85	0.91	0.96	1.01	1.02	1.03	1.05	1.05	1.05	1.00
6	R	17.41	17.63	18.20	18.51	19.82	24.94	32.62	40.09	48.17	49.64	45.19
	P	27.28	27.21	32.51	38.29	40.98	45.19	45.54	48.18	50.41	50.19	45.19
	C	46.39	46.82	47.52	47.99	50.75	58.78	71.57	81.95	95.00	98.58	100.00
	TPL	0.85	0.86	0.91	0.97	1.02	1.03	1.05	1.06	1.07	1.07	1.00

Tabla 4.4: Tabla de porcentajes de precisión (P), recuerdo (R), cobertura (c) y el tamaño promedio de la lista de sentidos propuestos (“TPL”) obtenidos usando diferentes umbral para el segundo enfoque.

Los bajos valores de los porcentajes son debido a que el clasificador NGD posee una lista muy pequeña o en ocasiones la lista está vacía cuando el umbral es muy bajo. Por el contrario, al utilizar umbrales muy altos, el ensamble supera el *baseline* en un 10.4% en promedio en la precisión con una cobertura de 94%. Estos umbrales altos (0.9 y 1) hacen que NGD permita rechazar los sentidos más improbables para una palabra objetivo dentro de la lista y tener el mayor número de sentidos aceptables para determinar la votación en el ensamble.

4.3 Discusiones

Los experimentos realizados en este capítulo, demuestran que el clasificador NGD puede aportar beneficios en ambos enfoques: reforzar los votos dentro de un ensamble como un clasificador más, o bien, rechazar los sentidos menos probables filtrando los votos. En otras palabras, se puede afirmar que la información de la Web puede ser aprovechada en la tarea de WSD bajo un esquema híbrido.

La diferencia entre los enfoques consiste en dónde interviene el clasificador NGD en el proceso de clasificación dentro de los ensambles: en la fase de clasificación o en la fase de votación. Las métricas indican que en el mejor de los casos, ambos enfoques son comparables. En ambos casos superan el *baseline* hasta en un 15% en precisión.

En términos de cobertura, el primer enfoque es mejor, ya que las listas que generan nunca estarán vacías y siempre propondrán un sentido. Mientras que el segundo supera al primero con ligeras diferencias en precisión y recuerdo.

Una propuesta para mejorar estos enfoques es realizar una intervención en la clasificación individual de los clasificadores. De esta manera, se espera que la precisión y recuerdo sea mayor que el clasificador normal. Sin embargo, esto implica un estudio más detallado sobre los algoritmos de aprendizaje para implementar esta intervención.

Capítulo 5

Conclusiones y Trabajo Futuro

La desambiguación del sentido de las palabras (WSD) no es una tarea fácil dentro del PLN. Muchos sistemas han intentado seleccionar el sentido más apropiado para una palabra polisémica usando técnicas estadísticas y/o aprendizaje automático. A diferencia de estos sistemas, el método propuesto es capaz de aprovechar el enorme potencial de la Web, al considerarlo un enorme corpus no etiquetado. Para aprovechar esta información dentro de la tarea de WSD, se utilizó la función NGD para extraer distancias de relación semántica entre dos palabras. Esta función se basa en información estadística proporcionada por Google. Así, al no depender de un corpus, la posibilidad de desambiguar un amplio vocabulario es posible con este método. El primer intento de resolver la tarea de WSD con el método propuesto, forzando a seleccionar un sólo sentido, no dió resultados satisfactorios (23.5%). Esto quiere decir que el método no sirve para desambiguar el sentido de una palabra. Sin embargo, el objetivo de la tesis no es proponer un método de desambiguación directa sino un método que apoye la tarea de WSD usando la Web como un recurso. Esto se aprecia en los experimentos detallados después del primer intento, al usar un umbral de cercanía. Las observaciones, al usar diferentes valores para este umbral, indican que la Web puede ser usada para obtener evidencia del uso de un sentido en específico gracias a su contexto mostrando los sentidos más probables. Por supuesto, entre esta lista de sentidos más probables encontramos el sentido correcto. Como se mostró en el capítulo 3, mientras más alejado es un sentido menos evidencia encontramos en de su uso en la Web. De ahí la utilidad de la función NGD. Al usar un umbral de 0.9, el método llega a tener en promedio una precisión de 94.3%, un recuerdo de 94.2 y una cobertura de 99.9%. Con estos valores, el método propone 5.63 sentidos en promedio

de 5.83 sentidos originales. A pesar de que sólo el 0.20 de los sentidos son rechazados el método es muy confiable. Así, podemos afirmar que la información extraída de la Web utilizada de manera aislada es útil para descartar los sentidos más improbables. Bajo esta premisa, se integró el método NGD con métodos supervisados, en este caso, un ensamble de clasificadores. El propósito de esta integración es averiguar si el rechazo de los sentidos poco probables puede ayudar a la tarea de WSD. Usando un ensamble como referencia, al integrar el método bajo distintos enfoques, los experimentos demostraron que la precisión de los métodos supervisados puede mejorarse hasta un 10.4% en promedio al incorporar la información extraída de la Web.

5.1 Trabajo futuro

El presente método tiene varias limitantes y por supuesto es perfectible. A continuación se proponen algunas ideas para mejorar el método.

- El método sólo crea el mapa de distancias entre una palabra del contexto y la palabra objetivo, lo ideal es realizar un mapa más completo entre la palabra a desambiguar, incluyendo las palabras del contexto y los sinónimos de los sentidos de la palabra objetivo. De esta manera el análisis de este mapa será más completo.
- En la práctica, la extracción de la distancia semántica con la función NGD de cada par de palabras requiere de numerosas consultas en la Web, y por tanto, la espera de la respuesta desde la Web puede hacer lento los cálculos. Por lo que es necesario buscar un método o una alternativa en donde se pueda disminuir el tiempo total para los cálculos.
- Definir un nuevo esquema para permitir la intervención del clasificador NGD en el proceso de clasificación de un algoritmo de aprendizaje. Si la primera opción del clasificador no está dentro de los valores posibles reportados por NGD tomar la opción en orden descendente que sí esté reportada por NGD.

- Proponer un esquema completo que permita realizar experimentos con las otras categorías gramaticales de SENSEVAL: verbos y adjetivos.
- Experimentar el método con otros idiomas condicionado a la disponibilidad de un tesoro a la manera de WordNet y siempre y cuando su presencia en la Web sea abundante.

Apéndice 1

Comparación de clasificadores supervisados

En este apéndice se muestra un resumen de los resultados de un estudio sobre el comportamiento de los clasificadores supervisados para la tarea de WSD [64]. El objetivo de este estudio fue identificar los mejores clasificadores para esta tarea y cómo se comportan con distintos tamaños de contexto.

La tabla A1.1 se muestra las palabras utilizadas (sustantivos) para este estudio, extraídas del SENSEVAL-3 All-Words y SEMCOR 2.0.

Sustantivo	Número de sentidos	Número de ejemplos	Promedio de ejemplos por sentido
adult	2	10	5.0
link	2	10	5.0
formation	5	18	3.6
dirt	2	20	10.0
stone	3	25	8.3
hope	4	46	11.5
discussion	2	49	24.5
activity	3	92	30.7
plant	2	99	49.5
experience	3	125	41.7
state	4	200	50.0
thing	10	271	27.1

Tabla A1.1: Estadísticas del conjunto de datos.

Se ha seleccionado los siguientes clasificadores para la comparación: Naïve Bayes (NB), tabla de decisiones (DT), regresión localmente ponderada (LWR), k-nn y las máquinas de soporte de vectores (SVM). La tabla A1.2 muestra los resultados obtenidos en este estudio (porcentajes de precisión) utilizando diferentes tamaños de contexto (2, 4 o 6 palabras en total del contexto) y usando la *validación cruzada con 10 pliegues*.

Clasificador	N.B.			DT			LWL			SVM			KNN; K=1		
	2	4	6	2	4	6	2	4	6	2	4	6	2	4	6
adult	.40	.60	.50	.40	.40	.10	.50	.40	.40	.30	.40	.60	.40	.50	.50
link	.60	.80	.80	.60	.30	.30	.40	.30	.50	.20	.50	.50	.30	.70	.60
formation	.38	.61	.66	.11	.05	.05	.38	.16	.22	.33	.16	.16	.28	.28	.28
dirt	.80	.70	.60	.70	.70	.70	.80	.75	.75	.65	.75	.80	.65	.60	.55
stone	.60	.64	.64	.36	.40	.40	.44	.44	.48	.48	.44	.48	.48	.48	.48
hope	.37	.39	.37	.37	.34	.32	.34	.32	.32	.54	.45	.39	.33	.30	.22
discussion	.59	.63	.69	.53	.51	.61	.49	.49	.53	.55	.57	.57	.57	.63	.55
activity	.57	.59	.51	.50	.45	.46	.56	.48	.47	.59	.56	.60	.60	.61	.48
plant	.68	.59	.56	.60	.61	.60	.63	.64	.66	.57	.62	.59	.63	.64	.66
experiencie	.52	.45	.46	.44	.43	.42	.42	.42	.43	.50	.48	.51	.46	.47	.49
state	.65	.66	.64	.68	.65	.65	.68	.68	.68	.69	.66	.65	.66	.65	.62
thing	.29	.23	.24	.21	.21	.21	.26	.25	.24	.25	.22	.22	.28	.23	.24
Average precision	.50	.48	.47	.45	.43	.43	.47	.45	.46	.48	.47	.47	.48	.47	.45

Tabla A1.2: Porcentajes de precisión de los clasificadores supervisados.

Los resultados de este estudio han mostrados que entre más información contextual (más palabras del contexto) pueden aportar beneficios. Además entre los clasificadores, NB ha superado a la mayoría de las palabras.

Por ultimo, la realización de este experimento ha proporcionado experiencia de primera mano. Esta experiencia ha permitido el desarrollo del ensamble supervisado que se muestra en el presente trabajo.

Apéndice 2

Resultados de los ensambles usando el primer enfoque del capítulo 4.

En las siguientes páginas de este apéndice se detallan las tablas de los resultados de los ensambles usando el primer enfoque descrito en el capítulo 4 para cada palabra y combinaciones de clasificadores. . Donde los porcentajes utilizan las métricas P y TPL que representan respectivamente a la precisión y tamaño promedio de la lista de sentidos propuestos por el ensamble.

Palabra	Métrica	Umbral										B. L.
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	
argument	P	47.29	47.29	47.29	47.29	47.29	44.96	42.64	41.09	44.96	45.74	41.86
	TPL	1.33	1.33	1.33	1.33	1.31	1.26	1.09	1.04	1.16	1.16	1.00
arm	P	82.96	82.96	82.96	82.96	82.96	83.70	83.70	83.70	83.70	83.70	80.74
	TPL	1.00	1.00	1.00	1.00	1.00	1.01	1.02	1.02	1.02	1.02	1.00
atmosphere	P	59.80	59.80	59.80	59.80	59.80	58.82	50.98	51.96	57.84	58.82	53.92
	TPL	1.35	1.35	1.35	1.35	1.35	1.33	1.04	1.06	1.14	1.18	1.00
audience	P	80.37	80.37	80.37	80.37	80.37	80.37	80.37	80.37	80.37	80.37	66.36
	TPL	1.19	1.19	1.19	1.19	1.19	1.19	1.19	1.19	1.19	1.19	1.00
bank	P	5.07	5.07	5.07	5.07	5.07	5.07	5.80	5.80	5.80	5.80	4.35
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.02	1.03	1.03	1.00
degree	P	20.00	20.00	20.00	20.00	20.00	20.71	22.14	22.14	21.43	20.00	13.57
	TPL	1.40	1.40	1.40	1.40	1.40	1.17	1.17	1.17	1.15	1.00	1.00
difference	P	43.85	43.85	43.85	43.85	43.85	46.15	50.00	50.00	50.00	50.00	41.54
	TPL	1.05	1.05	1.05	1.05	1.06	1.11	1.22	1.27	1.27	1.27	1.00
difficulty	P	54.55	54.55	54.55	54.55	54.55	54.55	54.55	54.55	54.55	54.55	13.64
	TPL	1.09	1.09	1.09	1.09	1.09	1.18	1.18	1.18	1.23	1.23	1.00
disc	P	66.97	66.97	66.97	66.97	66.97	66.06	66.06	66.97	66.97	66.97	45.87
	TPL	1.31	1.31	1.31	1.31	1.31	1.28	1.21	1.18	1.15	1.16	1.00
image	P	33.33	33.33	33.33	33.33	34.67	34.67	34.67	33.33	33.33	33.33	29.33
	TPL	1.32	1.32	1.32	1.32	1.32	1.21	1.15	1.08	1.19	1.13	1.00
interest	P	57.89	57.89	57.89	56.84	57.89	53.68	54.74	53.68	54.74	56.84	52.63
	TPL	1.44	1.44	1.45	1.42	1.37	1.29	1.23	1.11	1.15	1.22	1.00
judgment	P	37.50	37.50	37.50	37.50	37.50	34.38	34.38	37.50	37.50	37.50	50.00
	TPL	1.19	1.19	1.19	1.19	1.16	1.09	1.06	1.09	1.09	1.09	1.00
organization	P	76.36	76.36	76.36	76.36	76.36	76.36	76.36	78.18	80.00	80.00	76.36
	TPL	1.04	1.04	1.04	1.04	1.04	1.05	1.05	1.13	1.24	1.25	1.00
paper	P	49.26	49.26	49.26	49.26	50.00	50.00	50.74	51.47	51.47	51.47	44.12
	TPL	1.27	1.23	1.23	1.23	1.24	1.24	1.24	1.18	1.20	1.20	1.00
party	P	64.00	64.00	64.00	64.00	64.00	64.00	64.00	64.00	64.80	65.60	62.40
	TPL	1.06	1.06	1.06	1.06	1.06	1.08	1.09	1.04	1.06	1.06	1.00
performance	P	37.36	37.36	37.36	37.36	37.36	38.46	37.36	37.36	39.56	39.56	38.46
	TPL	1.14	1.14	1.14	1.14	1.14	1.18	1.08	1.07	1.11	1.11	1.00
plan	P	72.73	72.73	72.73	72.73	72.73	72.73	72.73	68.69	72.73	72.73	72.73
	TPL	1.23	1.22	1.22	1.22	1.22	1.22	1.22	1.12	1.11	1.12	1.00
shelter	P	31.30	31.30	31.30	31.30	31.30	31.30	29.57	27.83	25.22	23.48	14.78
	TPL	1.56	1.56	1.56	1.56	1.56	1.57	1.50	1.30	1.17	1.08	1.00
sort	P	54.13	54.13	54.13	54.13	55.96	55.96	55.96	56.88	57.80	57.80	55.96
	TPL	1.01	1.01	1.01	1.01	1.05	1.06	1.07	1.08	1.11	1.11	1.00
Promedio	P	51.30	51.30	51.30	51.25	51.51	51.16	50.88	50.82	51.72	51.80	45.19
	TPL	1.21	1.21	1.21	1.21	1.20	1.19	1.15	1.12	1.14	1.14	1.00

Tabla A2.1: Tabla de los porcentajes de precisión (P) y tamaño promedio de la lista (TPL) usando varios umbrales para cada palabra usando 2 clasificadores supervisados para el ensamble.

Palabra	Métrica	Umbral										B. L.
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	
argument	P	40.31	40.31	40.31	40.31	41.09	42.64	44.96	44.19	40.31	40.31	41.86
	TPL	1.02	1.02	1.02	1.02	1.03	1.08	1.12	1.13	1.02	1.02	1.00
arm	P	82.96	82.96	82.96	82.96	82.96	82.96	82.96	82.96	82.96	82.96	80.74
	TPL	1.03	1.03	1.03	1.03	1.03	1.01	1.00	1.00	1.00	1.00	1.00
atmosphere	P	53.92	53.92	53.92	53.92	53.92	53.92	57.84	55.88	54.90	53.92	53.92
	TPL	1.00	1.00	1.00	1.00	1.00	1.01	1.14	1.11	1.04	1.00	1.00
audience	P	71.96	71.96	71.96	71.96	71.96	71.96	71.96	71.96	71.96	71.96	66.36
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
bank	P	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	4.35
	TPL	1.10	1.10	1.10	1.10	1.10	1.09	1.07	1.06	1.00	1.00	1.00
degree	P	17.14	17.14	17.14	17.14	17.14	19.29	19.29	19.29	19.29	20.00	13.57
	TPL	1.00	1.00	1.00	1.00	1.00	1.06	1.07	1.07	1.07	1.09	1.00
difference	P	46.15	46.15	46.15	46.15	46.15	46.15	46.15	46.15	46.15	46.15	41.54
	TPL	1.19	1.19	1.19	1.19	1.20	1.16	1.10	1.08	1.08	1.08	1.00
difficulty	P	54.55	54.55	54.55	54.55	54.55	50.00	50.00	50.00	50.00	50.00	13.64
	TPL	1.14	1.14	1.14	1.14	1.14	1.14	1.14	1.14	1.18	1.18	1.00
disc	P	60.55	60.55	60.55	60.55	60.55	60.55	62.39	61.47	60.55	60.55	45.87
	TPL	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.07	1.05	1.04	1.00
image	P	30.67	30.67	30.67	30.67	32.00	33.33	33.33	33.33	32.00	30.67	29.33
	TPL	1.04	1.04	1.04	1.04	1.04	1.08	1.12	1.09	1.05	1.03	1.00
interest	P	49.47	49.47	49.47	49.47	50.53	49.47	51.58	51.58	51.58	49.47	52.63
	TPL	1.06	1.06	1.06	1.07	1.07	1.07	1.08	1.04	1.05	1.04	1.00
judgment	P	34.38	34.38	34.38	34.38	34.38	34.38	34.38	34.38	34.38	34.38	50.00
	TPL	1.00	1.00	1.00	1.00	1.03	1.00	1.00	1.00	1.00	1.00	1.00
organization	P	76.36	76.36	76.36	76.36	76.36	76.36	76.36	78.18	80.00	80.00	76.36
	TPL	1.04	1.04	1.04	1.04	1.04	1.05	1.05	1.13	1.24	1.25	1.00
paper	P	50.00	50.00	50.00	50.00	50.00	50.00	49.26	47.79	47.06	47.06	44.12
	TPL	1.19	1.19	1.19	1.19	1.19	1.19	1.11	1.04	1.04	1.04	1.00
party	P	66.40	66.40	66.40	66.40	66.40	65.60	64.80	64.00	64.00	64.00	62.40
	TPL	1.07	1.07	1.07	1.07	1.07	1.06	1.05	1.04	1.02	1.02	1.00
performance	P	39.56	39.56	39.56	39.56	39.56	39.56	40.66	39.56	39.56	38.46	38.46
	TPL	1.07	1.07	1.07	1.07	1.07	1.07	1.13	1.15	1.05	1.02	1.00
plan	P	70.71	70.71	70.71	70.71	70.71	70.71	70.71	70.71	71.72	70.71	72.73
	TPL	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.08	1.02	1.02	1.00
shelter	P	21.74	21.74	21.74	21.74	21.74	21.74	21.74	21.74	22.61	23.48	14.78
	TPL	1.10	1.10	1.10	1.10	1.10	1.10	1.11	1.10	1.12	1.13	1.00
sort	P	57.80	57.80	57.80	57.80	56.88	56.88	55.05	55.05	54.13	54.13	55.96
	TPL	1.10	1.10	1.10	1.10	1.06	1.06	1.04	1.03	1.00	1.00	1.00
Promedio	P	48.93	48.93	48.93	48.93	49.05	48.98	49.39	49.12	48.85	48.59	45.19
	TPL	1.07	1.07	1.07	1.07	1.07	1.07	1.08	1.07	1.05	1.05	1.00

Tabla A2.2: Tabla de los porcentajes de precisión (P) y tamaño promedio de la lista (TPL) usando varios umbrales para cada palabra usando 3 clasificadores supervisados para el ensamble.

Palabra	Métrica	Umbral										B. L.
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	
argument	P	45.74	45.74	46.51	46.51	46.51	44.96	44.19	42.64	45.74	46.51	41.86
	TPL	1.14	1.14	1.16	1.16	1.16	1.14	1.10	1.09	1.14	1.16	1.00
arm	P	80.74	80.74	80.74	80.74	80.74	81.48	82.96	82.96	82.96	82.96	80.74
	TPL	1.00	1.00	1.00	1.00	1.00	1.01	1.03	1.03	1.03	1.03	1.00
atmosphere	P	60.78	60.78	60.78	60.78	60.78	60.78	56.86	56.86	58.82	60.78	53.92
	TPL	1.16	1.16	1.16	1.16	1.16	1.16	1.04	1.05	1.09	1.16	1.00
audience	P	78.50	78.50	78.50	78.50	78.50	78.50	78.50	78.50	78.50	78.50	66.36
	TPL	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.00
bank	P	4.35	4.35	4.35	4.35	4.35	4.35	4.35	5.07	5.07	5.07	4.35
	TPL	1.01	1.01	1.01	1.01	1.01	1.01	1.04	1.05	1.11	1.11	1.00
degree	P	17.14	17.14	17.14	17.14	17.14	17.14	17.14	17.14	17.14	17.14	13.57
	TPL	1.16	1.16	1.16	1.16	1.16	1.11	1.06	1.06	1.02	1.00	1.00
difference	P	41.54	41.54	41.54	41.54	41.54	44.62	46.15	46.15	46.15	46.15	41.54
	TPL	1.08	1.08	1.08	1.08	1.08	1.13	1.20	1.20	1.20	1.20	1.00
difficulty	P	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	13.64
	TPL	1.14	1.14	1.14	1.14	1.14	1.32	1.32	1.36	1.41	1.41	1.00
disc	P	63.30	63.30	63.30	63.30	63.30	63.30	65.14	62.39	62.39	63.30	45.87
	TPL	1.16	1.16	1.16	1.16	1.16	1.16	1.15	1.09	1.15	1.16	1.00
image	P	30.67	30.67	30.67	30.67	30.67	30.67	30.67	30.67	30.67	30.67	29.33
	TPL	1.05	1.05	1.05	1.05	1.05	1.05	1.04	1.07	1.08	1.08	1.00
interest	P	53.68	53.68	53.68	53.68	53.68	52.63	55.79	54.74	54.74	53.68	52.63
	TPL	1.06	1.06	1.06	1.06	1.06	1.06	1.08	1.07	1.07	1.06	1.00
judgment	P	53.13	53.13	53.13	53.13	53.13	53.13	53.13	53.13	53.13	53.13	50.00
	TPL	1.38	1.38	1.38	1.38	1.34	1.38	1.38	1.38	1.38	1.38	1.00
organization	P	76.36	76.36	76.36	76.36	76.36	76.36	76.36	78.18	80.00	80.00	76.36
	TPL	1.04	1.04	1.04	1.04	1.04	1.05	1.05	1.13	1.24	1.25	1.00
paper	P	47.79	47.79	47.79	47.79	47.79	47.79	48.53	48.53	48.53	48.53	44.12
	TPL	1.22	1.21	1.21	1.21	1.21	1.21	1.20	1.12	1.10	1.10	1.00
party	P	64.00	64.00	64.00	64.00	64.00	64.00	64.00	64.80	65.60	65.60	62.40
	TPL	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.08	1.10	1.10	1.00
performance	P	39.56	39.56	39.56	39.56	39.56	39.56	41.76	41.76	40.66	39.56	38.46
	TPL	1.15	1.15	1.15	1.15	1.15	1.14	1.23	1.25	1.12	1.05	1.00
plan	P	75.76	75.76	75.76	75.76	75.76	75.76	75.76	73.74	75.76	75.76	72.73
	TPL	1.09	1.09	1.09	1.09	1.09	1.09	1.09	1.04	1.10	1.09	1.00
shelter	P	21.74	21.74	21.74	21.74	21.74	21.74	21.74	20.87	21.74	21.74	14.78
	TPL	1.18	1.19	1.19	1.19	1.19	1.19	1.18	1.14	1.12	1.10	1.00
sort	P	54.13	54.13	54.13	54.13	54.13	56.88	59.63	59.63	59.63	59.63	55.96
	TPL	1.00	1.01	1.01	1.01	1.01	1.05	1.10	1.10	1.10	1.10	1.00
Promedio	P	50.47	50.47	50.51	50.51	50.51	50.72	51.19	50.93	51.43	51.51	45.19
	TPL	1.12	1.12	1.12	1.12	1.12	1.13	1.13	1.13	1.14	1.14	1.00

Tabla A2.3: Tabla de los porcentajes de precisión (P) y tamaño promedio de la lista (TPL) usando varios umbrales para cada palabra usando 4 clasificadores supervisados para el ensamble.

Palabra	Métrica	Umbral										B. L.
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	
argument	P	42.64	42.64	42.64	42.64	42.64	42.64	43.41	45.74	43.41	42.64	41.86
	TPL	1.02	1.02	1.00	1.00	1.00	1.02	1.07	1.09	1.01	1.00	1.00
arm	P	80.74	80.74	80.74	80.74	80.74	80.74	80.74	80.74	80.74	80.74	80.74
	TPL	1.00	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.01	1.01	1.00
atmosphere	P	54.90	54.90	54.90	54.90	54.90	54.90	56.86	57.84	56.86	54.90	53.92
	TPL	1.02	1.02	1.02	1.02	1.02	1.02	1.09	1.10	1.06	1.02	1.00
audience	P	67.29	67.29	67.29	67.29	67.29	67.29	67.29	67.29	67.29	67.29	66.36
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
bank	P	4.35	4.35	4.35	4.35	4.35	4.35	4.35	5.07	5.07	5.07	4.35
	TPL	1.01	1.01	1.01	1.01	1.01	1.01	1.04	1.05	1.11	1.11	1.00
degree	P	16.43	16.43	16.43	16.43	16.43	17.14	17.14	17.14	17.14	17.86	13.57
	TPL	1.06	1.06	1.06	1.06	1.06	1.07	1.07	1.07	1.06	1.08	1.00
difference	P	41.54	41.54	41.54	41.54	41.54	41.54	41.54	41.54	41.54	41.54	41.54
	TPL	1.05	1.05	1.05	1.05	1.05	1.04	1.03	1.04	1.04	1.04	1.00
difficulty	P	45.45	45.45	45.45	45.45	45.45	36.36	36.36	31.82	31.82	31.82	13.64
	TPL	1.23	1.23	1.23	1.23	1.23	1.18	1.18	1.14	1.18	1.18	1.00
disc	P	61.47	61.47	61.47	61.47	61.47	61.47	61.47	61.47	61.47	61.47	45.87
	TPL	1.12	1.12	1.12	1.12	1.12	1.10	1.08	1.15	1.10	1.10	1.00
image	P	30.67	30.67	30.67	30.67	30.67	30.67	30.67	30.67	30.67	30.67	29.33
	TPL	1.08	1.08	1.08	1.08	1.08	1.07	1.05	1.08	1.07	1.05	1.00
interest	P	51.58	51.58	51.58	51.58	52.63	51.58	52.63	51.58	51.58	51.58	52.63
	TPL	1.08	1.08	1.08	1.09	1.09	1.09	1.07	1.05	1.05	1.07	1.00
judgment	P	53.13	53.13	53.13	53.13	53.13	53.13	53.13	53.13	53.13	53.13	50.00
	TPL	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.00
organization	P	76.36	76.36	76.36	76.36	76.36	76.36	76.36	78.18	80.00	80.00	76.36
	TPL	1.04	1.04	1.04	1.04	1.04	1.05	1.05	1.13	1.24	1.25	1.00
paper	P	48.53	48.53	48.53	48.53	48.53	48.53	48.53	47.79	47.79	47.79	44.12
	TPL	1.21	1.19	1.19	1.19	1.19	1.19	1.17	1.09	1.08	1.08	1.00
party	P	64.00	64.00	64.00	64.00	64.00	63.20	63.20	64.00	64.80	64.80	62.40
	TPL	1.04	1.04	1.04	1.04	1.04	1.02	1.02	1.05	1.06	1.06	1.00
performance	P	39.56	39.56	39.56	39.56	39.56	39.56	40.66	40.66	40.66	40.66	38.46
	TPL	1.11	1.11	1.11	1.11	1.11	1.08	1.09	1.07	1.08	1.08	1.00
plan	P	76.77	76.77	76.77	76.77	76.77	76.77	76.77	74.75	76.77	76.77	72.73
	TPL	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.05	1.06	1.06	1.00
shelter	P	20.87	20.00	20.00	20.00	20.00	20.00	20.87	24.35	26.09	26.09	14.78
	TPL	1.08	1.06	1.07	1.07	1.07	1.08	1.10	1.15	1.20	1.18	1.00
sort	P	58.72	58.72	58.72	58.72	57.80	57.80	55.96	55.96	55.05	55.05	55.96
	TPL	1.09	1.09	1.09	1.09	1.07	1.06	1.03	1.03	1.00	1.00	1.00
Promedio	P	49.21	49.16	49.16	49.16	49.17	48.63	48.84	48.93	49.05	48.94	45.19
	TPL	1.07	1.07	1.07	1.07	1.07	1.06	1.07	1.07	1.08	1.07	1.00

Tabla A2.4: Tabla de los porcentajes de precisión (P) y tamaño promedio de la lista (TPL) usando varios umbrales para cada palabra usando 5 clasificadores supervisados para el ensamble.

Palabra	Métrica	Umbral										B. L.
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	
argument	P	42.64	42.64	42.64	42.64	42.64	41.86	41.09	41.09	42.64	42.64	41.86
	TPL	1.04	1.04	1.04	1.04	1.04	1.03	1.01	1.02	1.04	1.04	1.00
arm	P	80.74	80.74	80.74	80.74	80.74	80.74	80.74	80.74	80.74	80.74	80.74
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
atmosphere	P	57.84	57.84	57.84	57.84	57.84	57.84	54.90	54.90	55.88	57.84	53.92
	TPL	1.07	1.07	1.07	1.07	1.07	1.07	1.02	1.02	1.04	1.07	1.00
audience	P	71.03	71.03	71.03	71.03	71.03	71.03	71.03	71.03	71.03	71.03	66.36
	TPL	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.00
bank	P	4.35	4.35	4.35	4.35	4.35	4.35	4.35	4.35	4.35	4.35	4.35
	TPL	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.03	1.01	1.01	1.00
degree	P	14.29	14.29	14.29	14.29	14.29	15.00	16.43	16.43	16.43	16.43	13.57
	TPL	1.04	1.04	1.04	1.04	1.04	1.05	1.08	1.08	1.06	1.06	1.00
difference	P	42.31	42.31	42.31	42.31	42.31	42.31	44.62	44.62	44.62	44.62	41.54
	TPL	1.04	1.04	1.04	1.04	1.04	1.05	1.08	1.09	1.09	1.09	1.00
difficulty	P	31.82	31.82	31.82	31.82	31.82	31.82	31.82	31.82	31.82	31.82	13.64
	TPL	1.09	1.09	1.09	1.09	1.14	1.27	1.32	1.32	1.36	1.36	1.00
disc	P	56.88	56.88	56.88	56.88	56.88	57.80	59.63	55.05	56.88	56.88	45.87
	TPL	1.18	1.18	1.18	1.18	1.17	1.19	1.17	1.13	1.20	1.20	1.00
image	P	30.67	30.67	30.67	30.67	30.67	30.67	30.67	32.00	32.00	30.67	29.33
	TPL	1.07	1.07	1.07	1.07	1.07	1.01	1.03	1.07	1.08	1.07	1.00
interest	P	54.74	54.74	54.74	54.74	54.74	52.63	55.79	54.74	54.74	54.74	52.63
	TPL	1.12	1.12	1.12	1.12	1.12	1.06	1.12	1.12	1.09	1.11	1.00
judgment	P	56.25	56.25	56.25	56.25	56.25	56.25	56.25	56.25	56.25	56.25	50.00
	TPL	1.09	1.09	1.09	1.09	1.09	1.09	1.09	1.09	1.09	1.09	1.00
organization	P	76.36	76.36	76.36	76.36	76.36	76.36	76.36	78.18	80.00	80.00	76.36
	TPL	1.04	1.04	1.04	1.04	1.04	1.05	1.05	1.13	1.24	1.25	1.00
paper	P	44.12	44.12	44.12	44.12	44.12	44.12	44.12	46.32	46.32	46.32	44.12
	TPL	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.12	1.11	1.11	1.00
party	P	63.20	63.20	63.20	63.20	63.20	64.00	64.80	64.00	63.20	63.20	62.40
	TPL	1.06	1.06	1.06	1.06	1.06	1.07	1.08	1.06	1.04	1.04	1.00
performance	P	46.15	46.15	46.15	46.15	46.15	46.15	41.76	39.56	39.56	39.56	38.46
	TPL	1.38	1.37	1.37	1.37	1.37	1.36	1.25	1.16	1.13	1.09	1.00
plan	P	74.75	74.75	74.75	74.75	74.75	74.75	74.75	75.76	75.76	74.75	72.73
	TPL	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.07	1.04	1.03	1.00
shelter	P	20.87	20.87	20.87	20.87	20.87	20.87	20.87	19.13	20.87	21.74	14.78
	TPL	1.14	1.15	1.16	1.16	1.16	1.16	1.15	1.12	1.15	1.14	1.00
sort	P	55.05	55.05	55.05	55.05	55.05	56.88	58.72	58.72	58.72	58.72	55.96
	TPL	1.00	1.01	1.01	1.01	1.01	1.04	1.08	1.08	1.08	1.08	1.00
Promedio	P	48.63	48.63	48.63	48.63	48.63	48.71	48.88	48.67	49.04	49.07	45.19
	TPL	1.08	1.08	1.08	1.08	1.08	1.09	1.09	1.09	1.10	1.10	1.00

Tabla A2.5: Tabla de los porcentajes de precisión (P) y tamaño promedio de la lista (TPL) usando varios umbrales para cada palabra usando 6 clasificadores supervisados para el ensamble.

Apéndice 3

Resultados de los ensambles usando el segundo enfoque del capítulo 4.

En las siguientes páginas de este apéndice se detallan las tablas de los resultados de los ensambles usando el segundo enfoque descrito en el capítulo 4 para cada palabra y combinaciones de clasificadores. Donde los porcentajes utilizan las métricas P, R, C y TPL que representan respectivamente a la precisión, el recuerdo, la cobertura y el tamaño promedio de la lista de sentidos propuestos por el ensamble. Y los valores en **negrita** indican que los porcentajes superan al *baseline* (B.L.).

Palabra	Métrica	Umbral										B. L.
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	
argument	R	0.78	1.55	1.55	1.55	2.33	4.65	9.30	17.83	41.86	45.74	41.86
	P	33.33	25.00	25.00	25.00	30.00	25.00	27.27	35.94	45.00	45.74	41.86
	C	2.33	6.20	6.20	6.20	7.75	18.60	34.11	49.61	93.02	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.08	1.17	1.16	1.00
arm	R	80.74	80.74	80.74	80.74	80.74	82.22	83.70	83.70	83.70	83.70	80.74
	P	83.85	83.85	83.85	83.85	83.85	84.09	84.33	84.33	84.33	84.33	80.74
	C	96.30	96.30	96.30	96.30	96.30	97.78	99.26	99.26	99.26	99.26	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.02	1.02	1.02	1.02	1.02	1.00
atmosphere	R	0.00	0.00	2.94	2.94	2.94	7.84	17.65	33.33	50.98	58.82	53.92
	P	0.00	0.00	75.00	75.00	50.00	40.00	36.73	46.58	57.78	58.82	100.00
	C	0.00	0.00	3.92	3.92	5.88	19.61	48.04	71.57	88.24	100.00	100.00
	TPL	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.08	1.16	1.18	1.00
audience	R	77.57	77.57	77.57	77.57	77.57	79.44	80.37	80.37	80.37	80.37	66.36
	P	79.81	79.81	79.81	79.81	79.81	80.19	80.37	80.37	80.37	80.37	66.36
	C	97.20	97.20	97.20	97.20	97.20	99.07	100.00	100.00	100.00	100.00	100.00
	TPL	1.19	1.19	1.19	1.19	1.19	1.19	1.19	1.19	1.19	1.19	1.00
bank	R	3.62	3.62	3.62	3.62	3.62	3.62	4.35	5.07	5.80	5.80	4.35
	P	4.27	4.27	4.27	4.27	4.27	4.24	4.92	5.43	5.80	5.80	4.35
	C	84.78	84.78	84.78	84.78	84.78	85.51	88.41	93.48	100.00	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.02	1.02	1.03	1.03	1.00
degree	R	0.71	1.43	2.14	2.14	2.14	6.43	9.29	9.29	15.00	20.00	13.57
	P	100.00	100.00	100.00	100.00	100.00	45.00	44.83	44.83	47.73	50.91	13.57
	C	0.71	1.43	2.14	2.14	2.14	14.29	20.71	20.71	31.43	39.29	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
difference	R	34.62	34.62	34.62	34.62	34.62	40.77	47.69	49.23	49.23	49.23	41.54
	P	43.69	43.69	43.69	43.69	43.69	47.32	50.00	50.00	50.00	50.00	41.54
	C	79.23	79.23	79.23	79.23	79.23	86.15	95.38	98.46	98.46	98.46	100.00
	TPL	1.07	1.07	1.07	1.07	1.08	1.13	1.23	1.27	1.27	1.27	1.00
difficulty	R	45.45	45.45	45.45	45.45	54.55	54.55	54.55	54.55	54.55	54.55	13.64
	P	83.33	76.92	76.92	76.92	75.00	63.16	57.14	54.55	54.55	54.55	13.64
	C	54.55	59.09	59.09	59.09	72.73	86.36	95.45	100.00	100.00	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.05	1.05	1.05	1.05	1.05	1.00
disc	R	0.00	0.00	0.00	0.00	1.83	4.59	21.10	46.79	66.06	66.97	45.87
	P	0.00	0.00	0.00	0.00	50.00	55.56	71.88	73.91	67.92	68.22	45.87
	C	0.00	0.00	0.00	0.00	3.67	8.26	29.36	63.30	97.25	98.17	100.00
	TPL	0.00	0.00	0.00	0.00	1.00	1.11	1.16	1.20	1.15	1.16	1.00
image	R	2.67	2.67	5.33	6.67	6.67	16.00	26.67	30.67	30.67	32.00	29.33
	P	50.00	40.00	50.00	55.56	62.50	60.00	55.56	50.00	37.10	32.43	29.33
	C	5.33	6.67	10.67	12.00	10.67	26.67	48.00	61.33	82.67	98.67	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.07	1.13	1.14	1.00

Continúa en la tabla A4.2...

Tabla A3.1: Primera parte de dos de la tabla de los porcentajes de las métricas P, R, C y TPL usando varios umbrales para cada palabra usando 2 clasificadores supervisados para el ensamble.

Palabra	Métrica	Umbral										B. L.
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	
interest	R	0.00	0.00	0.00	0.00	6.32	20.00	43.16	49.47	54.74	56.84	52.63
	P	0.00	0.00	0.00	0.00	60.00	50.00	56.16	53.41	55.91	56.84	52.63
	C	0.00	0.00	0.00	1.05	10.53	40.00	76.84	92.63	97.89	100.00	100.00
	TPL	0.00	0.00	0.00	1.00	1.10	1.13	1.12	1.11	1.15	1.22	1.00
judgment	R	0.00	0.00	0.00	0.00	0.00	12.50	34.38	37.50	37.50	37.50	50.00
	P	0.00	0.00	0.00	0.00	0.00	22.22	34.38	37.50	37.50	37.50	50.00
	C	0.00	0.00	0.00	0.00	3.13	56.25	100.00	100.00	100.00	100.00	100.00
	TPL	0.00	0.00	0.00	0.00	2.00	1.06	1.06	1.09	1.09	1.09	1.00
organization	R	0.00	0.00	0.00	1.82	1.82	1.82	18.18	54.55	74.55	76.36	76.36
	P	0.00	0.00	0.00	100.00	100.00	100.00	71.43	76.92	78.85	77.78	76.36
	C	0.00	0.00	0.00	1.82	1.82	1.82	25.45	70.91	94.55	98.18	100.00
	TPL	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
paper	R	31.62	33.09	33.09	33.09	33.82	33.82	37.50	48.53	51.47	51.47	44.12
	P	43.88	45.00	45.00	45.00	46.00	46.00	47.66	50.77	51.85	51.85	44.12
	C	72.06	73.53	73.53	73.53	73.53	73.53	78.68	95.59	99.26	99.26	100.00
	TPL	1.13	1.15	1.15	1.15	1.17	1.17	1.18	1.18	1.20	1.20	1.00
party	R	9.60	9.60	10.40	10.40	13.60	24.00	31.20	60.80	64.80	65.60	62.40
	P	52.17	52.17	54.17	54.17	60.71	73.17	67.24	66.09	65.85	66.67	62.40
	C	18.40	18.40	19.20	19.20	22.40	32.80	46.40	92.00	98.40	98.40	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.05	1.05	1.04	1.06	1.07	1.00
performance	R	9.89	9.89	9.89	9.89	9.89	13.19	23.08	32.97	38.46	39.56	38.46
	P	30.00	29.03	29.03	29.03	29.03	33.33	35.59	38.96	40.23	39.56	38.46
	C	32.97	34.07	34.07	34.07	34.07	39.56	64.84	84.62	95.60	100.00	100.00
	TPL	1.03	1.03	1.03	1.03	1.03	1.06	1.07	1.08	1.11	1.11	1.00
plan	R	1.01	2.02	5.05	6.06	6.06	6.06	6.06	14.14	61.62	72.73	72.73
	P	100.00	100.00	100.00	75.00	75.00	75.00	75.00	51.85	70.93	72.73	72.73
	C	1.01	2.02	5.05	8.08	8.08	8.08	8.08	27.27	86.87	100.00	100.00
	TPL	2.00	2.00	1.40	1.25	1.25	1.25	1.25	1.07	1.13	1.12	1.00
shelter	R	7.83	7.83	7.83	7.83	7.83	7.83	8.70	18.26	22.61	22.61	14.78
	P	23.08	23.08	22.50	22.50	22.50	22.50	23.26	33.87	32.91	30.59	14.78
	C	33.91	33.91	34.78	34.78	34.78	34.78	37.39	53.91	68.70	73.91	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.10	1.11	1.11	1.00
sort	R	7.34	7.34	7.34	8.26	14.68	33.94	52.29	54.13	57.80	57.80	55.96
	P	33.33	33.33	33.33	34.62	40.00	52.11	55.34	56.19	57.80	57.80	55.96
	C	22.02	22.02	22.02	23.85	36.70	65.14	94.50	96.33	100.00	100.00	100.00
	TPL	1.04	1.04	1.04	1.04	1.13	1.08	1.08	1.09	1.11	1.11	1.00
Promedio	R	16.50	16.71	17.24	17.51	19.00	23.86	32.06	41.11	49.57	51.46	45.19
	P	40.04	38.75	43.29	47.60	53.28	51.52	51.53	52.18	53.81	53.81	45.19
	C	31.62	32.36	33.06	33.54	36.07	47.07	62.68	77.42	91.14	94.93	100.00
	TPL	0.81	0.81	0.84	0.93	1.10	1.07	1.08	1.09	1.11	1.12	1.00

Tabla A3.2: Segunda parte de dos de la tabla de los porcentajes de las métricas P, R, C y TPL usando varios umbrales para cada palabra usando 2 clasificadores supervisados para el ensamble.

Palabra	Métrica	Umbral										B. L.
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	
argument	R	0.78	1.55	1.55	1.55	2.33	4.65	9.30	16.28	36.43	40.31	41.86
	P	33.33	25.00	25.00	25.00	30.00	25.00	27.27	32.81	39.17	40.31	41.86
	C	2.33	6.20	6.20	6.20	7.75	18.60	34.11	49.61	93.02	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.02	1.02	1.02	1.00
arm	R	80.74	80.74	80.74	80.74	80.74	81.48	82.96	82.96	82.96	82.96	80.74
	P	81.34	81.34	81.34	81.34	81.34	82.09	83.58	83.58	83.58	83.58	80.74
	C	99.26	99.26	99.26	99.26	99.26	99.26	99.26	99.26	99.26	99.26	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
atmosphere	R	0.00	0.00	2.94	2.94	2.94	7.84	19.61	31.37	47.06	53.92	53.92
	P	0.00	0.00	75.00	75.00	50.00	40.00	38.46	43.84	53.33	53.92	53.92
	C	0.00	0.00	3.92	3.92	5.88	19.61	50.98	71.57	88.24	100.00	100.00
	TPL	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
audience	R	69.16	69.16	69.16	69.16	69.16	71.03	71.96	71.96	71.96	71.96	66.36
	P	71.15	71.15	71.15	71.15	71.15	71.70	71.96	71.96	71.96	71.96	66.36
	C	97.20	97.20	97.20	97.20	97.20	99.07	100.00	100.00	100.00	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
bank	R	3.62	3.62	3.62	3.62	3.62	3.62	3.62	4.35	5.07	5.07	4.35
	P	3.82	3.82	3.82	3.82	3.82	3.82	3.79	4.38	5.07	5.07	4.35
	C	94.93	94.93	94.93	94.93	94.93	94.93	95.65	99.28	100.00	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
degree	R	0.71	1.43	2.14	2.14	2.14	6.43	9.29	9.29	15.00	20.00	13.57
	P	100.00	100.00	100.00	100.00	100.00	45.00	44.83	44.83	47.73	50.91	13.57
	C	0.71	1.43	2.14	2.14	2.14	14.29	20.71	20.71	31.43	39.29	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
difference	R	36.92	36.92	36.92	36.92	36.92	40.77	43.85	45.38	45.38	45.38	41.54
	P	39.02	39.02	39.02	39.02	39.02	42.06	44.19	45.38	45.38	45.38	41.54
	C	94.62	94.62	94.62	94.62	94.62	96.92	99.23	100.00	100.00	100.00	100.00
	TPL	1.03	1.03	1.03	1.03	1.04	1.05	1.06	1.06	1.06	1.06	1.00
difficulty	R	45.45	45.45	45.45	45.45	54.55	50.00	50.00	50.00	50.00	50.00	13.64
	P	83.33	76.92	76.92	76.92	75.00	57.89	52.38	50.00	50.00	50.00	13.64
	C	54.55	59.09	59.09	59.09	72.73	86.36	95.45	100.00	100.00	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
disc	R	0.00	0.00	0.00	0.00	1.83	3.67	20.18	41.28	59.63	60.55	45.87
	P	0.00	0.00	0.00	0.00	50.00	44.44	64.71	62.50	60.75	61.68	45.87
	C	0.00	0.00	0.00	0.00	3.67	8.26	31.19	66.06	98.17	98.17	100.00
	TPL	0.00	0.00	0.00	0.00	1.00	1.00	1.03	1.03	1.04	1.04	1.00
image	R	2.67	2.67	5.33	6.67	6.67	16.00	26.67	32.00	30.67	29.33	29.33
	P	50.00	40.00	50.00	55.56	62.50	60.00	54.05	50.00	35.94	29.33	29.33
	C	5.33	6.67	10.67	12.00	10.67	26.67	49.33	64.00	85.33	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.02	1.01	1.00

Continúa en la tabla A4.4...

Tabla A3.3: Primera parte de dos de la tabla de los porcentajes de las métricas P, R, C y TPL usando varios umbrales para cada palabra usando 3 clasificadores supervisados para el ensamble.

Palabra	Métrica	Umbral										B. L.
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	
interest	R	0.00	0.00	0.00	0.00	6.32	18.95	41.05	47.37	51.58	49.47	52.63
	P	0.00	0.00	0.00	0.00	60.00	46.15	52.00	50.00	51.58	49.47	52.63
	C	0.00	0.00	0.00	1.05	10.53	41.05	78.95	94.74	100.00	100.00	100.00
	TPL	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.02	1.03	1.04	1.00
judgment	R	0.00	0.00	0.00	0.00	0.00	12.50	34.38	34.38	34.38	34.38	50.00
	P	0.00	0.00	0.00	0.00	0.00	22.22	34.38	34.38	34.38	34.38	50.00
	C	0.00	0.00	0.00	0.00	6.25	56.25	100.00	100.00	100.00	100.00	100.00
	TPL	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
organization	R	0.00	0.00	0.00	1.82	1.82	1.82	18.18	54.55	74.55	76.36	76.36
	P	0.00	0.00	0.00	100.00	100.00	100.00	71.43	76.92	78.85	77.78	76.36
	C	0.00	0.00	0.00	1.82	1.82	1.82	25.45	70.91	94.55	98.18	100.00
	TPL	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
paper	R	31.62	33.09	33.09	33.09	33.09	33.09	36.76	44.85	47.06	47.06	44.12
	P	38.05	39.13	39.13	39.13	39.13	39.13	42.37	46.21	47.06	47.06	44.12
	C	83.09	84.56	84.56	84.56	84.56	84.56	86.76	97.06	100.00	100.00	100.00
	TPL	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.03	1.03	1.00
party	R	10.40	10.40	11.20	11.20	14.40	24.00	30.40	59.20	63.20	63.20	62.40
	P	52.00	52.00	53.85	53.85	60.00	69.77	63.33	62.71	63.20	63.20	62.40
	C	20.00	20.00	20.80	20.80	24.00	34.40	48.00	94.40	100.00	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
perfomance	R	9.89	9.89	9.89	9.89	9.89	13.19	24.18	32.97	38.46	38.46	38.46
	P	29.03	28.13	28.13	28.13	28.13	32.43	33.85	34.88	38.89	38.46	38.46
	C	34.07	35.16	35.16	35.16	35.16	40.66	71.43	94.51	98.90	100.00	100.00
	TPL	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.02	1.02	1.02	1.00
plan	R	1.01	2.02	5.05	6.06	6.06	6.06	6.06	14.14	60.61	70.71	72.73
	P	100.00	100.00	100.00	75.00	75.00	75.00	75.00	51.85	69.77	70.71	72.73
	C	1.01	2.02	5.05	8.08	8.08	8.08	8.08	27.27	86.87	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.02	1.00
shelter	R	7.83	7.83	7.83	7.83	7.83	7.83	8.70	16.52	21.74	22.61	14.78
	P	23.08	23.08	22.50	22.50	22.50	22.50	23.26	30.65	31.25	30.23	14.78
	C	33.91	33.91	34.78	34.78	34.78	34.78	37.39	53.91	69.57	74.78	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.02	1.00
sort	R	7.34	7.34	7.34	8.26	13.76	33.03	49.54	51.38	54.13	54.13	55.96
	P	33.33	33.33	33.33	34.62	37.50	50.70	52.43	53.33	54.13	54.13	55.96
	C	22.02	22.02	22.02	23.85	36.70	65.14	94.50	96.33	100.00	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Promedio	R	16.22	16.43	16.96	17.23	18.63	22.94	30.88	38.96	46.84	48.20	45.19
	P	38.82	37.52	42.06	46.37	51.85	48.94	49.12	48.96	50.63	50.40	45.19
	C	33.84	34.58	35.28	35.76	38.46	48.98	64.55	78.93	91.86	95.25	100.00
	TPL	0.74	0.74	0.79	0.90	1.00	1.00	1.01	1.01	1.01	1.01	1.00

Tabla A3.4: Segunda parte de dos de la tabla de los porcentajes de las métricas P, R, C y TPL usando varios umbrales para cada palabra usando 3 clasificadores supervisados para el ensamble.

Palabra	Métrica	Umbral										B. L.
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	
argument	R	0.78	3.10	3.88	3.88	4.65	6.98	11.63	18.60	41.86	46.51	41.86
	P	33.33	50.00	62.50	62.50	60.00	37.50	34.09	37.50	45.00	46.51	41.86
	C	2.33	6.20	6.20	6.20	7.75	18.60	34.11	49.61	93.02	100.00	100.00
	TPL	1.00	1.38	1.63	1.63	1.50	1.21	1.11	1.14	1.15	1.16	1.00
arm	R	80.74	80.74	80.74	80.74	80.74	81.48	82.96	82.96	82.96	82.96	80.74
	P	81.34	81.34	81.34	81.34	81.34	82.09	83.58	83.58	83.58	83.58	80.74
	C	99.26	99.26	99.26	99.26	99.26	99.26	99.26	99.26	99.26	99.26	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.01	1.03	1.03	1.03	1.03	1.00
atmosphere	R	0.00	0.00	2.94	2.94	2.94	7.84	19.61	34.31	51.96	60.78	53.92
	P	0.00	0.00	75.00	75.00	50.00	40.00	38.46	47.95	58.89	60.78	53.92
	C	0.00	0.00	3.92	3.92	5.88	19.61	50.98	71.57	88.24	100.00	100.00
	TPL	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.07	1.10	1.16	1.00
audience	R	75.70	75.70	75.70	75.70	75.70	77.57	78.50	78.50	78.50	78.50	66.36
	P	77.88	77.88	77.88	77.88	77.88	78.30	78.50	78.50	78.50	78.50	66.36
	C	97.20	97.20	97.20	97.20	97.20	99.07	100.00	100.00	100.00	100.00	100.00
	TPL	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.00
bank	R	3.62	3.62	3.62	3.62	3.62	3.62	3.62	4.35	5.07	5.07	4.35
	P	3.76	3.76	3.76	3.76	3.76	3.76	3.73	4.35	5.07	5.07	4.35
	C	96.38	96.38	96.38	96.38	96.38	96.38	97.10	100.00	100.00	100.00	100.00
	TPL	1.01	1.01	1.01	1.01	1.01	1.02	1.04	1.05	1.11	1.11	1.00
degree	R	1.43	2.14	2.86	2.86	2.86	7.14	10.00	10.00	15.71	20.71	13.57
	P	100.00	100.00	100.00	100.00	100.00	47.62	46.67	46.67	48.89	51.79	13.57
	C	1.43	2.14	2.86	2.86	2.86	15.00	21.43	21.43	32.14	40.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
difference	R	40.00	40.00	40.00	40.00	40.00	43.08	45.38	46.15	46.15	46.15	41.54
	P	40.94	40.94	40.94	40.94	40.94	43.75	45.74	46.15	46.15	46.15	41.54
	C	97.69	97.69	97.69	97.69	97.69	98.46	99.23	100.00	100.00	100.00	100.00
	TPL	1.07	1.07	1.07	1.07	1.07	1.12	1.19	1.20	1.20	1.20	1.00
difficulty	P	45.45	45.45	45.45	45.45	54.55	50.00	50.00	50.00	50.00	50.00	13.64
	R	83.33	76.92	76.92	76.92	75.00	57.89	52.38	50.00	50.00	50.00	13.64
	C	54.55	59.09	59.09	59.09	72.73	86.36	95.45	100.00	100.00	100.00	100.00
	TPL	1.08	1.08	1.08	1.08	1.06	1.21	1.19	1.23	1.23	1.23	1.00
disc	R	0.00	0.00	0.00	0.00	2.75	5.50	22.94	42.20	62.39	63.30	45.87
	P	0.00	0.00	0.00	0.00	50.00	60.00	62.50	60.53	63.55	64.49	45.87
	C	1.83	1.83	1.83	1.83	5.50	9.17	36.70	69.72	98.17	98.17	100.00
	TPL	1.00	1.00	1.00	1.00	1.17	1.10	1.05	1.07	1.15	1.16	1.00
image	R	2.67	2.67	5.33	6.67	6.67	16.00	26.67	32.00	30.67	29.33	29.33
	P	50.00	40.00	50.00	55.56	62.50	60.00	54.05	50.00	35.94	29.33	29.33
	C	5.33	6.67	10.67	12.00	10.67	26.67	49.33	64.00	85.33	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.06	1.08	1.07	1.00

Continúa en la tabla A4.6...

Tabla A3.5: Primera parte de dos de la tabla de los porcentajes de las métricas P, R, C y TPL usando varios umbrales para cada palabra usando 4 clasificadores supervisados para el ensamble.

Palabra	Métrica	Umbral										B. L.
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	
interest	R	0.00	0.00	0.00	0.00	6.32	18.95	43.16	50.53	54.74	53.68	52.63
	P	0.00	0.00	0.00	0.00	54.55	41.86	51.25	51.61	54.74	53.68	52.63
	C	0.00	0.00	0.00	1.05	11.58	45.26	84.21	97.89	100.00	100.00	100.00
	TPL	0.00	0.00	0.00	1.00	1.00	1.00	1.05	1.06	1.06	1.06	1.00
judgment	R	0.00	0.00	0.00	0.00	0.00	31.25	53.13	53.13	53.13	53.13	50.00
	P	0.00	0.00	0.00	0.00	0.00	55.56	53.13	53.13	53.13	53.13	50.00
	C	0.00	0.00	0.00	0.00	6.25	56.25	100.00	100.00	100.00	100.00	100.00
	TPL	0.00	0.00	0.00	0.00	1.00	1.44	1.38	1.38	1.38	1.38	1.00
organization	R	0.00	0.00	0.00	1.82	1.82	1.82	18.18	54.55	74.55	76.36	76.36
	P	0.00	0.00	0.00	100.00	100.00	100.00	71.43	76.92	78.85	77.78	76.36
	C	0.00	0.00	0.00	1.82	1.82	1.82	25.45	70.91	94.55	98.18	100.00
	TPL	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
paper	R	33.09	34.56	34.56	34.56	34.56	34.56	37.50	46.32	48.53	48.53	44.12
	P	34.88	36.43	36.43	36.43	36.43	36.43	39.23	47.01	48.53	48.53	44.12
	C	94.85	94.85	94.85	94.85	94.85	94.85	95.59	98.53	100.00	100.00	100.00
	TPL	1.13	1.13	1.13	1.13	1.13	1.13	1.13	1.10	1.10	1.10	1.00
party	R	10.40	10.40	11.20	11.20	14.40	24.00	30.40	60.80	64.80	64.80	62.40
	P	50.00	50.00	51.85	51.85	58.06	69.77	63.33	64.41	64.80	64.80	62.40
	C	20.80	20.80	21.60	21.60	24.80	34.40	48.00	94.40	100.00	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.06	1.08	1.08	1.00
perfomance	R	26.37	26.37	26.37	26.37	26.37	29.67	35.16	39.56	39.56	39.56	38.46
	P	28.24	28.24	28.24	28.24	28.24	31.40	36.36	39.56	39.56	39.56	38.46
	C	93.41	93.41	93.41	93.41	93.41	94.51	96.70	100.00	100.00	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.10	1.14	1.09	1.05	1.00
plan	R	1.01	2.02	5.05	6.06	6.06	6.06	6.06	14.14	65.66	75.76	72.73
	P	100.00	100.00	100.00	75.00	75.00	75.00	75.00	51.85	75.58	75.76	72.73
	C	1.01	2.02	5.05	8.08	8.08	8.08	8.08	27.27	86.87	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.10	1.09	1.00
shelter	R	7.83	7.83	7.83	7.83	7.83	7.83	8.70	17.39	22.61	23.48	14.78
	P	23.08	23.08	22.50	22.50	22.50	22.50	23.26	32.26	32.50	31.40	14.78
	C	33.91	33.91	34.78	34.78	34.78	34.78	37.39	53.91	69.57	74.78	100.00
	TPL	1.00	1.03	1.03	1.03	1.03	1.03	1.02	1.08	1.08	1.08	1.00
sort	P	7.34	7.34	7.34	8.26	13.76	35.78	55.05	56.88	59.63	59.63	55.96
	R	33.33	33.33	33.33	34.62	37.50	54.93	58.25	59.05	59.63	59.63	55.96
	C	22.02	22.02	22.02	23.85	36.70	65.14	94.50	96.33	100.00	100.00	100.00
	TPL	1.00	1.04	1.04	1.04	1.03	1.07	1.11	1.10	1.10	1.10	1.00
Promedio	P	17.71	18.00	18.57	18.84	20.29	25.74	33.61	41.70	49.92	51.49	45.19
	R	38.95	39.05	44.25	48.56	53.35	52.55	51.10	51.63	53.84	53.71	45.19
	C	38.00	38.60	39.31	39.78	42.54	52.82	67.03	79.73	91.95	95.28	100.00
	TPL	0.81	0.84	0.90	1.01	1.06	1.08	1.08	1.10	1.12	1.12	1.00

Tabla A3.6: Segunda parte de dos de la tabla de los porcentajes de las métricas P, R, C y TPL usando varios umbrales para cada palabra usando 4 clasificadores supervisados para el ensamble.

Palabra	Métrica	Umbral										B. L.
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	
argument	R	0.78	3.10	3.88	3.88	4.65	6.98	11.63	18.60	39.53	42.64	41.86
	P	25.00	44.44	55.56	55.56	54.55	36.00	33.33	36.92	42.50	42.64	41.86
	C	3.10	6.98	6.98	6.98	8.53	19.38	34.88	50.39	93.02	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.03	1.00	1.00	1.00
arm	R	80.74	80.74	80.74	80.74	80.74	80.74	80.74	80.74	80.74	80.74	80.74
	P	81.34	81.34	81.34	81.34	81.34	81.34	81.34	81.34	81.34	81.34	80.74
	C	99.26	99.26	99.26	99.26	99.26	99.26	99.26	99.26	99.26	99.26	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.01	1.01	1.00
atmosphere	R	0.00	0.00	2.94	2.94	2.94	7.84	19.61	34.31	50.98	54.90	53.92
	P	0.00	0.00	75.00	75.00	50.00	40.00	37.74	47.30	57.14	54.90	53.92
	C	0.00	0.00	3.92	3.92	5.88	19.61	51.96	72.55	89.22	100.00	100.00
	TPL	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.01	1.02	1.02	1.00
audience	R	64.49	64.49	64.49	64.49	64.49	66.36	67.29	67.29	67.29	67.29	66.36
	P	64.49	64.49	64.49	64.49	64.49	66.36	67.29	67.29	67.29	67.29	66.36
	C	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
bank	R	3.62	3.62	3.62	3.62	3.62	3.62	3.62	4.35	5.07	5.07	4.35
	P	3.76	3.76	3.76	3.76	3.76	3.76	3.73	4.35	5.07	5.07	4.35
	C	96.38	96.38	96.38	96.38	96.38	96.38	97.10	100.00	100.00	100.00	100.00
	TPL	1.01	1.01	1.01	1.01	1.01	1.02	1.04	1.05	1.11	1.11	1.00
degree	R	1.43	2.14	2.86	3.57	3.57	7.86	14.29	14.29	21.43	27.86	13.57
	P	66.67	50.00	57.14	62.50	55.56	40.74	43.48	40.82	33.71	36.11	13.57
	C	2.14	4.29	5.00	5.71	6.43	19.29	32.86	35.00	63.57	77.14	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.02	1.06	1.03	1.04	1.00
difference	R	38.46	38.46	38.46	38.46	38.46	39.23	40.77	41.54	41.54	41.54	41.54
	P	39.06	39.06	39.06	39.06	39.06	39.53	41.09	41.54	41.54	41.54	41.54
	C	98.46	98.46	98.46	98.46	98.46	99.23	99.23	100.00	100.00	100.00	100.00
	TPL	1.02	1.02	1.02	1.02	1.02	1.02	1.03	1.04	1.04	1.04	1.00
difficulty	R	40.91	40.91	40.91	40.91	50.00	36.36	36.36	31.82	31.82	31.82	13.64
	P	75.00	69.23	69.23	69.23	68.75	42.11	38.10	31.82	31.82	31.82	13.64
	C	54.55	59.09	59.09	59.09	72.73	86.36	95.45	100.00	100.00	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
disc	R	0.00	0.00	0.00	0.00	2.75	5.50	22.94	42.20	61.47	61.47	45.87
	P	0.00	0.00	0.00	0.00	30.00	50.00	55.56	60.53	62.62	62.62	45.87
	C	5.50	5.50	5.50	5.50	9.17	11.01	41.28	69.72	98.17	98.17	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.08	1.10	1.10	1.00
image	R	4.00	4.00	6.67	8.00	8.00	17.33	26.67	32.00	30.67	29.33	29.33
	P	6.52	6.67	10.42	12.24	12.50	31.71	48.78	48.98	35.38	29.33	29.33
	C	61.33	60.00	64.00	65.33	64.00	54.67	54.67	65.33	86.67	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.06	1.08	1.07	1.00


Continúa en la tabla A4.8... 

Tabla A3.7: Primera parte de dos de la tabla de los porcentajes de las métricas P, R, C y TPL usando varios umbrales para cada palabra usando 5 clasificadores supervisados para el ensamble.

Palabra	Métrica	Umbral										B. L.
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	
interest	R	0.00	0.00	0.00	0.00	6.32	18.95	42.11	47.37	51.58	51.58	52.63
	P	0.00	0.00	0.00	0.00	46.15	41.86	50.00	48.39	51.58	51.58	52.63
	C	1.05	1.05	1.05	2.11	13.68	45.26	84.21	97.89	100.00	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.02	1.06	1.08	1.05	1.07	1.00
judgment	R	0.00	0.00	0.00	0.00	0.00	31.25	53.13	53.13	53.13	53.13	50.00
	P	0.00	0.00	0.00	0.00	0.00	55.56	53.13	53.13	53.13	53.13	50.00
	C	0.00	0.00	0.00	0.00	6.25	56.25	100.00	100.00	100.00	100.00	100.00
	TPL	0.00	0.00	0.00	0.00	1.00	1.06	1.03	1.03	1.03	1.03	1.00
organization	R	0.00	0.00	0.00	1.82	1.82	5.45	18.18	54.55	74.55	76.36	76.36
	P	0.00	0.00	0.00	100.00	100.00	100.00	71.43	76.92	78.85	77.78	76.36
	C	0.00	0.00	0.00	1.82	1.82	5.45	25.45	70.91	94.55	98.18	100.00
	TPL	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
paper	R	32.35	33.82	33.82	33.82	33.82	33.82	36.76	45.59	47.79	47.79	44.12
	P	34.11	35.66	35.66	35.66	35.66	35.66	38.46	46.27	47.79	47.79	44.12
	C	94.85	94.85	94.85	94.85	94.85	94.85	95.59	98.53	100.00	100.00	100.00
	TPL	1.15	1.16	1.16	1.16	1.16	1.16	1.15	1.10	1.08	1.08	1.00
party	R	10.40	10.40	11.20	11.20	14.40	24.00	30.40	60.80	64.80	64.80	62.40
	P	43.33	43.33	45.16	45.16	51.43	69.77	63.33	64.41	64.80	64.80	62.40
	C	24.00	24.00	24.80	24.80	28.00	34.40	48.00	94.40	100.00	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.06	1.08	1.08	1.00
perfomance	R	27.47	26.37	26.37	26.37	26.37	29.67	35.16	39.56	41.76	40.66	38.46
	P	29.07	27.91	27.91	27.91	27.91	31.40	36.36	39.56	41.76	40.66	38.46
	C	94.51	94.51	94.51	94.51	94.51	94.51	96.70	100.00	100.00	100.00	100.00
	TPL	1.01	1.00	1.00	1.01	1.01	1.01	1.11	1.14	1.12	1.08	1.00
plan	R	2.02	3.03	6.06	7.07	7.07	7.07	7.07	14.14	65.66	76.77	72.73
	P	3.17	4.92	9.38	10.45	10.45	10.45	11.29	40.00	75.58	76.77	72.73
	C	63.64	61.62	64.65	67.68	67.68	67.68	62.63	35.35	86.87	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.06	1.06	1.00
shelter	R	7.83	6.96	6.96	6.96	6.96	6.96	8.70	20.00	28.70	29.57	14.78
	P	21.43	18.60	18.18	18.18	18.18	18.18	21.28	29.49	31.73	30.36	14.78
	C	36.52	37.39	38.26	38.26	38.26	38.26	40.87	67.83	90.43	97.39	100.00
	TPL	1.00	1.00	1.02	1.02	1.02	1.05	1.06	1.13	1.16	1.15	1.00
sort	R	11.93	11.93	11.93	12.84	15.60	34.86	50.46	52.29	55.05	55.05	55.96
	P	30.23	30.23	30.23	31.82	32.08	49.35	53.40	54.29	55.05	55.05	55.96
	C	39.45	39.45	39.45	40.37	48.62	70.64	94.50	96.33	100.00	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Promedio	P	17.18	17.37	17.94	18.25	19.56	24.41	31.89	39.71	48.08	49.39	45.19
	R	27.54	27.35	32.76	38.55	41.15	44.41	44.69	48.07	50.46	50.03	45.19
	C	46.04	46.46	47.17	47.63	50.24	58.55	71.30	81.76	94.83	98.43	100.00
	TPL	0.85	0.85	0.91	0.96	1.01	1.02	1.03	1.05	1.05	1.05	1.00

Tabla A3.8: Segunda parte de dos de la tabla de los porcentajes de las métricas P, R, C y TPL usando varios umbrales para cada palabra usando 5 clasificadores supervisados para el ensamble.

Palabra	Métrica	Umbral										B. L.
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	
argument	R	0.78	3.10	3.88	3.88	4.65	6.98	12.40	17.83	39.53	42.64	41.86
	P	16.67	36.36	45.45	45.45	46.15	36.00	34.78	35.38	42.50	42.64	41.86
	C	4.65	8.53	8.53	8.53	10.08	19.38	35.66	50.39	93.02	100.00	100.00
	TPL	1.00	1.09	1.09	1.09	1.08	1.04	1.02	1.03	1.04	1.04	1.00
arm	R	80.74	80.74	80.74	80.74	80.74	80.74	80.74	80.74	80.74	80.74	80.74
	P	80.74	80.74	80.74	80.74	80.74	80.74	80.74	80.74	80.74	80.74	80.74
	C	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
atmosphere	R	0.00	0.00	2.94	2.94	2.94	7.84	20.59	34.31	51.96	57.84	53.92
	P	0.00	0.00	75.00	75.00	50.00	40.00	38.89	47.30	58.24	57.84	53.92
	C	0.00	0.00	3.92	3.92	5.88	19.61	52.94	72.55	89.22	100.00	100.00
	TPL	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.01	1.04	1.07	1.00
audience	R	68.22	68.22	68.22	68.22	68.22	70.09	71.03	71.03	71.03	71.03	66.36
	P	68.22	68.22	68.22	68.22	68.22	70.09	71.03	71.03	71.03	71.03	66.36
	C	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	TPL	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.00
bank	R	3.62	3.62	3.62	3.62	3.62	3.62	3.62	3.62	4.35	4.35	4.35
	P	3.76	3.76	3.76	3.76	3.76	3.76	3.73	3.62	4.35	4.35	4.35
	C	96.38	96.38	96.38	96.38	96.38	96.38	97.10	100.00	100.00	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.01	1.00
degree	R	1.43	2.14	2.86	3.57	3.57	7.86	14.29	14.29	21.43	27.86	13.57
	P	66.67	50.00	57.14	62.50	55.56	40.74	43.48	40.82	32.97	35.14	13.57
	C	2.14	4.29	5.00	5.71	6.43	19.29	32.86	35.00	65.00	79.29	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.02	1.06	1.03	1.04	1.00
difference	R	40.00	40.00	40.00	40.00	40.00	40.77	43.85	44.62	44.62	44.62	41.54
	P	40.63	40.63	40.63	40.63	40.63	41.09	44.19	44.62	44.62	44.62	41.54
	C	98.46	98.46	98.46	98.46	98.46	99.23	99.23	100.00	100.00	100.00	100.00
	TPL	1.05	1.05	1.05	1.05	1.05	1.05	1.10	1.11	1.11	1.11	1.00
difficulty	R	40.91	40.91	40.91	40.91	50.00	36.36	36.36	31.82	31.82	31.82	13.64
	P	75.00	69.23	69.23	69.23	68.75	42.11	38.10	31.82	31.82	31.82	13.64
	C	54.55	59.09	59.09	59.09	72.73	86.36	95.45	100.00	100.00	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.06	1.16	1.19	1.18	1.18	1.18	1.00
disc	R	0.00	0.00	0.00	0.00	2.75	5.50	22.94	39.45	56.88	56.88	45.87
	P	0.00	0.00	0.00	0.00	30.00	50.00	55.56	56.58	57.94	57.94	45.87
	C	5.50	5.50	5.50	5.50	9.17	11.01	41.28	69.72	98.17	98.17	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.08	1.19	1.19	1.00
image	R	4.00	4.00	6.67	8.00	8.00	17.33	26.67	32.00	30.67	29.33	29.33
	P	6.52	6.67	10.42	12.24	12.50	31.71	48.78	48.98	35.38	29.33	29.33
	C	61.33	60.00	64.00	65.33	64.00	54.67	54.67	65.33	86.67	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.06	1.06	1.05	1.00

Continúa en la tabla A4.10...

Tabla A3.9: Primera parte de dos de la tabla de los porcentajes de las métricas P, R, C y TPL usando varios umbrales para cada palabra usando 6 clasificadores supervisados para el ensamble.

Palabra	Métrica	Umbral										B. L.
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	
interest	R	0.00	0.00	0.00	0.00	6.32	18.95	43.16	50.53	54.74	54.74	52.63
	P	0.00	0.00	0.00	0.00	46.15	41.86	51.25	51.61	54.74	54.74	52.63
	C	1.05	1.05	1.05	2.11	13.68	45.26	84.21	97.89	100.00	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.06	1.09	1.09	1.11	1.00
judgment	R	0.00	0.00	0.00	0.00	0.00	34.38	56.25	56.25	56.25	56.25	50.00
	P	0.00	0.00	0.00	0.00	0.00	61.11	56.25	56.25	56.25	56.25	50.00
	C	0.00	0.00	0.00	0.00	9.38	56.25	100.00	100.00	100.00	100.00	100.00
	TPL	0.00	0.00	0.00	0.00	1.00	1.11	1.09	1.09	1.09	1.09	1.00
organization	R	0.00	0.00	0.00	1.82	1.82	5.45	18.18	54.55	74.55	76.36	76.36
	P	0.00	0.00	0.00	100.00	100.00	100.00	71.43	76.92	78.85	77.78	76.36
	C	0.00	0.00	0.00	1.82	1.82	5.45	25.45	70.91	94.55	98.18	100.00
	TPL	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
paper	R	30.88	30.88	30.88	30.88	30.88	30.88	33.82	44.12	46.32	46.32	44.12
	P	32.56	32.56	32.56	32.56	32.56	32.56	35.38	44.78	46.32	46.32	44.12
	C	94.85	94.85	94.85	94.85	94.85	94.85	95.59	98.53	100.00	100.00	100.00
	TPL	1.08	1.08	1.08	1.08	1.08	1.08	1.08	1.13	1.11	1.11	1.00
party	R	10.40	10.40	11.20	11.20	14.40	24.00	31.20	60.80	62.40	62.40	62.40
	P	41.94	41.94	43.75	43.75	50.00	69.77	65.00	64.41	62.40	62.40	62.40
	C	24.80	24.80	25.60	25.60	28.80	34.40	48.00	94.40	100.00	100.00	100.00
	TPL	1.03	1.03	1.03	1.03	1.03	1.05	1.05	1.03	1.02	1.02	1.00
perfomance	R	26.37	26.37	26.37	26.37	26.37	29.67	32.97	37.36	38.46	39.56	38.46
	P	27.91	27.91	27.91	27.91	27.91	31.40	34.09	37.36	38.46	39.56	38.46
	C	94.51	94.51	94.51	94.51	94.51	94.51	96.70	100.00	100.00	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.01	1.05	1.05	1.09	1.09	1.00
plan	R	2.02	3.03	6.06	7.07	7.07	7.07	7.07	14.14	64.65	74.75	72.73
	P	3.13	4.84	9.23	10.29	10.29	10.29	11.29	38.89	73.56	74.75	72.73
	C	64.65	62.63	65.66	68.69	68.69	68.69	62.63	36.36	87.88	100.00	100.00
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.03	1.03	1.00
shelter	R	9.57	9.57	9.57	9.57	9.57	9.57	10.43	18.26	26.09	26.96	14.78
	P	24.44	23.91	23.40	23.40	23.40	23.40	24.00	26.25	28.85	27.68	14.78
	C	39.13	40.00	40.87	40.87	40.87	40.87	43.48	69.57	90.43	97.39	100.00
	TPL	1.00	1.02	1.04	1.04	1.04	1.04	1.04	1.08	1.12	1.11	1.00
sort	R	11.93	11.93	11.93	12.84	15.60	36.70	54.13	55.96	58.72	58.72	55.96
	P	30.23	30.23	30.23	31.82	32.08	51.95	57.28	58.10	58.72	58.72	55.96
	C	39.45	39.45	39.45	40.37	48.62	70.64	94.50	96.33	100.00	100.00	100.00
	TPL	1.00	1.02	1.02	1.02	1.02	1.05	1.09	1.09	1.08	1.08	1.00
Promedio	R	17.41	17.63	18.20	18.51	19.82	24.94	32.62	40.09	48.17	49.64	45.19
	P	27.28	27.21	32.51	38.29	40.98	45.19	45.54	48.18	50.41	50.19	45.19
	C	46.39	46.82	47.52	47.99	50.75	58.78	71.57	81.95	95.00	98.58	100.00
	TPL	0.85	0.86	0.91	0.97	1.02	1.03	1.05	1.06	1.07	1.07	1.00

Tabla A3.10: Segunda parte de dos de la tabla de los porcentajes de las métricas P, R, C y TPL usando varios umbrales para cada palabra usando 6 clasificadores supervisados para el ensamble.

Lista de figuras

Figura 2.1: Ejemplo del algoritmo original de Lesk:	20
Figura 2.2: El problema de Lesk: la cantidad de combinaciones posibles.....	20
Figura 2.3: Ejemplo del algoritmo simplificado de Lesk: Desambiguación del sentido de la palabra “PINE”.....	21
Figura 2.4: Esquema general de los métodos supervisados.	23
Figura 2.5: Esquema general de los métodos no supervisados.	27
Figura 2.6: Problema de clasificación linealmente separable	41
Figura 2.7: Mapeo de datos no lineales a un espacio de mayor dimensionalidad.	41
Figura 3.1: Esquema general del método previo.....	47
Figura 3.2: Ejemplo de desambiguación de la palabra “gato” con el método propuesto.	52
Figura 3.3: Ejemplo del algoritmo propuesto	55
Figura 3.4: Gráfica de Precisión/Recuerdo/Cobertura vs Umbral.	59
Figura 3.5: Gráfica de TPL vs Umbral.....	59
Figura 4.1: Estructura del ensamble del primer enfoque.	64
Figura 4.2: Esquema del ensamble usando el clasificador NGD para filtrar los votos.....	66

Lista de Tablas

Tabla 2.1: Ejemplo de resultados de la búsqueda en Google.....	17
Tabla 2.2: Definiciones de los sentidos de la palabra en inglés “plant” obtenidas desde WordNet 2.0.	19
Tabla 2.3: Algoritmo original de Lesk.....	19
Tabla 2.4: Algoritmo simplificado de Lesk.	21
Tabla 2.5: Ejemplos de synsets y definiciones de la palabra “plant”	29
Tabla 2.6: Ejemplos de synsets hiperonímicos y sus definiciones sobre el segundo sentido de la palabra “plant”.....	30
Tabla 2.7: Extracto del corpus de la tarea All-Words en inglés.....	31
Tabla 2.8: Extracto del corpus de la tarea English Lexical Simple en inglés.....	32
Tabla 2.9: Inventario de palabras que contiene SENSEVAL-3 ELS.....	33
Tabla 2.10: Algoritmo de ID3.....	37
Tabla 2.11: Algoritmo de k vecinos más cercanos.	39
Tabla 3.1: Porcentajes de precisión/recuerdo del experimento previo.	49
Tabla 3.2: Valores que devuelve NGD en ciertas condiciones.	51
Tabla 3.3: Algoritmo para el clasificador NGD.....	54
Tabla 3.4: Palabras en inglés que fueron utilizadas para este experimento.....	56
Tabla 3.5: Descripción de las métricas utilizadas para evaluar el algoritmo.....	57
Tabla 3.6: Porcentajes de los resultados para cada palabra con umbrales de 0.1 al 0.5.....	58
Tabla 3.7: Porcentajes de los resultados para cada palabra con umbrales de 0.6 al 1.0.....	58
Tabla 4.1: Palabras en inglés que fueron utilizadas para este experimento.....	62
Tabla 4.2: Porcentajes promedios de la precisión y clasificadores usados.....	63
Tabla 4.3: Tabla de los porcentajes de precisión (P) y el tamaño promedio de la lista de sentidos propuestos (TPL) obtenidos usando el primer enfoco.....	65

Tabla 4.4: Tabla de porcentajes de precisión (P), recuerdo (R), cobertura (c) y el tamaño promedio de la lista de sentidos propuestos (“TPL”) obtenidos usando diferentes umbral para el segundo enfoque.....	66
Tabla A1.1: Estadísticas del conjunto de datos.....	71
Tabla A1.2: Porcentajes de precisión de los clasificadores supervisados.....	72
Tabla A2.1: Tabla de los porcentajes de precisión (P) y tamaño promedio de la lista (TPL) usando varios umbrales para cada palabra usando 2 clasificadores supervisados para el ensamble.	74
Tabla A2.2: Tabla de los porcentajes de precisión (P) y tamaño promedio de la lista (TPL) usando varios umbrales para cada palabra usando 3 clasificadores supervisados para el ensamble.	75
Tabla A2.3: Tabla de los porcentajes de precisión (P) y tamaño promedio de la lista (TPL) usando varios umbrales para cada palabra usando 4 clasificadores supervisados para el ensamble.	76
Tabla A2.4: Tabla de los porcentajes de precisión (P) y tamaño promedio de la lista (TPL) usando varios umbrales para cada palabra usando 5 clasificadores supervisados para el ensamble.	77
Tabla A2.5: Tabla de los porcentajes de precisión (P) y tamaño promedio de la lista (TPL) usando varios umbrales para cada palabra usando 6 clasificadores supervisados para el ensamble.	78
Tabla A3.1: Primera parte de dos de la tabla de los porcentajes de las métricas P, R, C y TPL usando varios umbrales para cada palabra usando 2 clasificadores supervisados para el ensamble.	80
Tabla A3.2: Segunda parte de dos de la tabla de los porcentajes de las métricas P, R, C y TPL usando varios umbrales para cada palabra usando 2 clasificadores supervisados para el ensamble.	81
Tabla A3.3: Primera parte de dos de la tabla de los porcentajes de las métricas P, R, C y TPL usando varios umbrales para cada palabra usando 3 clasificadores supervisados para el ensamble.	82
Tabla A3.4: Segunda parte de dos de la tabla de los porcentajes de las métricas P, R, C y TPL usando varios umbrales para cada palabra usando 3 clasificadores supervisados para el ensamble.	83

Tabla A3.5: Primera parte de dos de la tabla de los porcentajes de las métricas P, R, C y TPL usando varios umbrales para cada palabra usando 4 clasificadores supervisados para el ensamble.	84
Tabla A3.6: Segunda parte de dos de la tabla de los porcentajes de las métricas P, R, C y TPL usando varios umbrales para cada palabra usando 4 clasificadores supervisados para el ensamble.	85
Tabla A3.7: Primera parte de dos de la tabla de los porcentajes de las métricas P, R, C y TPL usando varios umbrales para cada palabra usando 5 clasificadores supervisados para el ensamble.	86
Tabla A3.8: Segunda parte de dos de la tabla de los porcentajes de las métricas P, R, C y TPL usando varios umbrales para cada palabra usando 5 clasificadores supervisados para el ensamble.	87
Tabla A3.9: Primera parte de dos de la tabla de los porcentajes de las métricas P, R, C y TPL usando varios umbrales para cada palabra usando 6 clasificadores supervisados para el ensamble.	88
Tabla A3.10: Segunda parte de dos de la tabla de los porcentajes de las métricas P, R, C y TPL usando varios umbrales para cada palabra usando 6 clasificadores supervisados para el ensamble.	89

Referencias

- [1] Y. Wilks y R. Catizone, "Can we make information extraction more adaptive?" University of Sheffield, Computer Science Department, Memoranda en *Computer and Cognitive Science*, 2000. <http://www.dcs.shef.ac.uk/~yorick/papers/doc/doc.html>
- [2] R. Mihalcea, y P. Edmonds, *Proceedings of Senseval-3: The 3rd. Int. Workshop on the Evaluation of Systems for the Semantic Analysis of Text*. Barcelona, Spain, 2004. <http://acl.ldc.upenn.edu/acl2004/senseval/>
- [3] T. Mitchell. *Machine Learning*. McGraw-Hill, 1997.
- [4] G. Paliouras, V. Karkaletsis, I. Androutsopoulos y C. D. Spyropoulos, "Learning rules for large-vocabulary Word Sense Disambiguation: a comparison of various classifiers," en *Proceedings of the 2nd International Conference on Natural Language Processing*, Patra, Grecia, 2000.
- [5] J. Gonzalo, F. Verdejo e I. Chugar, "The Web as a resource for WSD," en *1st MEANING Workshop*, España, 2003.
- [6] E. Agirre y D. Martínez, "Exploring automatic Word Sense Disambiguation with decision lists and the Web," en *Proceedings of the COLING 2000 Workshop on Semantic Annotation and Intelligent Content*, 2000.
- [7] F. Heylighen, "Mining associative meanings from the Web: from word disambiguation to the global brain," en *Proceedings of the International Colloquium: Trends in Special Language & Language Technology*, Standard Editions, pp. 15-44.
- [8] M. Stevenson, *Word Sense Disambiguation. The Case for Combinations of Knowledge Sources*. CSLI Publications, 2003.
- [9] W. Weaver, *Translation in Machine Translation of Languages: Fourteen Essays*. The MIT Press, Cambridge, Massachusetts, 1949, pp. 15-23.
- [10] N. Ide y J. Véronis, "Word Sense Disambiguation: the state of the art," en *Computational Linguistics*, Vol. 24, 1998, pp. 1-40.
- [11] B. Hillel, "The present status of automatic translation of languages," en *Advances in Computers*, Vol. 1, Academic Press, New York, N.Y., 1960, pp. 91-163.

- [12] R. Mihalcea y T. Pedersen, "Advances on Word Sense Disambiguation," Notas del Tutorial, *IBERAMIA-2004*, Puebla, México, 2004.
- [13] R. Mihalcea y D. Moldovan, "Automatic acquisition of sense tagged corpora," en *Proceedings of Florida Artificial Intelligence Research Society Conference (FLAIRS 1999)*, Orlando, FL, May 1999.
- [14] A. Blum y T. Mitchell, "Combining labeled and unlabeled data with co-training," en *Proceedings of the 11th Annual Conference on Computational Learning Theory*, pp. 92-100, 1998.
- [15] Y. Wilks, D. Fass, C.-M. Guo, J. McDonald, T. Plate, y B. Sator, "Providing machine tractable dictionary tools," en *Machine Translation*, Vol. 5, 1990, pp. 99-154.
- [16] R. Cilibrasi y P. Vitanyi, "Automatic meaning discovery using Google," Manuscrito, CWI, 2004, <http://arxiv.org/abs/cs.CL/0412098>.
- [17] H. Bennett, P. Gács, M. Li, P. M. B. Vitányi y W. H. Zurek. "Information distance," en *EEETIT: IEEE Transactions on Information Theory*, 1997.
- [18] S. Weiss, "Learning to disambiguate," en *Information Storage and Retrieval*. Vol. 9, 1973, pp. 33-41.
- [19] E. F. Kelley y P. J. Stone, *Computer Recognition of English Word Senses*. Holanda – Norte, Amsterdam. 1975.
- [20] T. Pedersen, "Learning probabilistic models of Word Sense Disambiguation," Tesis doctoral, Universidad Metodista del Sur, Dallas, 1998.
- [21] T. Pedersen y R. Bruce, "Distinguishing word senses in untagged text," en *Proceeding of the Second Conference on Empirical Methods in Natural Language Processing*, Providence, RI, 1997, pp. 197-207.
- [22] D. Yarowsky, "Unsupervised word-sense disambiguation rivaling supervised methods," en *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (ACL-95)*, Cambridge, MA, 1995, pp. 189-196.
- [23] E. Black, "An experiment in computational discrimination of English word senses," en *IBM Journal of Research and Development*, Vol. 32, No. 2, 1998, pp. 185-194.
- [24] W. A. Gale, K. W. Church y D. Yarowsky, "A method for disambiguating word senses in a large corpus," en *Computers and the Humanities*, Vol. 26, pp. 415-439.
- [25] C. Leacock, G. Towell y E. M. Voorhees, "Corpus-based statistical sense resolution," en *Proceedings of the ARPA Human languages Technology Workshop*, 1993.

- [26] G. Towell y E. M. Voorhees, “Disambiguating highly ambiguous words. Computational Linguistics,” en *Computational Linguistics*, Vol. 24, No. 1, 1998, pp. 125-146.
- [27] D. Yarowsky, “Decision list for lexical ambiguity resolution: Application to accent restoration in spanish and French,” en *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 1994, pp. 88-95.
- [28] G. Rigau, J. Atserias y E. Agirre “Combining unsupervised lexical knowledge methods for word sense disambiguation,” en *35th Meeting of the Association for Computational Linguistic (ACL '97)* y en *8th Meeting of the European Chapter of the Association for Computational Linguistics (EACL '97)*, Madrid, España, 1997.
- [29] C. Manning y H. Schütze, *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.
- [30] A. Purandare y T. Pedersen, “SenseClusters - Finding clusters that represent word senses,” en *Proceedings of the Nineteenth National Conference on Artificial Intelligence (AAAI-04)*, San José, California y en *Proceedings of 5th Annual Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL-04)*, Boston, Massachusetts, 2004.
- [31] A. Kulkarni y T. Pedersen, “SenseClusters: unsupervised clustering and labeling of similar contexts,” en the *Proceedings of the Demonstration and Interactive Poster Session of the 43rd Annual Meeting of the Association for Computational Linguistics*, Ann Arbor, Michigan, 2005.
- [32] A. Kulkarni, “*Word Sense Discrimination by clustering similarity contexts*,” Tesis de Maestría, Departamento de Ciencias Computacionales, Universidad de Minnesota, Duluth, Agosto, 2004.
- [33] M. Lesk, “Automated sense disambiguation using machine-readable dictionaries: how to tell a pine cone from an ice cream cone,” en *Proceedings of the 1986 SIGDOG Conference*, Toronto, Canadá, 1986, pp. 24-26.
- [34] Y. A. Wilks, D. Fass, C.-M. Guo, J. E. MacDonald, T. Plate y B. A. Slator, “Providing machine tractable dictionary tools,” en *Semantics and the Lexicon*, Pustejovsky, James (ED), MIT Press, Cambridge, Massachusetts, 1990.
- [35] A. K. Luk, “Statistical sense disambiguation with relatively small corpora using dictionary definitions,” en *Proceedings of the 33rd Meeting of the Association for Computational Linguistics (ACL)*, Cambridge, Massachusetts, 1995, pp. 181-188.
- [36] E. Agirre, O. Ansa, E. Hovy y D. Martínez, “Enriching very large ontologies using the WWW,” en *Proceedings of the Ontology Learning Workshop, ECAI*, Berlin, Alemania, 2000.
- [37] J. Gonzalo, F. Verdejo e I. Chugar, “The Web as a Resource for WSD,” en *1st MEANING Workshop*, España, 2003.

- [38] P. Rosso, M. Montes-y-Gómez, D. Buscaldi, A. Pancardo-Rodríguez y L. Villaseñor Pineda. “Two Web-based approaches for noun sense disambiguation,” en *Computational Linguistics and Intelligent Text Processing (CICLing 2005)*, Ciudad de México, México, 2005
- [39] K. Aas y L. Eikvil, “Text categorization: A survey,” Norwegian Computing Center, reporte técnico, 1999.
- [40] T. Joachims, “Text categorization with support vector machines: learning with many relevant features,” en *Proceedings of the Tenth European Conference on Machine Learning (ECML’98), Lecture Notes in Computer Science*, No. 1398, 1998, pp. 137-142.
- [41] D. Lewis y M. Ringuette, “A comparison of two learning algorithms for text classification,” en *Proceedings of the 3rd Annual Symposium on Document Analysis and Information Retrieval*, 1994, pp. 81-93.
- [42] D. Lewis, “Naive (Bayes) at forty: The independence assumption in information retrieval,” en *Proceedings of 10th European Conference on Machine Learning*, Springer Verlag, 1998, pp. 4-15.
- [43] J. R. Quinlan, “Induction of decision trees,” en *Machine Learning*, Vol. 1, 1986, pp. 81-106.
- [44] J. R. Quinlan, *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 1993.
- [45] W. Cohen y H. Hirsh, “Joins that generalize: text classification using WHIRL,” en *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining (KDD’98)*, AAAI Press, 1998, pp. 169 –173.
- [46] W. Cohen y Y. Singer, “Context-sensitive learning methods for text categorization,” en *ACM Transactions on Information Systems*, Vol. 17, No. 2, 1999, pp.141 –173.
- [47] Kilgarriff y J. Rosenzweig, “Framework and results for English SENSEVAL,” en *Computers and the Humanities*, Vol. 34, No. 1-2, 2000, pp 15-48.
- [48] F. Vasilescu, P. Langlais y G. Lapalme, “Evaluating variants of the Lesk approach for disambiguating words”, en *Proceedings of LREC 2004*, 2004, pp. 633-636.
- [49] J. R. Hobbs, “World knowledge and word meaning,” en *Proceedings of the Third Workshop on Theoretical Issues in Natural Language Processing (TINLAP-3)*, Las Cruces, Nuevo México, 1987, pp. 20-25.
- [50] A. Miller, R. Beckwith, C. Fellbaum, D. Gross y K. J. Miller, “Introduction to WordNet: An on-line lexical database,” en *International Journal of Lexicography*, Vol. 3, No.4, 1990, pp.235-312.
- [51] C. Fellbaum, *WordNet: an Electronic Lexical Database*. The MIT Press, 1998.

- [52] S. M. Harabagiu, G. A. Miller y D.I. Moldovan, “WordNet 2 - A Morphologically and Semantically Enhanced Resource,” en *Proceedings of the SIGLEX Workshop*, 1999.
- [53] A. Kilgarriff, “SENSEVAL: An exercise in evaluating Word Sense Disambiguation programs,” en *Proceedings of LREC*, Vol. 1, Granada, España, 1998, pp. 581-588. <http://citeseer.ist.psu.edu/kilgarriff98senseval.html>
- [54] A. Kilgarriff y M. Palmer, Eds., “SENSEVAL. Evaluating Word Sense Disambiguation programs,” en *Computer and the Humanities*, Vol. 34, 2000.
- [55] J. Preiss y D. Yarowsky, Eds., *Proceedings of SENSEVAL-2*, Association for Computational Linguistics Workshop, Toulouse, France, 2001.
- [56] M. Snyder y M. Palmer, “The English all-words task,” en *Proceedings of Senseval-3: The Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, 2004, pp. 41-43.
- [57] R. Mihalcea, T. Chklovski y A. Kilgarriff, “The Senseval-3 english lexical sample task,” en *Proceedings of Senseval-3: The Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, 2004, pp. 25-28.
- [58] T. Chklovski y R. Mihalcea, “Building a sense tagged corpus with Open Mind Word Expert,” en *Proceedings of the ACL 2002 Workshop on “Word Sense Disambiguation: Recent Successes and Future Directions”*, Philadelphia, 2002.
- [59] W. Gale, K. W. Church y D. Yarowsky, “Estimating upper and lower bounds on the performance of word-sense disambiguation programs,” en *Proceedings of the 30th annual meeting on Association for Computational Linguistics*, 1992, pp. 249-256.
- [60] T. Joachims, “A statistical learning model of text classification with support vector machines,” en *Proceedings of the 24th ACM International Conference on Research and Development in Information Retrieval*, ACM Press, 2001.
- [61] S. Dumais, J. Platt, D. Heckerman, y M. Sahami, “Inductive learning algorithms and representations for text categorization,” en *Proceedings of the Seventh International Conference on Information Retrieval and Knowledge Management (ACM-CIKM’98)*, 1998, pp. 148-155.
- [62] M. Hearst, B. Schölkopf, S. Dumais, E. Osuna, y J. Platt, “Trends and controversies - Support vector machines,” *IEEE Intelligent systems*, Vol. 13, No. 4, pp. 18-28, 1998.
- [63] A. Pancardo-Rodríguez, M. Montes-y-Gómez, P. Rosso, D. Buscaldi y L. Villaseñor-Pineda, “Desambiguación léxica de sustantivos usando la Web,” en *Workshop on Lexical Resources and the Web for Word Sense Disambiguation. IX Ibero-American Conference on Artificial Intelligence IBERAMIA 2004*, 2004, pp. 118-122.

- [64] A. Pancardo-Rodríguez, M. Montes-y-Gómez, L. Villaseñor-Pineda y P. Rosso, “A mapping between Classifiers and Training Conditions for WSD,” en *Computational Linguistics and Intelligent Text Processing (CICLing 2005)*, Ciudad de México, México, 2005, pp. 246-249.
- [65] J. C. Platt, “Fast training of SVMs using sequential minimal optimization”, en *Advances in Kernel Methods - Support Vector Learning*, B. Schölkopf, C. Burges, y A. Smola, Eds. MIT Press, Cambridge, Mass., 1998.
- [66] E. Frank, M. Hall, y B. Pfahringer, “Locally Weighted Naive Bayes,” en *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, Acapulco, Mexico. Morgan Kaufmann, 2003.

Publicaciones

A. Pancardo-Rodríguez, M. Montes-y-Gómez, P. Rosso, D. Buscaldi y L. Villaseñor-Pineda, “Desambiguación léxica de sustantivos usando la Web,” en *Workshop on Lexical Resources and the Web for Word Sense Disambiguation. IX Ibero-American Conference on Artificial Intelligence IBERAMIA 2004*, 2004, pp. 246-249.

A. Pancardo-Rodríguez, M. Montes-y-Gómez, L. Villaseñor-Pineda y P. Rosso, “A mapping between classifiers and training conditions for WSD,” en *Computational Linguistics and Intelligent Text Processing (CICLing 2005)*, Ciudad de México, México, 2005, pp. 246-249.

P. Rosso, M. Montes-y-Gómez, D. Buscaldi, A. Pancardo-Rodríguez y L. Villaseñor Pineda, “Two Web-based approaches for noun sense disambiguation,” en *Computational Linguistics and Intelligent Text Processing (CICLing 2005)*, Ciudad de México, México, 2005, pp. 267-279.