Universidad Central “Marta Abreu” de Las Villas

Facultad de Matemática, Física y Computación

Trabajo para optar por el Título Académico

Máster en Ciencia de la Computación

Detección de la polaridad de las opiniones basada en nuevos recursos léxicos

Autor:

Lic. Mario Alberto Amores Fernández

Tutora:

Dra. Leticia Arco García

2016

Dedicatoria

Agradecimientos

Resumen

Las opiniones son una parte importante en la vida de los seres humanos. Para extraer el sentimiento acerca de objetos, productos o servicios, se necesitan sistemas de minería de opinión automatizados. La herramienta PosNeg Opinion detecta de manera no supervisada la polaridad de las opiniones basándose en recursos léxicos, de ahí que sea sensible a la calidad de éstos, que en su mayoría están concebidos para el idioma Inglés, aquellos que fueron anotados automáticamente tienen muchos errores, los anotados manualmente recogen muy pocos términos, y el formato que presentan muchas veces limita la interacción entre ellos. Por otro lado, las opiniones generalmente presentan varios problemas que aún no son tratados eficazmente por las herramientas existentes. El objetivo de esta investigación consiste en desarrollar un sistema, a partir de las características de PosNeg Opinion, para la detección no supervisada de la polaridad de las opiniones a partir del empleo de nuevos recursos léxicos y que sea capaz de tratar la mayoría de los problemas presentes en las opiniones. Los resultados obtenidos son: la creación de los recursos SentiWordNet 4.0 y 4.1 para el idioma Inglés y SpanishSentiWordNet que es pionero en la puntuación de términos en español, los dos esquemas para la detección no supervisada de la polaridad de las opiniones, los recursos que permiten el manejo de jergas, emoticonos, palabras modificadoras y la negación, la herramienta PosNeg Opinion 3.0 que implementa el esquema finalmente propuesto auxiliándose de la biblioteca desarrollada PolarityDetection obteniendo satisfactorios valores de Exactitud, Precisión, Recall y F1.

Abstract

Opinions are an important part in the life of human beings. To remove the feeling about objects, products or services, automated mining systems review are required. The PosNeg Opinion tool detects unsupervised polarity of opinions based on lexical resources, and is therefore sensitive to the quality of these, most of which are designed for the English language, those were scored automatically have many errors, annotated manually collect very few terms, and format have often limits the interaction between them. On the other hand, opinions generally have several problems that are not effectively treated by existing tools. The objective of this research is to develop a system, based on the characteristics of PosNeg Opinion for detection unsupervised polarity of opinions from the use of new lexical resources and be able to treat most problems present in the opinions. The results obtained are: the creation of SentiWordNet 4.0 and 4.1 for the English and SpanishSentiWordNet language is a pioneer in the score of terms in Spanish resources, the two schemes for detection unsupervised polarity of opinions, resources that allow finally proposed management jargons, emoticons, modifying words and denial, PosNeg Opinion 3.0 tool that implements the scheme using for this library developed with satisfactory values PolarityDetection Accuracy, Precision, Recall and F1.

Tabla de contenidos

Introducción

Las personas al prepararse para tomar alguna decisión siempre tienen en cuenta, además de su propia experiencia, lo que otras personas puedan aportar, por lo que las opiniones son una parte importante en la vida de los seres humanos. La creciente popularidad de sitios ricos en opiniones, como blogs, redes sociales y sitios de recomendaciones, abre nuevas oportunidades y retos al poderse usar diferentes técnicas para entender y extraer las opiniones de otros.

Las opiniones son los estados subjetivos que reflejan los sentimientos y la percepción de una persona sobre un suceso o un objeto (Kaur & Duhan 2015). Según (Liu 2012), una opinión es una emoción sobre una entidad o un aspecto de la entidad expresado por un usuario, esta entidad puede ser un producto, persona, evento, organización, o tópico, y aparecer en una jerarquía de componentes y subcomponentes; cada nodo representa un componente y se asocia con un conjunto de atributos del componente. Una opinión puede expresarse en cualquier nodo o atributo del nodo. Por simplicidad, en las publicaciones de este tema se utiliza el término aspecto (característica) para representar ambos: componentes y atributos. Una opinión también ha sido definida matemáticamente como una quíntupla (Liu 2012), donde se incluye el nombre de una entidad, los aspectos de la entidad, los sentimientos de cada uno de los aspectos de la entidad, el titular (usuario) de la opinión, y el tiempo en que la opinión fue expresada. El sentimiento es positivo, negativo o neutro, o se expresa con distinta fuerza o niveles de intensidad; por ejemplo, de 1 a 5 estrellas como es usado en la mayoría de sitios de críticas que se encuentran en la Web.

Las opiniones son una información muy valiosa tanto desde el punto de vista de un usuario particular, instituciones y organizaciones políticas, así como desde el punto de vista del departamento de marketing de una empresa. A un usuario le es interesante antes de realizar una compra comprobar qué opinan otros usuarios sobre dicha marca, producto o servicio. En el entorno empresarial tiene una especial importancia debido a que para una empresa resulta vital conocer lo que piensan los usuarios sobre las virtudes y debilidades de sus productos y servicios, así como a qué grupos llegan. Para realizar este estudio, necesitarían tener analistas que se dediquen a obtener estadísticas sobre las opiniones positivas y negativas de los productos y obtener conclusiones que le permitan a la empresa tomar las decisiones adecuadas (bajar los precios, cambiar componentes, rediseñar el producto, etc.). Esta tarea puede resultar tediosa, pues se tendrían que procesar miles de opiniones, extraer cuales son los rasgos

positivos y cuales los rasgos negativos de un producto, entre otros análisis. Por tanto, con la llegada de la web 2.0, enormes volúmenes de textos con opiniones están disponibles en la web. Para extraer el sentimiento acerca de un objeto, producto o servicio de esta enorme red, se necesitan sistemas de minería de opinión automatizados.

Existen varias definiciones de minería de opinión o análisis de sentimientos (Kaur & Duhan 2015; Liu 2012; Vasantharaj et al. 2015; Zuva & Zuva 2012; Sharma & Chitre 2014; Angulakshmi & ManickaChezian 2014; Singh et al. 2010). Según (Liu 2012), el análisis de sentimientos o la minería de opinión es el estudio computacional de las opiniones, valoraciones, actitudes y emociones expresadas por los usuarios hacia entidades, personas, temas, eventos, productos y sus atributos. Bing Liu en (Liu 2012), define la minería de opinión como un tipo de procesamiento del lenguaje natural para el seguimiento del estado de ánimo de la opinión pública sobre un tema en particular; por tanto, implica la construcción de un sistema para recoger y examinar opiniones sobre un aspecto o tópico en entradas de blog, comentarios, críticas o tweets. En (Vasantharaj et al. 2015) se ha definido la minería de opinión de la siguiente manera: Dado un conjunto de documentos textuales D que contiene las opiniones (o sentimientos) en relación con un objeto, la minería de opinión pretende extraer atributos y elementos de lo que se ha comentado en cada documento d que pertenece a D y conocer si los comentarios son o no positivos, negativos o neutros.

Aunque en la actualidad se hace referencia indistintamente a minería de opinión y análisis de sentimiento, estos términos tuvieron orígenes diferentes según (Zuva & Zuva 2012). Por una parte, la minería de opinión se originó en la comunidad de la recuperación de información, y su objetivo es extraer y procesar aún más la opinión de los usuarios sobre los productos, películas u otras entidades. El análisis de sentimientos, por otra parte, se formuló inicialmente como una tarea del procesamiento del lenguaje natural de los sentimientos expresados en los textos (Zuva & Zuva 2012). Sin embargo, estos dos problemas son similares en su propia esencia y caen bajo el ámbito del análisis de la subjetividad. El análisis de la subjetividad implica diversos métodos y técnicas que se originan a partir de la recuperación de información, la inteligencia artificial y procesamiento del lenguaje natural. Esta confluencia de diferentes enfoques se explica por la naturaleza de los datos que están siendo procesados y los requisitos de la aplicación (Liu 2012).

En (Kaur & Duhan 2015) definen la minería de opinión como la técnica para extraer la información subjetiva del texto y determinar la polaridad global de la opinión. Sin embargo, existen oraciones objetivas que expresan opinión y oraciones subjetivas que no expresan ninguna opinión. Por otro lado, no solo es deseado determinar la polaridad global de la opinión, puede ser interesante en la toma de decisiones calcular la polaridad local, por aspectos o tópicos. De ahí que esta definición ofrece una visión limita del análisis de sentimiento.

Varios autores coinciden en que la minería opinión es una disciplina muy difícil y prometedora (Liu 2012; Sharma & Chitre 2014; Angulakshmi & ManickaChezian 2014). Se define como una intersección de las técnicas de lingüística computacional y de recuperación de información para hacer frente a las opiniones expresadas en un documento. Su objetivo principal es la solución de los problemas relacionados con las opiniones acerca de los productos, criticas de películas, los mensajes de grupos de noticias de un político, sitios de revisión, etc (Sharma & Chitre 2014). Una forma de realizar el análisis de sentimiento consiste en transitar por cuatro etapas: la extracción de datos de diversas fuentes, la clasificación del texto, el agrupamiento y la evaluación con valores positivo o negativo (Sharma & Chitre 2014).

Un elemento importante en el análisis de sentimiento es el resumen de las opiniones para contribuir a la toma de decisiones. Una definición que incorpora explícitamente el resumen de las opiniones es la presentada en (Singh et al. 2010), donde se define la minería de opinión como un área de la minería de textos encargada de extraer los conjuntos de datos de opiniones y resumirlos de forma comprensible para el usuario final. Por tanto, involucra el procesamiento del lenguaje natural y la tarea de extracción de las opiniones para averiguar la polaridad de cualquier información de los consumidores de productos. En (Angulakshmi & ManickaChezian 2014) también se hace referencia explícitamente al resumen de las opiniones y definen la minería de opinión como la secuencia de tres pasos principales: (1) recuperación de la opinión, (2) clasificación de la opinión y (3) la creación del resumen.

Existen varias definiciones de minería de opinión (Kaur & Duhan 2015; Liu 2012; Vasantharaj et al. 2015; Zuva & Zuva 2012; Sharma & Chitre 2014; Angulakshmi & ManickaChezian 2014; Singh et al. 2010), unas más generales, otras muy específicas, pero todas tienen como elemento común la aplicación de técnicas del procesamiento del lenguaje

natural, lingüística computacional y minería de textos, que tienen como objetivo la extracción de información subjetiva a partir de contenidos generados por los usuarios, como puedan ser comentarios en blogs, evaluaciones de productos o servicios, y respuestas a una encuesta (Liu 2012), y de esta forma contribuir a la toma de decisiones.

Una de las principales tareas de la minería de opinión es la clasificación de la polaridad de la opinión, que consiste en determinar si la opinión es positiva o negativa con respecto a la entidad a la que se esté refiriendo, que puede ser una persona, un producto, una temática, etc.

La mayoría de las aproximaciones existentes para la determinación de la polaridad aplican técnicas de aprendizaje supervisado. Estas técnicas, aun cuando hasta el momento obtienen mejores resultados, presentan varias desventajas, entre ellas: están sujetas al sobreentrenamiento y son altamente dependientes de la calidad, tamaño y dominio de los datos de entrenamiento. Por su parte, las aproximaciones no supervisadas no requieren de la existencia de colecciones previamente clasificadas, son menos dependientes del lenguaje pero se basan en recursos externos, y actualmente existen pocos recursos disponibles y son mayormente dependientes del idioma, sobre todo del idioma Inglés.

El procesamiento automático de opiniones no es una tarea sencilla. Algunos de los problemas presentes en el tratamiento de las opiniones son: el uso de lenguaje informal, las abreviaturas, los errores ortográficos y tipográficos, el lenguaje irónico y sarcástico, el nivel de conocimiento del lenguaje, el nivel cultural, entre otros. Estos problemas, en comparación con el procesamiento de documentos en otras tareas de la minería de textos, imponen una mayor dificultad a la minería de opiniones.

Teniendo en cuenta la actualidad e importancia de la minería de opinión, así como los intereses expresados por algunos usuarios en el tema, especialistas del laboratorio de Inteligencia Artificial del Centro de Estudios Informáticos (CEI) de la Universidad Central “Marta Abreu” de Las Villas (UCLV), se dieron a la tarea de desarrollar la aplicación PosNeg Opinion para la detección no supervisada de la polaridad de opiniones en Inglés y Español (Amores 2013; Amores, Arco & Artiles 2015a). La aplicación implementa un esquema general compuesto por cinco etapas. PosNeg Opinion permite que el usuario analice un gran cúmulo de opiniones de manera sencilla. Esta aplicación puede utilizarse como un módulo para una aplicación más general de minería de opinión pues resuelve una de las fases de este proceso y es fácilmente reutilizable, además, se puede comunicar con otras aplicaciones

mediante ficheros XML. Fue desarrollada completamente en JAVA, por lo que es multiplataforma. Necesita como entrada un fichero XML con todas las opiniones a analizar y como salida muestra cuántas fueron positivas y cuántas negativas. A petición del usuario también retorna el porcentaje de las opiniones negativas y positivas así como una lista con las opiniones negativas y otra con las opiniones positivas, además de destacar cuáles fueron las opiniones de mayor puntuación en cada caso (positivas/negativas). Desafortunadamente, PosNeg Opinion es sensible a la calidad de los recursos léxicos que utiliza en su análisis y los recursos léxicos disponibles en internet tienen algunas limitaciones, tales como términos incorrectamente anotados. Por otra parte, PosNeg Opinion no trata algunos de los problemas presentes en las opiniones, limitándose de esta forma la calidad de la determinación de la polaridad y consecuentemente la futura toma de decisiones a partir de los resultados que la herramienta ofrece. Estas desventajas de PosNeg Opinion constituyen una problemática a la cual aún no se le ha dado respuesta, lo cual justifica el **planteamiento del problema de investigación** siguiente:

La herramienta PosNeg Opinion detecta de manera no supervisada la polaridad de las opiniones basándose en recursos léxicos, de ahí que sea sensible a la calidad de éstos, que en su mayoría están concebidos para el idioma Inglés, aquellos que fueron anotados automáticamente tienen muchos errores, los anotados manualmente recogen muy pocos términos, y el formato que presentan muchas veces limita la interacción entre ellos. Por otro lado, las opiniones generalmente presentan varios problemas que aún no son tratados eficazmente por las herramientas existentes.

El **objetivo general** de esta investigación consiste en desarrollar un sistema, a partir de las características de PosNeg Opinion, para la detección no supervisada de la polaridad de las opiniones a partir del empleo de nuevos recursos léxicos y que sea capaz de tratar la mayoría de los problemas presentes en las opiniones. Este se desglosa en los siguientes **objetivos específicos**:

- Identificar, a partir del estudio de las principales herramientas existentes que contribuyen a la detección de la polaridad de las opiniones, aquellas que pueden ser sustituidas por herramientas más eficientes y cuáles se pueden actualizar.
- Crear recursos léxicos que permitan obtener mejores resultados en la detección de la polaridad de las opiniones.

- Diseñar un esquema general que establezca los pasos a seguir para minar las opiniones de manera no supervisada y utilizando los recursos léxicos creados.
- Desarrollar soluciones computacionales que permitan resolver varias de las problemáticas existentes en las opiniones, contribuyendo a un mejor procesamiento de las mismas y consecuentemente mayor efectividad en el cálculo de la polaridad.
- Aplicar de manera integrada aquellas herramientas creadas, las soluciones obtenidas para manejar algunas de las problemáticas de las opiniones, así como otras herramientas disponibles para las tareas del análisis de sentimientos, que faciliten la implementación del esquema propuesto.

Las **preguntas de investigación** planteadas son:

- ¿Cuáles de las herramientas que permiten realizar la detección de la polaridad de las opiniones pueden ser sustituidas por otras más eficientes y cuáles requieren ser actualizadas?
- ¿Qué elementos tener en cuenta en la creación y mejora de los recursos léxicos, así como en la implementación y validación de los mismos?
- ¿Cómo manejar aquellas características de las opiniones que más afectan la efectiva detección de la polaridad?
- ¿Cómo integrar los recursos léxicos y las soluciones computacionales para manejar características de las opiniones en un esquema que permita minar, de manera no supervisada, las opiniones?
- ¿Qué herramientas libres y de código abierto facilitan la implementación del esquema para la detección de la polaridad de las opiniones y cómo garantizar la integración entre las mismas?

Después de haber realizado el marco teórico se formuló la siguiente **hipótesis de investigación** como presunta respuesta a las preguntas de investigación: La actualización o creación de herramientas para minar opiniones y el tratamiento de los problemas de este proceso, contribuye a la creación de un efectivo sistema para la detección no supervisada de la polaridad de las opiniones.

El **valor teórico** de la investigación radica en la concepción y definición de dos esquemas para la detección de la polaridad de las opiniones. El **valor práctico** consiste en el sistema PosNeg

Opinion 3.0 que permite la detección no supervisada de la polaridad de las opiniones, los nuevos recursos léxicos SentiWordNet 4.0, y 4.1 y SpanishSentiWordNet, y la biblioteca PolarityDetection, que pueden ser útiles para el desarrollo de diversas herramientas para el análisis de sentimiento. Además, la creación de listas de emoticonos y jergas con polaridades asignadas, y una propuesta para el tratamiento de la negación en las opiniones, constituyen elementos que aportan valor práctico.

La tesis está estructurada en tres capítulos. En el primer capítulo se describen las características, ventajas y desventajas de los diferentes niveles que se han establecido para el análisis de sentimiento, se comentan las tareas básicas de la minería de opinión y las principales técnicas y algoritmos para realizarlas, abordan algunos de los problemas presentes en el tratamiento de las opiniones y finalmente se describen las principales técnicas desarrolladas para la detección de la polaridad de las opiniones. En el segundo capítulo se describen los recursos léxicos existentes, se presentan nuevos recursos léxicos inspirados en los ya existentes e intentando suplir las carencias de éstos, se presentan dos esquemas generales para la detección no supervisada de la polaridad de las opiniones y es descrito el sistema PosNeg Opinion 3.0 que fue diseñado siguiendo los esquemas propuestos para detectar la polaridad de opiniones. En el tercer capítulo se describen las colecciones empleadas en la validación de la propuesta que se presenta, las formas de evaluación para el cálculo de la polaridad de las opiniones, un análisis comparativo de los recursos léxicos creados respecto a aquellos publicados previamente, así como la evaluación de cada uno de los elementos novedosos que conforman la propuesta. Este documento culmina con las conclusiones, recomendaciones, referencias bibliográficas y los anexos.

Capítulo 1 Acerca del análisis de sentimientos

En este capítulo se describirán los diferentes niveles que se han establecido para el análisis de sentimiento y sus ventajas y desventajas. También se comentarán las tareas básicas de la minería de opinión y las principales técnicas y algoritmos que han sido propuestos para dar cumplimiento a estas tareas. Se abordarán algunos de los problemas presentes en el tratamiento de las opiniones, que hacen que trabajar con opiniones sea más complejo que manejar otros tipos de textos. Finalmente, se describirán las principales técnicas desarrolladas para la detección de la polaridad de las opiniones, identificando las características de las propuestas supervisadas y no supervisadas a tales efectos.

1.1 Diferentes niveles del análisis de sentimiento

Existen tres niveles para desarrollar el análisis de sentimiento (Liu 2012; Buche et al. 2013): nivel de documento, nivel de oración o frase, y nivel de aspecto. La clasificación del sentimiento se puede realizar en cualquiera de estos tres niveles. Como se verá a continuación, el nivel de aspectos impone mayores retos en el análisis de sentimiento, mientras que los niveles de frases y documentos requieren menos esfuerzos al aplicar las técnicas de procesamiento del lenguaje natural.

En el nivel de documento se identifica si el documento (críticas de productos, blogs, y mensajes de foros) expresa opiniones y si las opiniones son positivas, negativas o neutras; y finalmente se ofrece un valor de polaridad global de toda la opinión que se plasma en el documento (Liu 2012; Lin et al. 2006; Liu 2011; Liu 2010a). Por ejemplo, dada una revisión de un producto, se desea determinar si la revisión expresa una opinión positiva o negativa en general sobre el producto. Este nivel de análisis supone que cada documento expresa opiniones de una única entidad (por ejemplo, un solo producto). Por lo tanto, no es aplicable a los documentos que evalúan o comparan múltiples entidades.

En (Turney 2002) presentó un trabajo basado en una medida de distancia de los adjetivos que se encuentran en todo el documento con polaridad conocida, es decir, excelente o pobre. Para ello, diseñó un algoritmo de tres pasos. En el primer paso, los adjetivos se extraen junto con una palabra que proporciona la información adecuada. En el segundo paso, la orientación semántica es dada por la medición de la distancia a partir de palabras de polaridad conocida.

En el tercer paso, el algoritmo calcula el promedio de la orientación semántica para todos los pares de palabras y clasifica una revisión como recomendada o no. En experimentos realizados sobre 410 comentarios extraídos de Epinions[1], este trabajo presentó una efectividad del 74%. Otro enfoque es presentado por (Pang et al. 2002) basado en técnicas de clasificación clásicas. Este enfoque tiene como objetivo probar si un grupo seleccionado de algoritmos de aprendizaje automático puede producir buenos resultados cuando la minería opinión es percibida a nivel de documento, asociada a dos clases: positiva y negativa. Presentan los resultados utilizando redes bayesianas, máxima entropía y algoritmos de máquinas de vectores de soporte y muestra buenos resultados comparables a otros oscilando entre el 71% y el 85% de exactitud dependiendo de los conjuntos de datos y el conjunto de entrenamiento.

En el nivel de frase u oración se identifica si una frase es una opinión y si esta es positiva, negativa o neutra. Neutral por lo general significa que no hay opinión (Liu 2012). Este nivel de análisis está estrechamente relacionado con la clasificación de la subjetividad (Wiebe et al. 1999), que distingue frases (llamadas frases objetivas) que expresan información sobre hechos de las frases (llamada frases subjetivas) que expresan puntos de vista y opiniones. Sin embargo, subjetividad no es equivalente al sentimiento pues las frases objetivas pueden implicar opiniones, por ejemplo, “El teléfono se rompió en dos días”. También puede suceder que una oración sea subjetiva y no exprese opinión; por ejemplo, “Yo pienso que él regresó a la casa.”

La minería de opinión a nivel de la oración se asocia con dos tareas (Cardie et al. 2003; Pang et al. 2002; Liu 2010b). La primera es identificar si la oración expresa opinión o no. La segunda es encontrar el valor de una opinión como positiva, negativa o neutra. Se asume que a nivel de la oración una sentencia contiene una sola opinión, como por ejemplo, "La calidad de imagen de esta cámara es buena". Sin embargo, esto no es cierto en muchos casos, al igual que si tenemos en cuenta la oración compuesta, como por ejemplo, "La calidad de la imagen de esta cámara es increíble, así como la duración de la batería, pero el visor es demasiado pequeño para una cámara tan buena", expresa opiniones tanto positivas como negativas y podemos decir que es una opinión mixta. Para "calidad de imagen" y "duración de la batería", la oración es positiva, pero para “visor”, es negativa. La opinión sobre la cámara como un todo

[1] www.epinions.com

es positiva. Con el objetivo de analizar las opiniones compuestas los investigadores determinan las cláusulas que componen una oración (Wilson et al. 2004); no obstante, el nivel de la cláusula no siempre garantiza la correcta determinación de la polaridad; por ejemplo, "Aunque el servicio no es bueno, yo aún adoro ese restaurante". Wilson y colectivo de autores en (Wilson et al. 2004) señalaron que una sola frase no sólo puede contener múltiples opiniones, como ya se había expresado, sino que también tienen cláusulas tanto subjetivas como objetivas. Es útil identificar este tipo de cláusulas dentro de una frase.

Riloff y Wiebe (Riloff & Wiebe 2003) utilizan un enfoque bootstrap para identificar las frases subjetivas con una precisión de alrededor del 90%. Yu y Hatzivassiloglou (Yu & Hatzivassiloglou 2003) clasifican las frases en subjetivas y objetivas utilizando un clasificador bayesiano y determinan la orientación de las mimas (positiva / negativa / neutral). Este trabajo obtuvo una efectividad del 90%. Wilson y colectivo de autores en (Wilson et al. 2004) propusieron los primeros resultados experimentales de la clasificación de la subjetividad de las cláusulas anidadas de las oraciones. Utilizaron una amplia gama de características, incluyendo nuevas características sintácticas desarrolladas para el reconocimiento de opiniones. Sus experimentos obtienen una efectividad en el rango de un 23% a un 79%.

Al igual que la minería de opiniones a nivel de documento, la minería de opiniones a nivel de frase, no descubre lo que exactamente las personas les gusta o no (Liu 2012). El nivel de aspecto lleva a cabo el análisis con mayor granularidad; ya que extrae los atributos (por ejemplo, imagen, calidad, tamaño del lente) de la entidad de la cual se expresa una opinión y la orientación de dicha opinión (Kaur & Duhan 2015). Este fue anteriormente llamado nivel de rasgos o características (Hu & Liu 2004). En vez de enfocarse en las construcciones del lenguaje (documentos, párrafos, oraciones, cláusulas o frases), el nivel aspecto se enfoca directamente en la propia opinión. Se basa en la idea de que una opinión consiste en un sentimiento (positivo o negativo) y un objetivo (entidad). Una opinión sin su objetivo identificado es de uso limitado. El nivel de aspectos también ha sido formalizado como el proceso de extracción de las características del objeto comentado, la determinación de la polaridad de las características del objeto, y luego agrupar las características similares y producir un informe en forma de resumen (Mishra & Jha 2012).

El darse cuenta de la importancia de los objetivos de la opinión también ayuda a entender mejor el problema del análisis de sentimientos. Por ejemplo, aunque la frase "aunque el

servicio no es tan bueno, todavía me encanta este restaurante" tiene claramente un tono positivo, no podemos decir que esta frase es del todo positiva. De hecho, la sentencia es positiva sobre el restaurante, pero negativa acerca de su servicio. En muchas aplicaciones, los objetivos de opinión se describen por las entidades o sus diferentes aspectos. Por lo tanto, el objetivo de este nivel de análisis es descubrir los sentimientos de entidades o sus aspectos. Por ejemplo, la frase "la calidad de la llamada del iPhone es buena, pero la duración de la batería es corta" evalúa dos aspectos, la calidad de la llamada y duración de la batería, de la entidad iPhone. El sentimiento de calidad de la llamada del iPhone es positivo, pero el sentimiento de la vida de la batería es negativo. La calidad de la llamada y la duración de la batería del iPhone son los objetivos de la opinión. Basados en este nivel de análisis, se puede producir un resumen estructurado de opiniones sobre entidades y sus aspectos, lo que convierte el texto no estructurado en datos estructurados y puede ser utilizado para todo tipo de análisis cualitativos y cuantitativos. Tanto las clasificaciones a nivel de documento como a nivel de oración ya son altamente desafiante; el nivel de aspecto es incluso más difícil.

Liu y Cheng (Liu et al. 2005) utilizan un método de aprendizaje de patrones supervisado para extraer las características del objeto y consecuentemente calcular la polaridad. Para identificar la orientación de la opinión usó un enfoque basado en un recurso léxico. Este enfoque básicamente utiliza palabras de opinión y expresiones en una oración para determinar la polaridad siguiendo tres pasos (Ding et al. 2008): identificar las palabras con polaridad, identificar el rol de las palabras negadoras e identificar las cláusulas comparativas. Esta investigación obtuvo en varias colecciones altos valores de precisión. La precisión promedio en opiniones positivas es del 88.9% y del 79.1% en las negativas. Además, los valores promedio de recall fueron del 90.2% en las opiniones positivas y del 82.4% en las negativas.

Hu y Liu presentaron un análisis de comentarios de los clientes a través de la minería opinión basada en la frecuencia de las características, en el que las características más frecuentes son aceptadas para el procesamiento y son las que se evalúan durante la generación del resumen (Hu & Liu 2004). En este trabajo apoyado en métodos de minería de datos y del procesamiento del lenguaje natural, además del uso del diccionario WordNet (Miller 1995), obtienen un promedio de efectividad de un 84%. En oposición a Hu y Liu, Popescu y Etzioni (Popescu et al. 2005), también desde una perspectiva no supervisada mejoran el enfoque basado en la frecuencia con la introducción de la relación parte-de y eliminan las frecuentes

ocurrencias de frases nominales que pueden no ser características. Estos obtienen valores de 86% de precisión y un 89% de recall.

1.2 Tareas básicas del análisis de sentimiento

La complejidad del procesamiento de las opiniones, así como los distintos análisis que se pueden realizar a partir de ellas, han conllevado a la definición de tareas básicas del análisis de sentimientos. A continuación se describirán las características principales de cada una de ellas.

Detección de una opinión:

La determinación de si un documento contiene una opinión es algo diferente de la clasificación de un documento como una opinión. Este último, por lo general requiere de la aplicación de técnicas de aprendizaje supervisado, y depende de la característica general del documento (por ejemplo, grado de subjetividad), mientras que el primero, generalmente requiere el uso de léxicos de opinión y se basa en la detección de evidencias de opinión. A nivel de frase o párrafo; sin embargo, la distinción entre determinar si estamos en presencia de una opinión o clasificar la polaridad de la opinión, se vuelve intrascendente ya que las características de la frase o párrafo están fuertemente influenciadas por la presencia de evidencias de opinión (Yang et al. 2007). No obstante, existen oraciones que no expresan opinión alguna y tienen evidencias de opinión; por ejemplo: "Si las batidoras que sacaron están buenas, me voy a comprar una". Resulta más efectivo trabajar a nivel de frases para llevar a cabo la detección de una opinión. Realizar esta tarea a nivel de frases utilizando técnicas de aprendizaje supervisado, genera dos nuevos problemas. En primer lugar, es probable que produzca un clasificador no muy adecuado para la clasificación del nivel de frases con datos de entrenamiento con etiquetas de nivel de documento. En segundo lugar, las frases son documentos cortos descritos por muy pocos rasgos y esto puede disminuir la eficacia del clasificador. De ahí que generalmente se utilice para este nivel de clasificación un enfoque basado en recursos léxicos, que consiste en crear y consultar listas de palabras o frases subjetivas, las cuales pueden ser construidas de forma automática o manual (Mishne & others 2006; Oard et al. 2006; Yang et al. 2006; Zhang & Ye 2008). Algunas propuestas que detectan opiniones han sido presentadas por Hannah y colectivo de autores en (Hannah et al. 2007), Ounis y colectivo de autores en (Yang et al. 2007), Zhang y colectivo de autores en (Zhang & Yu 2006; Wei et al. 2007), Seki y Uehara en (Seki & Uehara 2009).

Detección de subjetividad:

La detección de la subjetividad consiste en determinar si una unidad textual tiene una naturaleza objetiva (hecho) o subjetiva (opinión). Un ejemplo de oración subjetiva es “Entre varios aspectos, es un título fascinante.” y uno de una objetiva que expresa un hecho o acontecimiento es “La nueva estación de trenes será inaugurada el próximo mes.” (Esuli & Sebastiani 2006a). La mayor parte de los análisis concuerdan que el texto objetivo constituye datos reales, mientras que el texto subjetivo representa puntos de vista individuales, creencias, opiniones o sentimientos. Por lo tanto, la mayoría de los sistemas de minería opinión utilizan texto subjetivo para determinar la orientación de la opinión (Archak et al. 2007).

Existe cierta confusión entre los investigadores de equiparar la subjetividad con opinión (Liu 2012). Se dice opinión a un documento o frase que expresa o implica un sentimiento positivo o negativo. Los dos conceptos no son equivalentes, aunque tienen una gran intersección. Una oración subjetiva puede no expresar sentimiento. Por ejemplo, "Creo que él fue a su casa" es una oración subjetiva, pero no expresa ningún sentimiento. Y, por el contrario, frases objetivas pueden implicar opiniones o sentimientos debido a que los hechos pueden ser deseables e indeseables (Zhang & Liu 2011). Por ejemplo, la siguiente frase establece un hecho que implica claramente una opinión negativa (opinión implícita) sobre el respectivo producto, porque el hecho es no deseable: "El auricular se partió en dos días.". Además de las opiniones explícitas como expresiones subjetivas, existen muchos otros tipos de subjetividad que han sido estudiados, aunque no tan extensamente; por ejemplo, el afecto, el juicio, la apreciación, la especulación, textos en perspectiva, con acuerdos y desacuerdos, posturas políticas (Alm 2008; Ganter & Strube 2009; Greene & Resnik 2009; Hardisty et al. 2010; Lin et al. 2006; Medlock & Briscoe 2007; Mukherjee & Liu 2012; Murakami & Raymond 2010; Neviarouskaya et al. 2010; Somasundaran & Wiebe 2009). Muchos de ellos también pueden implicar sentimientos. Por tanto, determinar si una oración o frase es subjetiva u objetiva no implica que sea o no una opinión, de ahí la diferencia de esta tarea con la de detección de opinión.

Hay una gran variedad de enfoques que se adoptan para esta tarea de la minería de opinión. Los más destacados son los basados en heurísticas y la estructura del discurso, el análisis de grano fino y grueso, las palabras claves y el análisis de la idea (Abdul-Mageed et al. 2011). La mayoría de estos enfoques se basan en el aprendizaje supervisado(Liu 2012). Por ejemplo, los

primeros trabajos reportados en (Wiebe et al. 1999) llevan a cabo la clasificación de la subjetividad utilizando un clasificador Bayesiano. Otras propuestas para detectar la subjetividad han sido presentadas en (Veloso & Meira Jr 2007; Esuli & Sebastiani 2006a; Yu & Hatzivassiloglou 2003; Wiebe & Riloff 2005; Wiebe et al. 2004; Riloff et al. 2006; Education 2004; Barbosa & Feng 2010; Raaijmakers & Kraaij 2010; Schapire & Singer 2000; Wilson et al. 2004; Benamara et al. 2011; Wiebe 2000). Otros trabajos adicionales sobre la clasificación de la subjetividad de frases han estado dirigidas a otros idiomas como árabe (Abdul-Mageed et al. 2011) y los idiomas urdu (Mukund & Srihari 2010) sobre la base de diferentes algoritmos de aprendizaje automático que utilizan las características generales y específicas del lenguaje.

Clasificación de la opinión:

El paso más importante en el análisis de opiniones es la clasificación de la polaridad del texto que consiste en determinar su polaridad, es decir, si la opinión es negativa, positiva o neutral. Dada una colección de documentos D= {d1...dn} y categorías predefinidas en el conjunto C = {positiva, negativa, neutral}, la clasificación del sentimiento es clasificar cada di en D, con una etiqueta expresada en C (Sindhu & Ch 2013). También se puede considerar la clasificación de la polaridad o del sentimiento como una tarea de clasificación binaria, donde un documento de opinión está marcado con un sentimiento positivo o negativo en general (Rashid et al. 2013; Mishra & Jha 2012; Chandrakala & Sindhu 2012). En (Pang et al. 2002; Turney 2002; Dave et al. 2003) consideran que la determinación de la polaridad, a partir de un texto subjetivo, es establecer si la opinión expresada es positiva, negativa o mixta. El aprendizaje automático y el enfoque basado en recursos léxicos son los más populares para clasificar las opiniones (Selvam & Abirami 2009).

Es objetivo de este trabajo centrarse en la tarea de clasificación de la opinión, es decir, determinar si es positiva o negativa, e incluso especificar los niveles de positividad y negatividad; de ahí que se abordará esta etapa de manera detallada en el epígrafe 1.4.

Determinación de la fuerza de la opinión:

Determinar la fuerza de la opinión consiste en calcular en qué medida es positiva o negativa, es decir, cuantificar el grado del sentimiento. Mientras que la asignación del sentimiento se ocupa de analizar si un texto tiene una orientación semántica positiva, negativa o neutra, la intensidad del sentimiento trata con el análisis de si los sentimientos positivos o negativos son

suaves o fuertes. Por ejemplo, las frases "no me gusta usted" y "te odio" tienen asignadas una orientación semántica negativa, pero la segunda se considera más intensa que la primera (Abbasi et al. 2011). También se puede definir esta tarea como un problema multiclases con clases que por lo general siguen un rango como: débilmente positivo, ligeramente positivo o muy positivo. También se conoce como clasificación afectiva, cuando se trata de predecir los estados de ánimo de los usuarios, como la felicidad, la tristeza, la bondad, la igualdad y así sucesivamente (Subasic & Huettner 2001; Grefenstette et al. 2004; Bollen et al. 2011).

La clasificación de las polaridades de sentimiento e intensidades generalmente implica el uso de métodos de clasificación supervisada aplicados a rasgos lingüísticos; dentro de ellos, el enfoque basado en máquinas de vectores de soporte (Support Vector Machine; SVM) ha superado diversas técnicas que incluyen clasificadores Bayesianos, los árboles de decisión, Winnow, etc. (Tatemura 2000; Terveen et al. 1997; Jin et al. 2007).

Algunas propuestas han dado soluciones al problema de predecir las puntuaciones de evaluación (por ejemplo, 1-5 estrellas) de los comentarios (Pang & Lee 2005). En este caso, el problema puede formularse como un problema de regresión ya que las puntuaciones de evaluación son ordinales, aunque no todos los investigadores resolvieron el problema mediante técnicas de regresión. En (Goldberg & Zhu 2006) mejoraron este enfoque modelando un índice de predicción como un problema de aprendizaje semi-supervisado basado en grafos que utiliza tanto opiniones marcadas (con puntuaciones) como no marcadas (sin calificación). En (Qu et al. 2010) introdujeron una representación de los documentos como bolsas de opiniones para captar la fuerza de n-gramas con opiniones, que es diferente de la representación tradicional de bolsa de palabras. Liu y Seneff (Liu & Seneff 2009) propusieron un método para extraer frases adverbio-adjetivo-sustantivo (por ejemplo, "*very nice car*"), basado en la estructura de la cláusula obtenida mediante el análisis de las oraciones en una representación jerárquica.

Determinación de la fuente de la opinión:

La fuente de una opinión puede ser una persona o una institución (Shelke et al. 2012). Identificar la fuente de las opiniones es vital para asociar una fuerza mayor o menor a las opiniones dependiendo del grado de experticia de quien la emite. Por ejemplo, si se está abordando un tema médico, las opiniones emitidas por doctores serán más tomadas en cuenta que aquellas emitidas por personas no especialistas en el tema en cuestión. La fuente de las

opiniones también aporta información adicional para el desarrollo de otras tareas como por ejemplo el resumen de opiniones, ya que facilita el agrupamiento por grupos de edades, regiones, profesiones y tipos de instituciones, entre otros (Deshpande 2012).

Esta tarea está estrechamente relacionada con la extracción del titular de la opinión. A esta subtarea se la ha prestado menor atención por parte de los investigadores de la minería de opinión. Uno de los enfoques existentes para hacer frente a este problema es realizar un seguimiento de la presencia de indicios de subjetividad con el fin de identificar los usos de verbos como opinión (Elarnaoty et al. 2012), mientras se usa el análisis semántico para localizar las fuentes de opinión (Bethard et al. 2004). Otro método consiste en tratar la tarea de la extracción del titular de la opinión como un problema de clasificación de etiquetado secuencial (Choi et al. 2005; Choi et al. 2006). Características estructurales del árbol de análisis sintáctico son seleccionadas por otros investigadores para modelar la relación estructural entre un titular y una opinión (Kim & Hovy 2006b). La identificación de los titulares de opinión puede beneficiarse de, o tal vez incluso requerir, representación de expresiones de opinión, ya sea de forma simultánea (Bethard et al. 2004; Choi et al. 2005; Choi et al. 2006) o como un pre-procesamiento (Kim & Hovy 2005; Kim & Hovy 2006b).

En (Choi et al. 2005) consideran la extracción del titular de la opinión como una tarea de extracción de información que usa la combinación de dos técnicas: reconocimiento de entidades y la extracción de información. Los modelos formales la identifican como una tarea de etiquetado secuencial, que después aprende mediante la extracción de patrones. En (Bloom et al. 2007) describen la extracción del titular basados en un lexicón construido a mano, una combinación de análisis heurístico superficial y análisis de dependencias, y la expectativa de maximización de la desambiguación del sentido de las palabras; ellos buscan frases en el texto con tipos de taxonomías de titulares dominio-dependientes. En (Kim & Hovy 2006a) utilizaron técnicas de aprendizaje automatizado para la extracción del titular de la opinión. Como características para su clasificador de máxima entropía, seleccionan características estructurales a partir de un análisis sintáctico de profundo, sobre la base de una representación en bloques de las opiniones. Los bloques se desarrollaron alrededor de las palabras de opinión, y se investigaron las relaciones semánticas entre éstos y el titular de la opinión y del objetivo de la opinión. En (Kim et al. 2008) se utilizan el recurso léxico SentiWordNet (Esuli & Sebastiani 2006b), un reconocedor de entidades nombradas, y un analizador sintáctico para la

extracción del titular de opinión. También en (Seki 2008) utilizan un enfoque de clasificación de autor y de autoridad como base para la detección del titular de la opinión. Este enfoque se basa en las características seleccionadas a partir de la significación de la frecuencia en un corpus de entrenamiento, y clasifica las oraciones con opiniones que sean expresadas desde un punto de vista de un autor o desde el punto de vista de una autoridad. Esta diferenciación es utilizada para la identificación del titular de la opinión, ya que trata estos dos tipos de opinión con diferentes reglas al extraer su titular.

Determinación del objetivo de la opinión:

Al determinar el objetivo de la opinión es necesario indagar de quién se habla en la opinión y con quién se está de acuerdo o no. El objetivo de la opinión hace referencia a la persona, objeto, función, evento o tema relacionado con la opinión que se expresa, el resultado de la identificación de los objetivos de opinión es una característica fundamental de la minería de la opinión. El análisis en profundidad de cada faceta de un producto contenido en una opinión del cliente es necesario para el público en general, los comerciantes y los fabricantes (Anwer et al. 2010).

Los clientes generalmente expresan sentimientos positivos, negativos o ambos sobre el objeto de la opinión y sus atributos (Sharma & Chitre 2014). La clasificación a nivel de documento y la clasificación a nivel de oración no permiten determinar los gustos y preferencias de los consumidores acerca de los atributos particulares del objeto (Fan & Wu 2012). Cuando un consumidor hace un comentario sobre un objeto (producto, persona, tema u organización), éste comenta sobre las características del objeto (Hiroshi et al. 2004). Por ejemplo, si los usuarios comentaron sobre un teléfono móvil, básicamente pueden hacer comentarios sobre la cámara, el tamaño del LCD, el altavoz, el peso, etc. Si en la salida sobre la cámara 125 comentarios expresan opiniones positivas y 25 comentarios pueden ser negativos, un nuevo cliente que esté interesado en la calidad de la cámara del móvil puede tomar fácilmente la decisión de comprar el producto o no (Liu 2012). Este ejemplo ilustra que la clasificación de los textos de opinión a nivel de documento o el nivel de la oración es a menudo insuficiente para algunas aplicaciones debido a que no se identifican los objetivos de opinión o los sentimientos que se asignan a tales objetivos (Liu 2012). Incluso si se asume que cada documento evalúa una sola entidad, un documento de opinión positiva sobre la entidad no quiere decir que el autor tiene una opinión positiva sobre todos los aspectos de la entidad. Del mismo modo, una opinión

negativa no significa que el autor es negativo acerca de todo. Para un análisis más completo, es necesario descubrir los aspectos y determinar si el sentimiento es positivo o negativo en cada aspecto (Hu & Liu 2004). La minería de opinión por aspectos facilita la determinación del objetivo de la opinión (Seerat & Azam 2012). Un estudio de fondo revela que la extracción del objetivo de la opinión implica la aplicación de diversas técnicas de procesamiento del lenguaje natural, tales como tokenización, etiquetado de las diferentes partes del discurso, eliminación de palabras vacías, la selección de entidades y en algunos casos hasta la aplicación de analizadores sintácticos (Vasantharaj et al. 2015).

Existe una estrecha relación entre la minería de opinión basada en aspectos y la determinación del objetivo de la opinión. El objetivo es a menudo el aspecto o tema que puede ser extraído de una frase. Sin embargo, también hay que señalar que algunas opiniones pueden jugar dos papeles, o sea, que indican un sentimiento positivo o negativo y éste implican un aspecto (objetivo) de manera implícita. Por ejemplo, en la oración "este coche es caro", "caro" es una palabra de sentimiento y también indica el aspecto precio. Para la extracción de un aspecto explícito existen cuatro enfoques principales: la extracción basada en la frecuencia de los sustantivos y las frases nominales (Hu & Liu 2004; Popescu et al. 2005; Blair-Goldensohn et al. 2008; Hu et al. 2010; Moghaddam & Ester 2010; Scaffidi et al. 2007; Zhu et al. 2009; Long et al. 2010) la extracción explotando la relación de la opinión y el objetivo (Hu & Liu 2004; Blair-Goldensohn et al. 2008; Zhuang et al. 2006; Somasundaran & Wiebe 2009; Kobayashi et al. 2006; Qiu et al. 2011; Wu et al. 2009; Kessler & Nicolov 2009), extracción usando aprendizaje supervisado (Indurkhya & Damerau 2010; Mooney & Bunescu 2005; Sarawagi 2008; Jakob & Gurevych 2010; Li, Han, et al. 2010; Choi & Cardie 2010; Yu et al. 2011) y extracción usando tópicos (Mei et al. 2007; Titov & McDonald 2008; Branavan et al. 2009; Lin & He 2009; Brody & Elhadad 2010; Li, Huang, et al. 2010; Zhao et al. 2010; Sauper et al. 2011; Griffiths et al. 2004; Lu & Zhai 2008; Jo & Oh 2011; Lu et al. 2009; Moghaddam & Ester 2011).

No solo se ha trabajado en la detección de aspectos explícitos, algunas propuestas han estado dirigidas a identificar aspectos implícitos. Existen muchos tipos de expresiones que indican aspectos implícitos. Adjetivos y adverbios son, quizás, los tipos más comunes porque la mayoría de los adjetivos describen algunos atributos o propiedades de entidades específicas; por ejemplo, “hermosa” describe "apariencia". Aspectos implícitos pueden también ser verbos.

En general, las expresiones de aspectos implícitos pueden ser muy complejas; por ejemplo, "Esta cámara no va a caber fácilmente en un bolsillo." "Cabe en un bolsillo" indica el aspecto de tamaño. Como la extracción de aspecto explícito se ha estudiado ampliamente, la investigación se ha hecho limitada a identificar las correspondencias entre los aspectos implícitos y los respectivos aspectos explícitos. En (Su et al. 2008) se propuso un método de agrupamiento para identificar aspectos implícitos y las reglas de asociación fueron aplicadas para la determinación de tales aspectos en (Hai et al. 2011).

Resumen de las opiniones y/o visualización gráfica de los resultados:

Una tarea muy importante y que contribuye a la efectiva toma de decisiones es el resumen de las opiniones y la visualización gráfica de los resultados. Esta tarea se puede llevar a cabo agregando votos (índices de 1 a 5, estrellas), sobresaltando algunas opiniones, representando acuerdo/desacuerdo, etc.

El resumen de texto se ha estudiado ampliamente en el procesamiento del lenguaje natural (Das & Martins 2007). Sin embargo, un resumen de opiniones es muy diferente del resumen de un documento tradicional, ya que al resumir opiniones generalmente es necesario centrarse en las entidades, los aspectos y los sentimientos acerca de ellos. Además, es necesario incluir elementos cuantitativos, que son la esencia del resumen basado en los aspectos de la opinión (Raut & Londhe 2014).

El resumen puede estar en una forma estructurada o en una forma no estructurada como un documento de texto corto, los componentes claves de un resumen deben incluir opiniones sobre diferentes entidades y sus aspectos y también deben tener un punto de vista cuantitativo (Liu 2012). La perspectiva cuantitativa es especialmente importante, ya que se desprenden análisis diferentes si se conoce que el 20% de los clientes tienen una opinión positiva que si fuera el 90% opinando de manera positiva.

El proceso de resumen de la opinión se centra principalmente en los siguientes dos enfoques: basado en características (Bahrainian & Dengel 2013; Ranade et al. 2013; Liu et al. 2012) cuando implica hallazgo de términos frecuentes (características) que están apareciendo en muchos comentarios y cuando se presenta mediante la selección de frases que contienen información característica particular, por ejemplo, utilizando análisis semántico latente (Latent Semantic Analysis, LSA) (Liu et al. 2012; Steinberger 2013).

Algunos autores han realizado el resumen de las opiniones basándose en aspectos. El resumen basado en los aspectos de la opinión tiene dos características principales. En primer lugar, se capta la esencia de las opiniones: objetivos de opinión (entidades y sus aspectos) y sentimientos acerca de ellos. En segundo lugar, es cuantitativa, lo que significa que da el número o porcentaje de personas que tienen opiniones positivas o negativas sobre las entidades y aspectos. Esta forma de resumen estructurado ha sido adoptado por investigadores para resumir críticas de películas (Zhuang et al. 2006), para resumir el texto de la opinión (Hu et al. 2010), y para resumir evaluaciones de servicios (Blair-Goldensohn et al. 2008). Sin embargo, hay que señalar que el resumen basado en el aspecto no tiene por qué ser de estructurado; pudiera ser en forma de un documento de texto basado en la misma idea. Algunas mejoras realizadas a las propuestas iniciales de la creación de resúmenes de las opiniones basándose en aspectos fueron publicadas en (Carenini et al. 2013; Tata & Di Eugenio 2010; Lu et al. 2010; Hu et al. 2010; Lerman et al. 2009; Nishikawa et al. 2010a; Nishikawa et al. 2010b; Ganesan et al. 2010; Yatani et al. 2011).

Varios investigadores también estudiaron el problema de resumir las opiniones mediante la búsqueda de puntos de vista contrastantes (Kim & Zhai 2009; Paul et al. 2010; Park et al. 2011; Lerman & McDonald 2009). Otros autores han estudiado el resumen de opiniones de la manera tradicional, es decir, la idea es ofrecer un resumen del texto sin especificar los aspectos (o temas) y sentimientos acerca de ellos (Seki et al. 2006; Wang & Liu 2011; Beineke et al. 2003). Una desventaja de los resúmenes tradicionales es que no consideran o consideran muy poco las entidades, aspectos, y sentimientos acerca de ellos, por lo tanto, pueden seleccionar frases que no estén relacionadas con los sentimientos. Además, no ofrecen un punto de vista cuantitativo, que a menudo es importante en la práctica.

1.3 Algunos de los problemas presentes en el tratamiento de las opiniones

Las opiniones están permeadas de diversos tipos de problemas que hacen que se torne difícil el desarrollo de técnicas de minería de opinión. La mayoría de los elementos inconvenientes al minar opiniones se encuentra en el suministro de la opinión, el objetivo de la opinión, y también las expresiones críticas o comentarios creados por los usuarios.

Algunas críticas de productos son altamente específicas y ricas en opiniones, que nos permiten ver diferentes cuestiones con mayor claridad que otras formas de textos de opinión.

Conceptualmente, no existe ninguna diferencia entre diversos tipos de textos de opinión. Las diferencias son principalmente superficiales y en el grado de dificultad de tratar con ellos. Por ejemplo, los comentarios en Twitter[2] (Tweets) son cortos (140 caracteres como máximo) y no formales, y utilizan muchas jergas de internet y emoticonos. Los tweets son, de hecho, más fácil de analizar debido al límite de longitud que hace a los autores ser directo e ir al grano. Por lo tanto, a menudo es más fácil de conseguir una alta exactitud en el análisis de sentimientos. Los comentarios de productos son también más fáciles, ya que están muy centrados y con poca información irrelevante. Los foros de discusión son quizás el más difícil de tratar debido a que los usuarios pueden hablar de cualquier tema y también interactuar unos con otros. En cuanto al grado de dificultad, existe también la dimensión de los diferentes dominios de aplicación. Opiniones sobre los productos y servicios son generalmente más fáciles de analizar. Los debates sociales y políticos son mucho más difíciles debido a lo complejo del tema y la presencia de expresiones de sentimientos, sarcasmos e ironías (Liu 2012).

Hay varios desafíos en el análisis de los sentimientos de las opiniones, en primer lugar, una palabra que se considera que es positiva en una situación puede ser considerada negativa en otra situación (Chandrakala & Sindhu 2012). Podemos tomar la palabra "largo", por ejemplo. Si un cliente dijo: “el tiempo de duración de la batería de un ordenador portátil fue largo”, eso sería una opinión positiva. Si el cliente dijo: “el tiempo de arranque del portátil fue muy largo”, eso sería una opinión negativa (Abbasi et al. 2011). Estas diferencias hacen que un sistema de minería de opinión entrenado para recoger opiniones sobre un tipo de producto o característica del producto puede no funcionar muy bien en otro. Otro reto surge debido a que las personas no siempre expresan las opiniones de la misma manera. Si un cliente A comenta de un teléfono móvil, "la calidad de la voz es excelente" y el cliente B comenta, "la calidad del sonido del teléfono es muy buena", ambos están hablando de la misma función pero con una redacción diferente. Identificar los sinónimos de las palabras es también una tarea difícil, sobre todo porque dos palabras pueden ser sinónimos en determinado contexto y en otro no (Sharma & Chitre 2014). La mayor parte de procesamiento de texto tradicional se basa en el hecho de que las pequeñas diferencias entre los dos fragmentos de texto no cambian mucho el

[2] www.twitter.com

significado. En la minería de opinión, sin embargo, "the movie was great" es muy diferente de "the movie was not great" (Wiegand et al. 2010).

Las personas pueden ser contradictorias en sus declaraciones. La mayoría de los comentarios tienen criterios positivos y negativos, lo cual es poco manejable mediante el análisis individual de oraciones. Sin embargo, mientras más informal es el medio (Twitter o blogs, por ejemplo), más probable es combinar diferentes opiniones en la misma frase. Por ejemplo, un ser humano puede entender con facilidad la siguiente opinión "the movie bombed even though the lead actor rocked it", pero es difícil de analizar automáticamente. En algunas situaciones, incluso, las personas tienen dificultades para entender lo que alguien pensó a partir del análisis de un pequeño fragmento de texto, porque carece de contexto. Por ejemplo, la expresión "That movie was as good as his last one" es totalmente dependiente de lo que la persona que expresa la opinión pensó de la película anterior. La pragmática es un subcampo de la lingüística que estudia las formas en que el contexto contribuye al significado. Es importante detectar la pragmática de la opinión del usuario que puede cambiar el sentimiento que se expresa. La capitalización se puede utilizar con sutileza para denotar el sentimiento. Por ejemplo, la oración “I just finished watching THE DESTROY” expresa un sentimiento positivo, mientras que la “That completely DESTROYED me” denota un sentimiento negativo.

Otro reto al minar opiniones es identificar la entidad. Un texto o frase puede tener múltiples entidades asociadas a él. Esto es extremadamente importante para encontrar la entidad en la que la opinión está enfocada (Mukherjee & Bhattacharyya 2013). Por ejemplo, en las oraciones “Sony is better than Samsung” y “Raman defeated Ravanan in football”, se expresa un sentimiento positivo para Sony y Raman pero negativo para Samsung y Ravanan.

Otro elemento a tener en cuenta al procesar opiniones es que la reseña del producto, comentarios y críticas podrían estar en diferentes idiomas (inglés, español, urdu, árabe, francés, etc), por lo tanto, hacer frente a cada idioma de acuerdo con su orientación es una tarea difícil (Sharma & Chitre 2014). Hasta la fecha, los analizadores de sentimiento que se han implementado son en su mayoría sólo para un idioma determinado (Kaur & Duhan 2015). Aunque algunos analizadores de sentimiento se han puesto en práctica para clasificar comentarios en varios idiomas, la mayoría de ellos se implementan para el idioma Inglés. Lograr un analizador de sentimiento no dependiente del lenguaje sería algo fructífero, ya que daría una visión amplia de la opinión hacia un producto.

Al detectar los aspectos generalmente se trabaja en función de identificar los sustantivos presentes en la opinión; sin embargo, algunos verbos, adverbios y adjetivos también pueden expresar aspectos y sentimientos asociados a los aspectos. De ahí que la selección de aquellas palabras importantes para el análisis a partir de la salida de un etiquetador, no es trivial. Por ejemplo, "like" es un verbo, pero también es una palabra de opinión (Sharma & Chitre 2014).

Ciertas palabras exhiben diferentes polaridades cuando se utilizan en diferentes dominios (Kaur & Duhan 2015). Por ejemplo, "The movie was inspired from a Tollywood movie" tiene una orientación negativa mientras que "I got inspired from the novel" tiene una orientación positiva. Aquí, la palabra "inspired" exhibe dos polaridades diferentes para dos contextos diferentes. El análisis de sentimiento se lleva a cabo habitualmente dirigido a un dominio específico, así se han logrado muy buenos resultados de precisión. Sin embargo, un analizador de sentimiento generalizado sigue siendo un reto debido a la diferencia en el significado de una palabra o frase en diferentes dominios.

En el formato libre, se pueden utilizar las abreviaturas, las palabras cortas, y el lenguaje romano (Sharma & Chitre 2014). Por ejemplo "u" para "you", "cam" para "camera", "pic" para "picture", "f9" para "fine", "b4" para "before", "gud" para "good" etc. Para hacer frente a este tipo de lenguaje se necesita mucho trabajo.

Las jergas son por lo general las formas cortas e informales de las palabras originales de uso frecuente en los mensajes de texto en línea (Kaur & Duhan 2015). Por ejemplo "gr8" es una palabra del argot usado para "great", "5n" o "fyn" para "fine". Estas palabras no son una parte de los diccionarios tradicionales, pero se encuentran ampliamente en los textos en línea. Si estas jergas se pueden asignar a las palabras originales, los resultados de rendimiento de los analizadores de los sentimientos pueden ser mejorados.

La negación juega un papel vital en la alteración de la polaridad del adjetivo asociado y por lo tanto la polaridad del texto. Una posible solución para manejar la negación es invertir la polaridad del adjetivo que aparece después de una palabra de negación, por ejemplo, "el restaurante es bueno", debe ser una opinión clasificada positiva, pero "el restaurante no es bueno", debe ser una opinión clasificada negativa. Sin embargo, esta solución falla en los casos como "No es una maravilla pero el restaurante es bueno" y "No sólo la comida era deliciosa, el interior y el servicio también eran excelentes". El uso de técnicas de

procesamiento de lenguaje natural o el uso de modelos matemáticos no resuelven por completo las negaciones (Kaur & Duhan 2015).

A veces, el texto contiene otra entidad para referirse a una entidad. En ese caso, se requiere el conocimiento de la entidad que se utiliza para hacer referencia a la otra para identificar el sentimiento. Por ejemplo, dada la afirmación "She looks as Snow White", es necesario saber acerca de “Snow White” para identificar la orientación del sentimiento del texto (Kaur & Duhan 2015).

En este epígrafe se han abordado varias características de las opiniones que evidencian la complejidad que tiene la minería de opinión respecto la minería de textos en general.

1.4 Detección de la polaridad de las opiniones

La clasificación de sentimientos o la clasificación de la polaridad es una de las tareas más importantes en el análisis de sentimientos, en la que cada parte del texto se etiqueta como positivo, negativo o neutro según la opinión general, expresada en dicho texto (Angulakshmi & ManickaChezian 2014). Otros autores consideran la clasificación de la polaridad como una tarea de clasificación binaria para etiquetar un documento con opiniones expresando un valor positivo o negativo (Sindhu & Ch 2013). Identificar la polaridad de los sentimientos puede parecer trivial para la mente humana. Sin embargo, la automatización de este proceso no es así de simple. Una serie de estudios han sido desarrollados para encontrar soluciones a este problema.

El método más sencillo para la clasificación del sentimiento es armar listas de palabras que se dividen en dos grupos, el primer grupo es una lista de palabras y sinónimos con polaridad positiva (por ejemplo, buena, óptimo, excelente) y el segundo es una lista de palabras y de sinónimos con polaridad negativa (por ejemplo, mala, ruin, desagradable). A partir de estos dos grupos, se busca en el texto la ocurrencia de las palabras que aparecen en la lista, y el texto se clasifica según el número de palabras positivas y negativas localizadas (Kim & Hovy 2004). Un ejemplo de este enfoque es Twitrratr[3], que es una herramienta de rastreo que mide la opinión pública a través de los mensajes publicados por los usuarios de la red social Twitter. A través de una palabra clave introducida en el campo de búsqueda del sistema, la aplicación

[3] http://twitrratr.com/

busca comentarios donde mencionan la palabra clave. Una vez encontrados, los comentarios se cruzan con las listas de adjetivos preseleccionados y son consecuentemente clasificados como positivos, negativos o neutros. Las listas de adjetivos, divididas en aspectos positivos y negativos, están disponibles en el sitio web para su visualización. Es un método simple y fácil, pero no cubre algunos problemas de las opiniones mencionados en el epígrafe 1.3.

A partir de esta propuesta inicial, se han desarrollado diversos enfoques para la detección de la polaridad de las opiniones, así como diferentes nomenclaturas para describir estos enfoques. En (Poirier et al. 2011) describen dos enfoques para clasificar las opiniones: el enfoque lingüístico y el estadístico. El primer enfoque consiste en aplicar técnicas de aprendizaje automático y el segundo consiste en calcular estadísticos a partir de los términos presentes en las opiniones. En general, los sistemas que aplican estas técnicas clasifican los comentarios textuales en dos clases (positivos y negativos). Según el autor los dos enfoques pueden ser combinados.

Abbasi y colectivo de autores (Abbasi et al. 2008) presentan técnicas para la clasificación de sentimientos en tres categorías. En diversas investigaciones, algoritmos de aprendizaje automatizado han sido utilizados, siendo los más comunes las SVM, ampliamente usadas para clasificar comentarios de películas y los clasificadores bayesianos que han sido aplicados en los comentarios disponibles en sitios de productos en la web. El clasificador SVM ha obtenido mejores resultados que los clasificadores bayesianos.

Según (Abbasi et al. 2008), otra categoría es la basada en el análisis de enlaces (link analysis). El autor comenta como ejemplo, el uso de análisis de co-citaciones para clasificar opiniones en sitios web, buscando cuantificar la relación entre dos documentos d_i y d_j, anotando los documentos citados por ambos. La idea del análisis de co-citación es que, si dos documentos poseen varias citaciones iguales, se puede inferir que ambos tienen alguna semejanza, aunque estos dos documentos no se citen directamente (Efron 2004). Una limitación de este enfoque es cuando la estructura de los enlaces no es clara o cuando los enlaces son escasos.

Una tercera categoría es basada en puntuaciones (score). Generalmente esta técnica clasifica textos de opinión basándose en la suma del total de las características extraídas determinadas como positivas o negativas. Para los términos positivos es atribuido el valor +1, mientras que para los términos negativos es atribuido el valor -1. La clasificación del sentimiento de una oración es dada como positiva si la suma de los valores de sus términos es positiva, por el

contrario, si la suma da un valor negativo la clasificación asignada es negativa (Abbasi et al. 2008).

En (Liu 2010b; Yessenov & Misailovic 2009; Chaovalit & Zhou 2005; Mukras 2009; NG 2010) dividen la clasificación de sentimientos en dos grupos: los que emplean algoritmos de aprendizaje supervisado y los que emplean algoritmos de aprendizaje no supervisado. Según (Chaovalit & Zhou 2005), el primer grupo utiliza aprendizaje automatizado donde los clasificadores son entrenados con documentos previamente etiquetados con el fin de clasificar una colección de datos. El otro grupo, denominado de orientación semántica o no supervisado, analiza clases de palabras, como adjetivos, adverbios, entre otros, y si esas palabras tienden a ser positivas o negativas, con el fin de clasificar las opiniones. De acuerdo con (Mukras 2009), el principal factor que distingue estos dos grupos es que el supervisado tiene acceso a ejemplos de datos ya etiquetados. Otros autores han seguido una clasificación similar a esta, donde asocian al grupo de aprendizaje supervisado a los algoritmos de aprendizaje automático, y al grupo de aprendizaje no supervisado, el enfoque basado en recursos léxicos (Peng & Shih 2010; Schrauwen 2010). Boiy y colectivo de autores (Boiy & Moens 2009) presentan una clasificación similar a las anteriores, determinando dos técnicas para la clasificación de sentimiento: simbólico y de aprendizaje automatizado. Según el autor, el enfoque simbólico usa reglas manualmente construidas y recursos léxicos. Las técnicas de aprendizaje automatizado utilizan métodos no supervisados, semi-supervisados o supervisados para construir un modelo de datos a partir de un extenso conjunto de entrenamiento.

Podemos notar que varias terminologías han sido usadas para describir las técnicas de clasificación de sentimiento. A pesar de utilizar clasificaciones diferentes ellas apuntan al uso de métodos o algoritmos semejantes. En este trabajo adoptaremos la clasificación citada por (Liu 2010b; Yessenov & Misailovic 2009; Chaovalit & Zhou 2005; Mukras 2009; NG 2010), o sea, hacemos la distinción entre técnicas no supervisadas que no hacen uso de documentos etiquetados y técnicas supervisadas que adoptan algoritmos de aprendizaje supervisado.

1.4.1 Clasificación supervisada

Diversos trabajos se destacan en cuanto al empleo de clasificadores basados en técnicas de aprendizaje supervisado para la detección de la polaridad. Los sistemas que utilizan estas técnicas en su mayoría clasifican los comentarios en dos clases (positiva o negativa) (Poirier et al. 2011). Algunos sistemas parten de un conjunto de entrenamiento donde cada opinión tiene

una nota o clasificación de 1 a 5 estrellas (Pontiki et al. 2014; Bespalov et al. 2011; Moghaddam & Ester 2013).

Primeramente, es realizada la colecta de una colección de documentos y se realiza un etiquetado sobre esta, donde estos documentos deben representar una muestra ideal para el entrenamiento, con el fin de alcanzar una excelente precisión para la clasificación. Después de seleccionada una muestra adecuada de documentos, o sea que abarca la ocurrencia de posibles términos existentes en la colección de documentos, el próximo paso es entrenar un clasificador con ese conjunto de entrenamiento. Este entrenamiento es un proceso iterativo con el objetivo de producir un mejor modelo. A continuación, el desempeño del clasificador entrenado es analizado sobre un conjunto de datos de pruebas (no etiquetados) y repetido varias veces en caso del que el resultado no sea satisfactorio (Chaovalit & Zhou 2005).

Los clasificadores parten de colecciones textuales representadas predominantemente usando n-gramas, bolsas de palabras y etiquetas de partes del discurso (Cabral Cavalcanti 2011). Los clasificadores más utilizados son SVM (Dave et al. 2003; Chesley et al. 2006; Kennedy & Inkpen 2006; Annett & Kondrak 2008; Zhang et al. 2008), redes Bayesianas (Gamon et al. 2005; Ferguson et al. 2009; Tan et al. 2009; Weichselbraun et al. 2010), máxima entropía (Pang et al. 2002; Schrauwen 2010), árboles de decisión (Annett & Kondrak 2008; Zhang et al. 2008; Gryc & Moilanen 2014; Schrauwen 2010) y redes neuronales (Chen & Chiu 2009; Tanawongsuwan 2010).

(Pang et al. 2002) es uno de los trabajos más citados y ofrece un estudio comparativo entre clasificadores bayesianas, los que usan máxima entropía y SVM para examinar la eficacia de la aplicación de técnicas de aprendizaje supervisado para el problema de la clasificación de la polaridad. Los autores extrajeron las opiniones de la fuente de datos del The Internet Movie Database (IMDb)[4] que contiene críticas de películas. Fueron seleccionados apenas los comentarios que estaban etiquetados con evaluación de estrellas o puntuaciones brindadas por el propio usuario. Las puntuaciones fueron automáticamente extraídas y convertidas en una de las tres categorías: positiva, negativa o neutra (en este trabajo los autores solo utilizaron las categorías positiva y negativa). Pang y colectivo de autores (Pang et al. 2002) produjeron una colección datos compuestos de 1301 revisiones positivas y 752 negativas. Para evitar el

[4] http://www.imdb.com/

dominio de un mismo autor o de pocos autores en las revisiones, limitaron los comentarios hasta 20 por autor, lo cual propició un total de 144 autores. La colección fue representada siguiendo la representación del texto en bolsa de palabras con sus respectivas frecuencias de aparición en cada comentario y consideraron la presencia de términos, unigramas, bigramas y etiquetas de las partes del discurso.

Tabla 1.1 Desempeño de los clasificadores en los experimentos de (Pang et al. 2002)**.**

	Representación	**# de atributos**	**Frecuencia o presencia**	**NB**	**ME**	**SVM**
1	Unigramas	16165	frecuencia	**78.7**	N/A	72.8
2	Unigramas	16165	presencia	81.0	80.4	**82.9**
3	Unigramas + Bigramas	32330	presencia	80.6	80.8	**82.7**
4	Bigramas	16165	presencia	77.3	**77.4**	77.1
5	Unigramas + POS	16695	presencia	81.5	80.4	**81.9**
6	Adjetivos	2633	presencia	77.0	**77.7**	75.1
7	Top 2633 unigramas	2633	presencia	80.3	81.0	**81.4**
8	Unigramas + posición	22430	presencia	81.0	80.1	**81.6**

La **Tabla 1.1** ilustra la precisión de la clasificación resultante para los clasificadores y representación textual adoptada. Los resultados en negrita representan los mejores resultados para cada categoría. Sus experimentos mostraron que los algoritmos de aprendizaje automatizado supervisados exhiben buenos resultados para la tarea de la clasificación de la polaridad en comparación con la clasificación manual realizada. El método basado en redes bayesianas obtuvo el 77,3% de precisión usando bigramas, pero mostró mejores resultados con unigrams y etiquetas de las partes del discurso. Para el método máxima entropía, se obtuvieron mejores resultados usando unigramas y bigramas con un 80,8% de precisión. Con SVM se lograron los mejores resultados, con un 82,9% de precisión usando unigramas. La representación con presencia o ausencia de términos demostró ser más informativa que la representación de la frecuencia de términos. Los experimentos también demostraron que la presencia de unigramas presenta una mejor eficacia. En términos de desempeño el clasificador bayesiano tiende a retornar los peores resultados, mientras que SVM presenta los mejores resultados, aunque las diferencias no son muy grandes (Pang et al. 2002).

Otros estudios que se destacan para la clasificación de sentimientos utilizando aprendizaje supervisado son (Xu 2010; Milidiú 2006), en los cuales se demostró que los datos entrenados para un dominio no pueden ser reusados para otro domino. También están (Cambria et al.

2013; Poria et al. 2014; Ghiassi et al. 2013; Singh, Piryani, Uddin & Waila 2013; Balahur & Turchi 2014; dos Santos & Gatti 2014; Moraes et al. 2013).

Las propuestas existentes para la clasificación de la polaridad mediante el empleo de técnicas de aprendizaje supervisado tienen una alta efectividad, pero tienen varias desventajas que limitan significativamente su uso, entre ellas: son dependientes del domino y del idioma, y requieren de un extenso y eficiente conjunto de datos etiquetados para el entrenamiento. Teniendo en cuenta las desventajas antes mencionadas, en este trabajo seguiremos enfoques no supervisados para la clasificación de la polaridad de las opiniones ya que no dependen del domino, son fácilmente escalables a otros lenguajes y no necesitan de un conjunto de entrenamiento previamente etiquetado.

1.4.2 Clasificación no supervisada

La clasificación de la polaridad siguiendo un enfoque no supervisado no requiere de un conjunto de opiniones previamente etiquetadas y se basada en la orientación semántica de los términos extraídos por medio de heurísticas lingüísticas, o de un léxico de palabras pre-seleccionadas. Uno de los grandes desafíos de este enfoque consiste en obtener la orientación semántica de términos, frases o múltiples expresiones, las cuales pueden ser clasificadas como positivas o negativas. Por ejemplo, excelente es positiva, mientras que malo es negativa. La orientación puede variar de acuerdo con el contexto (Hatzivassiloglou & Wiebe 2000).

De acuerdo con (Liu 2010b), las palabras y las oraciones son indicadores dominantes para la clasificación de sentimiento, y un posible método no supervisado puede consistir en realizar una clasificación basada en frases sintácticas fijas. Para (Yessenov & Misailovic 2009), aplicar técnicas de aprendizaje no supervisado es atribuir una clasificación basada en distancias entre puntos dados. Chaovalit y Zhou consideran en su trabajo (Chaovalit & Zhou 2005) el enfoque de orientación semántica para la minería de opinión, la clasificación se basa en medir cuánto un término tiende a ser positivo o cuánto tiende a ser negativo.

En la mayoría de los trabajos que siguen este enfoque se emplean etiquetadores de partes del discurso (Zhu et al. 2014; Amores, Arco & Artiles 2015a), algunos métodos utilizan la técnica Pointwise Mutual Information (PMI) para auxiliarse en la determinación de la orientación semántica de los documentos (Becker et al. 2013; Zhu et al. 2014; Singh, Piryani, Uddin, Waila, et al. 2013), y otros adoptan el uso de recursos léxicos para la clasificación (Amores,

Arco & Artiles 2015a; Xianghua et al. 2013; Abdul-Mageed & Diab 2014; Hogenboom et al. 2013). Por tanto, los trabajos publicados utilizan diversas estrategias para realizar la clasificación de la polaridad de manera no supervisada.

Uno de los trabajos más citados que usan PMI es el publicado en (Turney 2002). Turney (Turney 2002) presentó un algoritmo no supervisado para clasificar comentarios como recomendados (thumbs up) o no recomendados (thumbs down). La clasificación es dada por la media de la orientación semántica de las frases en los textos que contengan adjetivos y adverbios. La orientación semántica se calcula empleando el algoritmo PMI-IR que usa el método PMI y técnicas de recuperación de información (Information Retrieval; IR) para medir la semejanza entre palabras u oraciones. La orientación semántica de una determinada frase se calcula comparando la similitud de una palabra de referencia positiva (excelente) con la similitud de una palabra de referencia negativa (pobre), o sea, una frase es marcada con un valor numérico dado por la información mutua entre la frase y la palabra de referencia positiva substrayendo la información mutua entre la frase y la palabra de referencia negativa, esa clasificación numérica también indica la intensidad de la orientación semántica (Turney 2002). Este algoritmo fue evaluado en 410 comentarios extraídos de Epinions[5]. Los datos fueron extraídos aleatoriamente de cuatro dominios diferentes: automóviles, bancos, filmes y destinos turísticos, pertenecientes a diferentes autores.

Tabla 1.2 Patrones de etiquetas POS para la extracción de dos palabras

Primera Palabra	Segunda Palabra	Tercera Palabra
Adjetivo	Sustantivo (Singular o Plural)	----------
Adverbio, adverbio comparativo o adverbio superlativo	Adjetivo	No ocurrencia de sustantivo (Singular o Plural)
Adjetivo	Adjetivo	No ocurrencia de sustantivo (Singular o Plural)
Sustantivo (Singular o Plural)	Adjetivo	No ocurrencia de sustantivo (Singular o Plural)
Adverbio, adverbio comparativo o adverbio superlativo	Verbo (Infinitivo, Pretérito, Participio Pretérito o Gerundio)	----------

[5] http://www.epinions.com/

El algoritmo propuesto en (Turney 2002) tiene tres pasos fundamentales. En el primer paso se usa un etiquetador de partes del discurso para identificar frases que contengan adjetivos o adverbios. La motivación para esto es que algunas investigaciones demostraron que adjetivos y adverbios son buenos indicadores de subjetividad y opinión. Un adjetivo aislado puede indicar subjetividad, pero puede ser insuficiente para determinar la orientación de una opinión. Dos palabras consecutivas se extraen siempre y cuando la tercera palabra consecutiva no sea un sustantivo, como se describe en la **Tabla 1.2**.

El segundo paso consiste en estimar la orientación de las oraciones extraídas usando el método PMI mediante el cálculo de la probabilidad de co-ocurrencia de dos términos y la probabilidad de que los términos co-ocurran si son estadísticamente independientes. La orientación semántica de una oración se calcula a partir de la asociación de ésta con una palabra de referencia positiva y negativa. Las probabilidades se calculan a través de consultas en un motor de búsqueda. Para cada consulta, un motor de búsqueda generalmente devuelve una cantidad de documentos relevantes considerados hits. Así es posible estimar las probabilidades de ocurrencia de términos. El tercer paso consiste en calcular la media de la orientación semántica para todas las oraciones en un comentario y clasificar el texto como recomendado, si la orientación semántica es positiva, o no recomendado, si la orientación semántica es negativa.

Entre las limitaciones de esta investigación está que consume un gran tiempo para la ejecución, el desempeño del algoritmo depende del tamaño de la colección de documentos que son indexados y del lenguaje de consulta del motor de búsqueda, no obstante, es un algoritmo simple de implementar y no solamente restrictivo a adjetivos (Turney 2002). El algoritmo alcanzó una precisión promedio del 74% y la menor precisión fue del 66% para el dominio de películas. Los mejores valores de precisión se obtuvieron para los dominios acerca de bancos y automóviles, con valores entre el 80% y el 84% (Turney 2002).

Popescu y colectivo de autores presentaron OPINE (Popescu et al. 2005), un sistema no supervisado para la extracción de información de comentarios de productos. Este sistema está compuesto por las siguientes subtareas: identificar características explícitas e implícitas de productos, identificar opiniones sobre las características de los productos localizados, determinar la polaridad de cada opinión y clasificar la intensidad de ésta. El sistema OPINE fue construido a partir de una herramienta de extracción de información independiente del

dominio llamada KnowItAll [6] que adopta el algoritmo PMI para detectar relaciones candidatas entre las oraciones a partir de un patrón de reglas. El sistema utiliza también la herramienta MINIPAR (Lin 1999), que contiene recursos para etiquetar partes del discurso, lematizar, entre otros para evaluar los comentarios.

Otro trabajo que clasifica la polaridad de manera no supervisada es el publicado en (Abbasi et al. 2013). En esta propuesta se seleccionan las palabras relevantes para la clasificación de sentimiento de comentarios de productos escritos en chino. La orientación del sentimiento es dada calculando la diferencia entre las secuencias de términos positivos y negativos. Sus experimentos obtuvieron resultados semejantes a los métodos supervisados llegando a obtener una precisión del 92%. Otra propuesta especialmente diseñada para el idioma chino fue presentada por (Peng & Shih 2010) y alcanza una eficacia del 80%. Rothfels y Tibshirani (Rothfels & Tibshirani 2010) adaptaron el trabajo de (Abbasi et al. 2013) para la lengua inglesa en el dominio de filmes, y sorprendentemente solo alcanzaron una del 65.5%.

Entre las limitaciones de este enfoque está que consume un gran tiempo para la ejecución, el desempeño del algoritmo depende del tamaño de la colección de documentos que son indexados y del lenguaje de consulta del motor de búsqueda.

Teniendo en cuenta las desventajas de las propuestas presentadas anteriormente, varios investigadores se dieron a la tarea de clasificar la polaridad de las opiniones basándose en recursos léxicos (Hung & Lin 2013; Martín-Wanton et al. 2010; Xu 2010; Hamdan et al. 2013; Cernian et al. 2015). En (Martín-Wanton et al. 2010) desarrollaron un método que se basa en que una misma palabra puede no tener la misma polaridad en contextos diferentes. El autor adopta un algoritmo para reducir el sentido ambiguo de términos y localizar el sentido y la polaridad correcta de palabras a través de recursos como: Wordnet, una base de datos lexical donde las palabras son agrupadas en conjuntos de sinónimos; SentiWordNet, un recurso léxico para la minería de opinión, donde cada término en su base contiene un valor positivo, negativo y objetivo y el General Inquirer (GI)[7], un diccionario en Inglés que contiene palabras apuntadas en las categorías positiva, negativa o negación. Palabras con la categoría negación son conocidas como palabras modificadoras o negadoras, son términos que pueden alterar la

[6] http://www.cs.washington.edu/research/knowitall/
[7] http://www.wjh.harvard.edu/~inquirer/

orientación de otro término, por ejemplos: no, nunca, nada, ninguno. En el epígrafe 2.1 detallaremos las características del SentiWordNet y el General Inquirer.

El método de (Martín-Wanton et al. 2010) se caracteriza por reducir la ambigüedad para determinar la polaridad. Para reducir la ambigüedad utilizan un desambiguador del sentido de la palabra (Word Sense Disambiguation; WSD), que consiste en seleccionar el significado más adecuado para una palabra de acuerdo con el contexto en que ocurre. El significado de los términos se obtiene del Wordnet, y posteriormente se aplica un algoritmo de agrupamiento para identificar grupos de palabras de sentidos cohesionados. Posteriormente, los grupos que coinciden con el contexto son seleccionados. Después de obtener el sentido correcto de cada palabra de la opinión, se define la polaridad utilizando el SentiWordNet asociado con las categorías del GI. La polaridad de la opinión se calcula sumando la polaridad de las palabras positivas y negativas, en caso de que la sumatoria global de las palabras positivas sea mayor, la opinión es considerada positiva, en caso contrario es negativa, y si ambas sumatorias retornasen valores iguales, la opinión es considera neutra.

Para evaluar la técnica propuesta por (Martín-Wanton et al. 2010) fue utilizado un conjunto de datos compuesto de 1000 titulares de noticias obtenidas de periódicos. Fueron localizadas 4279 palabras de las cuales 3275 son ambiguas. La aplicación del método a la colección arrojó un 44.3% de exactitud (Accuracy), un 37.66% de precisión, y los valores recall y F1 fueron 72.11% y 49.41%, respectivamente. En (Martín-Wanton et al. 2010) confirmaron que la hipótesis de reducir ambigüedad del sentido de las palabras es útil para determinar la polaridad de los términos, obteniéndose resultados que superan algunos clasificadores supervisados y detectores de polaridad no supervisados.

En (Xu 2010) se presentó un método basado en recursos léxicos que aplica el algoritmo Promedio del Valor Sentimental (Average Sentimental Value; ASV), el cual construye funciones basadas en recursos fortalecidos por un léxico de sentimiento, tales como puntuaciones negativas o positivas a términos, para calcular la polaridad de comentarios. La polaridad de un texto se obtiene comparando la media de puntuaciones positivas con la media de puntuaciones negativas de términos. La polaridad de la revisión es dada por la media de los términos que obtuvieron mayor puntuación.

En (Xu 2010) se realizaron experimentos considerando 200 comentarios negativos y 200 positivos de cada uno de los dominios: libro, DVD, reproductor mp3 y video juegos. Fueron

seleccionadas también 400 opiniones positivas y 400 negativas del dominio restaurantes. La mayor precisión fue obtenida en el dominio de restaurantes con 72.30% y la menor precisión en el dominio de libros con 57.8%. En los dominios de DVD y juegos obtuvieron precisión de 62.8% y 62.5%, respectivamente.

En (Miranda et al. 2016) se muestra la construcción de un sistema de minería de opiniones en español sobre comentarios dados por clientes de diferentes hoteles. El sistema trabaja bajo el enfoque léxico utilizando la adaptación al español de las normas afectivas para las palabras en inglés (ANEW). Estas normas se basan en las evaluaciones que se realizaron en las dimensiones de valencia, excitación y el dominio. Para la construcción del sistema se tuvo en cuenta las fases de extracción, preprocesamiento de textos, identificación del sentimiento y la respectiva clasificación de la opinión utilizando ANEW. Los experimentos del sistema se hicieron sobre un corpus etiquetado proveniente de la versión en español de Tripadvisor. Como resultado final se obtuvo una precisión del 94% superando a sistemas similares. Cabe destacar la utilización de varios algoritmos complementarios para el cálculo de la polaridad. El uso de herramientas de PLN potentes como FREELING (Padró & Stanilovsky 2012) permite la construcción de sistemas de minería de opiniones muy robustos. Además, en (Muñiz-Cuza & Ortega-Bueno 2016) se presenta un método para la clasificación de la polaridad de aspectos de productos. La característica más relevante de la propuesta radica en la construcción automática de recursos de polaridad dependientes del dominio a través del empleo de la técnica de Análisis de Semántica Latente y de manera concreta se generaron estos recursos para los dominios de laptop y restaurante utilizando *n*-gramas de términos (*n* en el rango de 1 a 3) y *n*-gramas sintácticos no continuos (*n* en el rango de 2 a 3). Con excepción del recurso generado para los trigramas sintácticos, los recursos demostraron generar información valiosa que puede ser tomada en cuenta para la clasificación de la polaridad de los aspectos. El método permite generar recursos de polaridad para varias unidades textuales independiente del idioma. La clasificación de la polaridad de los aspectos se realiza en dos fases fundamentales: extracción de las palabras y frases de opinión, y la clasificación de la polaridad. La primera etapa consiste en extraer de un contexto lineal y sintáctico relacionado con el aspecto las unidades textuales para las cuales fueron generados los recursos de polaridad. Finalmente, la polaridad del aspecto, en una crítica dada, es determinada por los valores de polaridad positivo y negativo de cada una de las palabras y frases de opinión extraídas. Los resultados obtenidos

por la propuesta son alentadores si se considera que el proceso de construcción de los recursos se realiza completamente de manera automática. Este método alcanza valores de exactitud en un rango de 64,98% a un 73,63%.

En (Khoo et al. 2015) se presenta un nuevo lexicón de sentimiento de propósito general llamado WKWSCI y se compara con otros tres lexicones existentes. Este lexicón cubre adjetivos, adverbios y verbos. Las palabras fueron codificadas manualmente con un valor de intensidad del sentimiento en una escala de siete puntos. La eficacia de los cuatro recursos léxicos para la clasificación de la polaridad a nivel de documento y a nivel de oración se evaluó utilizando un conjunto de datos de cometarios sobre productos de Amazon[8]. El léxico WKWSCI obtuvo los mejores resultados para el nivel de documento, con una precisión del 75%. El léxico de Hu y Liu obtuvo los mejores resultados para el nivel de frase, con una precisión del 77%. El mejor modelo de aprendizaje automático basado en bolsas de palabras obtuvo una precisión del 82% para el nivel de documento.

Como se describió anteriormente, son diversos los métodos que siguen un enfoque no supervisado. Estos métodos han reportado buenos resultados y son esenciales en situaciones en que no se tiene un conjunto de datos previamente etiquetado. Otra de las ventajas de estos métodos es que tienen menor dependencia del dominio, problema este normalmente asociado a métodos supervisados por mantener un conjunto fijo de datos etiquetados (Abbasi et al. 2013). Algunas propuestas no supervisadas tienen las siguientes limitaciones: consumen un gran tiempo para la ejecución, el desempeño del algoritmo depende del tamaño de la colección de documentos que son indexados y del lenguaje de consulta del motor de búsqueda; por ello, en este trabajo asumimos la clasificación no supervisada, con el empleo de recursos léxicos para la clasificación de la opinión. En el próximo capítulo abordaremos sobre algunos recursos léxicos disponibles en la literatura para la clasificación de sentimientos.

1.5 Consideraciones finales del capítulo

La minería de opinión o llamado análisis de sentimientos se puede realizar en diferentes niveles, siendo el nivel de aspectos el que impone mayores retos en el análisis de sentimientos, mientras que los niveles de frases y documentos requieren menos esfuerzos al aplicar las

[8] www.amazon.com

técnicas de procesamiento del lenguaje natural. La minería de opinión a nivel de documentos y frases no descubre lo que exactamente las personas les gusta o no. El nivel de aspecto lleva a cabo el análisis con mayor granularidad; ya que extrae los atributos de la entidad de la cual se expresa una opinión y la orientación de dicha opinión.

Existen diversas tareas del análisis de sentimientos, entre ellas: la detección de una opinión, la detección de subjetividad, la clasificación de la opinión, la determinación de la fuerza de la opinión, determinación de la fuente de la opinión, determinación del objetivo de la opinión y resumen de las opiniones y/o visualización de los resultados. Considerando que la clasificación de la opinión se considera el paso más importante en el análisis de sentimientos, resulta de interés para el desarrollo de este trabajo la clasificación de la opinión, así como la determinación de la fuerza de la opinión, ya que mientras que la asignación del sentimiento se ocupa de analizar si un texto tiene una orientación semántica positiva, negativa o neutra, la intensidad del sentimiento trata con el análisis de si los sentimientos positivos o negativos son suaves o fuertes.

Al determinar la polaridad de las opiniones debemos ser capaces de manejar algunos problemas presentes en las opiniones. Muchos de ellos debido al suministro de la opinión, el objetivo de la opinión, y también las expresiones críticas o comentarios creados por los usuarios. Algunos de los problemas presentes son: una palabra que se considera que es positiva en una situación puede ser considerada negativa en otra situación por lo que es necesario tener en cuenta el contexto, las personas pueden ser contradictorias en sus declaraciones, algunas veces las personas tienen dificultades para entender lo que alguien pensó a partir del análisis de un pequeño fragmento de texto porque carece de contexto, la capitalización se puede utilizar con sutileza para denotar el sentimiento, las opiniones pueden aparecer en diferentes idiomas, ciertas palabras exhiben diferentes polaridades cuando se utilizan en diferentes dominios, el uso de abreviaturas y jergas, la presencia de palabras negadoras y la presencia de ironía y sarcasmo. Por tanto, ser capaces de manejar automáticamente estas situaciones que inevitablemente estarán presentes en las opiniones constituye un desafío y a pesar de que existen varios trabajos dirigidos a resolver tales problemas, aun es necesario seguir desarrollando este tema.

Las propuestas existentes para la clasificación de la polaridad mediante el empleo de técnicas de aprendizaje supervisado tienen una alta efectividad, pero tienen varias desventajas que

limitan significativamente su uso, entre ellas: son dependientes del domino y del idioma, y requieren de un extenso y eficiente conjunto de datos etiquetados para el entrenamiento. Teniendo en cuenta las desventajas antes mencionadas, en este trabajo seguiremos enfoques no supervisados para la clasificación de la polaridad de las opiniones ya que no dependen del domino, son fácilmente escalables a otros lenguajes y no necesitan de un conjunto de entrenamiento previamente etiquetado. Algunas limitaciones de este enfoque es que consume un gran tiempo para la ejecución, en el caso de usar PMI-IR, el desempeño del algoritmo depende del tamaño de la colección de documentos que son indexados y del lenguaje de consulta del motor de búsqueda y en el caso de los que usan recursos léxicos dependen de la calidad de estos, que la mayoría están enfocados para el idioma inglés, aquellos que fueron anotados automáticamente contienen muchos errores y los creados manualmente recogen muy pocos términos. Teniendo en cuenta las desventajas de estas variantes, se clasificará la polaridad de las opiniones basándose en nuevos recursos léxicos.

Capítulo 2 Esquemas para detectar la polaridad de las opiniones basados en nuevos recursos léxicos

En este capítulo se comentarán las características principales de los recursos léxicos existentes y las ventajas y desventajas que estos presentan. Se propondrán nuevos recursos léxicos inspirados en los ya existentes e intentando suplir las carencias de los precedentes. Dos esquemas generales para la detección no supervisada de la polaridad de las opiniones serán presentados, uno de ellos incorpora el uso de nuevos recursos diseñados para el tratamiento de algunos de los problemas que presentan las opiniones. Finalmente, es descrito el sistema PosNeg Opinion 3.0 que fue diseñado siguiendo los esquemas propuestos para detectar la polaridad de opiniones.

2.1 Recursos léxicos

De acuerdo con (Esuli 2008), uno de los factores principales para cualquier trabajo relacionado con la minería de opinión es la necesidad de identificar cuáles elementos del lenguaje contribuyen a expresar subjetividad en un texto. Esa identificación puede ser realizada a través de recursos léxicos que listan propiedades relevantes relacionadas a los ítems de opinión. Los ítems en un recurso léxico pueden ser palabras únicas o secuencias de palabras.

Según (Pasqualotti 2008), un recurso léxico es una estructura sistemática que explícitamente expresa determinadas características o significados asociados a las palabras. Cada término de un recurso léxico consiste de una forma ortográfica y fenológica con un formato de representación del significado.

En el contexto de la minería de opinión, los recursos léxicos han sido construidos o adaptados con la intención de aplicarlos en investigaciones para la detección de la subjetividad en textos y la clasificación de sentimiento. Algunos recursos son compuestos por un conjunto de términos, relacionados a las categorías, por ejemplo, positiva o negativa. Posiblemente los términos que no están relacionados a una de estas dos categorías pueden ser considerados como objetivos y no contribuir a la orientación global del texto (Esuli 2008).

Las categorías para recursos léxicos aplicados a la minería de opinión, en algunos casos, no se restringen solamente a la polaridad de palabras, sino también asocian otros atributos relacionados a la afectividad. En (Ohana & Tierney 2009) citan la terminología de léxicos de

opinión (Opinion Lexicons), para describir recursos que asocian orientación de sentimiento y palabras, basados en que los términos individuales pueden ser considerados como una unidad de información de opinión, y pueden sugerir subjetividad y sentimientos en textos. En (Yi et al. 2003) describen léxico de sentimiento (Sentiment Lexicon) como un recurso que contiene la definición del sentimiento de las palabras individuales, compuesto por la entrada léxica (es decir, uno o más términos que poseen connotación sentimental), la etiqueta POS y la polaridad léxica de la entrada. En (Remus et al. 2010) describen léxico afectivo (Affect Lexicon) como un recurso que se compone de un sumario de entradas léxicas para las palabras afectivas (cualquier expresión que tenga asociada una connotación afectiva) con su etiqueta POS correspondiente, categorías afectivas (polaridad, apreciación, juicio, afecto, etc.) e intensidad (fuerte, medio, débil).

Según (Valitutti et al. 2004), para organizar un léxico afectivo, es necesario obtener un modelo de conocimiento afectivo que represente emociones, humor y actitudes. A partir de este modelo es posible identificar un gran número de conceptos afectivos, organizarlos en una estructura jerárquica y conectarlos a través de un léxico.

Varios recursos léxicos han sido creados para ser usados en el análisis de sentimientos. Algunos fueron construidos manualmente, por ejemplo, en (Subasic & Huettner 2001) construyeron manualmente un recurso léxico asociando palabras con categorías afectivas, especificando la intensidad para la afectividad y el grado de relación con la categoría (Mejova 2009). Para (Ohana & Tierney 2009), los recursos léxicos construidos manualmente tienden a tener un número limitado de condiciones, pues construir listas manuales es un esfuerzo que consume tiempo y está sujeto a notaciones tendenciosas. Para superar problemas de extraer de forma automática las correspondencias léxicas de grupos de términos (inducción léxica), algunos trabajos fueron propuestos con la intención de extender el tamaño de los recursos léxicos a partir de términos considerados relevantes, explorando sus relaciones y similitudes (Garrido Merchán 2015; Cruz et al. 2011; Yu et al. 2013; Neviarouskaya et al. 2011).

Uno de los primeros trabajos en esta área fue propuesto por (Hatzivassiloglou & McKeown 1997), denominado HM lexicon. Esta investigación evalúa la orientación semántica de dos adjetivos que estén entre una conjunción, extraídos de un conjunto de 21 millones de palabras de artículos del Wall Street Journal. El HM lexicon consiste de una lista de 1336 adjetivos etiquetados, de los cuales 657 son positivos y 679 negativos. En (Hatzivassiloglou &

McKeown 1997) desarrollaron su método a partir de algoritmos de aprendizaje supervisado y no supervisado para inferir la orientación semántica de los adjetivos.

Otro enfoque común son los recursos construidos extendiendo recursos léxicos, para fines generales, ya existentes, normalmente evaluando las relaciones semánticas, tales como sinónimos y antónimos, de un término (Ohana & Tierney 2009). El WordNet de la Universidad de Princeton es uno de los recursos léxicos más populares adoptado para el análisis de sentimiento (Mejova 2009). A continuación, describimos el WordNet y otros recursos léxicos que localizamos para la clasificación de sentimiento.

WordNet:

El WordNet es una extensa base de datos léxica de la lengua inglesa, desarrollada por George A. Miller, a partir de la combinación de informaciones lexicográficas (relaciones léxicas y semánticas usadas para representar la organización del conocimiento léxico) y, recursos computacionales (Fellbaum 1998). En (Miller 1995) define el vocabulario de un lenguaje con un conjunto de pares palabra-sentido a partir de un conjunto de significados. En el WordNet un sentido es representado por un conjunto de uno o más sinónimos. La base posee más de 118000 palabras diferentes y más de 90000 sentidos (Miller 1995).

El WordNet está compuesto de sustantivos, verbos, adjetivos y adverbios agrupados en conjuntos de sinónimos (synsets), cada uno expresando un concepto distinto. Estos son organizados en un conjunto de lexicógrafos por categoría sintáctica y por otros criterios de organización. Los adverbios son mantenidos en apenas un archivo, mientras los sustantivos y verbos son agrupados de acuerdo con la semántica. Los adjetivos son divididos en adjetivos descriptivos y relacionales (Miller et al. 1990).

WordNet Affect:

A partir del WordNet fue desarrollado el recurso léxico WordNet Affect[9], un recurso lingüístico para la representación léxica de conocimiento afectivo de palabras, a través de la selección y anotación de synsets que representan conceptos afectivos. El WordNet Affect es parte de la base WordNet Domains[10], una extensión multilingüe (inglés e italiano) en el WordNet donde cada synset es apuntado en uno o más dominios (ejemplo, deporte, política,

[9] http://wndomains.fbk.eu/wnaffect.html, Fundazione Bruno Kessler (FBK-irst'09), 2009.
[10] http://wndomains.fbk.eu/, Fundazione Bruno Kessler (FBK-irst'09), 2009.

medicina). Un dominio puede incluir synsets de diferentes categorías sintácticas (Strapparava et al. 2004).

El WordNet Affect fue desarrollado en dos etapas. La primera etapa consistió en identificar synsets claves, nombrados affects. Esta etapa fue realizada manualmente con el objetivo de obtener un conjunto inicial de palabras afectivas, conteniendo 1903 términos, donde 539 son sustantivos, 517 adjetivos, 238 verbos y 5 adverbios. Inicialmente fueron colectados adjetivos con ayuda de diccionarios. Después, los sustantivos fueron adicionados a la base con la intención de correlacionarlos con los adjetivos y posteriormente fueron adicionados verbos y adverbios. Para cada ítem fue creado un cuadro para adicionar información léxica y afectiva. Informaciones léxicas incluyen correlación entre términos en inglés e italiano, etiquetas POS, definiciones, sinónimos y antónimos. La información afectiva corresponde a uno o más tipos de representación de emoción (ejemplo, comportamiento, personalidad, sentimiento, entre otros). La información de la base Affect fue sumada a los synsets deWordNet Affect, con una anotación afectiva denominada a-label; por ejemplo, el a-label "Emo" tiene la descripción "emoción" y ejemplos de synsets son "sustantivo anger#1, verbo fear#1" (Strapparava et al. 2004).

La segunda etapa consistió en seleccionar un subconjunto del WordNet conteniendo todos los synsets en que exista por lo menos una palabra en la lista de términos afectivos, donde las relaciones semánticas fueron utilizadas para alimentar el WordNet Affect. La versión 1.0 contiene el mapeado entre WordNet 1.6 y sus correspondientes dominios afectivos, un total de 2874 synsets y 4787 palabras (Strapparava et al. 2004; Pasqualotti 2008).

SentiWordNet:

Otro léxico existente, desarrollado por (Esuli 2008) explícitamente para colaborar en aplicaciones para la minería y clasificación de opiniones, es el SentiWordNet (SWN). Este recurso léxico, es resultado de anotaciones automatizadas en todos los synsets del WordNet, con grado de positividad, negatividad y neutralidad. Cada synset *s* es asociado a tres valores numéricos, Pos(*s*), Neg(*s*) y Obj(*s*) que indican cuan positivo, negativo u objetivo (neutro) son los términos contenidos en el synset *s*. Cada uno de los tres valores varían en el intervalo [0.0, 1.0] y la suma de ellos es 1.0 para cada synset.

Para construir los SWN1.0 (Esuli & Sebastiani 2006b) y 1.1 (Esuli & Sebastiani 2007) fueron empleados algoritmos de aprendizaje supervisado y semi-supervisados. Para obtener las

versiones 2.0 (Esuli & Sebastiani 2009) y 3.0 (Baccianella et al. 2010) los resultados de los algoritmos semi-supervisados fueron adoptados como una etapa intermedia en el proceso de etiquetamiento, y posteriormente se siguió un proceso iterativo de refinamiento de los resultados hasta lograr convergencia. La ambigüedad en el SWN 1.0 es tratada por la representación semántica de los synsets del WordNet. Para las demás versiones fueron adoptados recursos diseñados expresamente para manejar la ambigüedad de los términos. Así, el SWN 2.0 adoptó el eXtended WordNet[11] y el SWN 3.0 adoptó el Princeton WordNet Gloss Corpus[12], que asume ser más preciso que el eXtended WordNet.

El método usado para desarrollar el SentiWordNet es derivado de los trabajos de (Esuli & Sebastiani 2005; Esuli & Sebastiani 2006a), basado en el análisis cuantitativo de términos asociados a synsets y en el uso de la representación vectorial de términos resultantes para la clasificación semi-supervisada de synsets. El trío de puntuación en SWN se deriva de la combinación de los resultados producidos por un conjunto de ocho clasificadores ternarios, con similares niveles de precisión; sin embargo, con diferentes comportamientos de clasificación. Un clasificador difiere del otro teniendo en cuenta el conjunto de formación de partida y el aprendizaje adoptado para este conjunto, produciendo así distintas clasificaciones resultantes de cada synset del WordNet.

La puntuación en el SWN es dada por la proporción de clasificadores que señaló la correspondiente etiqueta al synset. Si todos los clasificadores ternarios resultan en atribuir la misma etiqueta a un synset, la etiqueta tendrá la máxima puntuación para el synset, de lo contrario tendrá puntuación proporcional al número de clasificadores que señaló (Esuli 2008; Baccianella et al. 2010). El SWN 3.0 posee versión web para consulta online[13]. La base de datos del SWN también está disponible en un archivo de datos (.txt). A continuación, se muestran algunos ejemplos de registros en SentiWordNet 3.0 siguiendo el formato:

POS	ID	PosScore	NegScore	SynsetTerms	Gloss
a	00016247	0.125	0.500	super abundant #1	most excessively abundant
a	00122245	0.375	0.125	prevenient #1 anticipatory #1	in anticipation
r	00124611	0.000	0.125	legally#2	in a legal manner; “he acted legally”

[11] http://www.hlt.utdallas.edu/~xwn/downloads.html
[12] http://wordnet.princeton.edu/glosstag.shtml
[13] http://sentiwordnet.isti.cnr.it/

El POS-ID identifica el synset, los valores PosScore y NegScore corresponden a la positividad y negatividad señalados por el SentiWordNet para el determinado synset. El valor de objetividad es dado por: Obj(s)=1-(Pos(s)+Neg(s)). SynsetTerms corresponde a los términos separados por espacio, pertenecientes al synset, con la clase gramatical y número correspondiente al sentido. Gloss describe el sentido del término.

SentiWordNet cubre todos los synsets de WordNet. Aunque este es quizás uno de sus mayores aportes, es también una de sus mayores debilidades, puesto que, en muchos casos, conceptos que no estén bien relacionados con el conjunto inicial de synsets etiquetados manualmente no obtendrán puntuaciones muy acordes al etiquetado esperado. Un ejemplo de esto es que el concepto representativo de "tumor" o "cáncer" obtiene una puntuación muy baja (0.125) en la categoría Neg.

El SentiWordNet fue construido semiautomáticamente; por tanto, todos los resultados no fueron validados manualmente, por lo que, algunas clasificaciones pueden ser incorrectas. Por ejemplo, el synset#1 del sustantivo "flu"({influenza, flu, grippe} - (*an acute febrile highly contagious viral disease*)) se clasificó como Positivo=0.75, Negativo=0.0, Objetivo=0.25, a pesar de tener varias palabras negativas en su glosario.

Micro –WNop:

Otro recurso léxico derivado del WordNet, es el llamado Micro-WNop[14]. Este recurso léxico consiste de 1105 synsets del WordNet 2.0 apuntados manualmente. Un grupo de cinco personas (denominadas J1,..., J5) etiquetó cada synset con tres valores Pos(s), Neg(s), y Obj(s), que suman 1. El total de synsets fue dividido en tres grupos para etiquetar: Micro-WNop (common), Micro-WNop (Group 1) y Micro-WNop (Group 2).

En el primer grupo, fueron apuntados los synsets del 1 al 110 (110 synsets) por todos los apuntadores trabajando juntos para desarrollar un entendimiento común de semántica de las tres categorías. En el segundo grupo, fueron apuntados los synsets del 111 al 606 (496 synsets) por J1, J2 y J3 independientemente. En el tercero grupo fueron apuntados los synsets del 607 al 1105 (499 synsets) por J4 y J5 independientemente (Baccianella et al. 2010; Cerini et al. 2007). La base del Micro-WNop está disponible en un único archivo texto con los valores separados por tabulación.

[14] http://www-3.unipv.it/wnop/

General Inquirer:

Más allá de recursos léxicos derivados del diccionario WordNet, localizamos otros léxicos que son aplicables a la minería de opinión. El General Inquirer (GI) es un diccionario de la lengua inglesa que contiene listas de palabras, clasificadas de acuerdo con categorías específicas (Taboada et al. 2006). El léxico fue desarrollado por Philip Stone y amigos, en la década del 60. Inicialmente fue proyectado como una herramienta para el análisis de contenido para identificar características específicas de mensajes. El léxico GI actual posee 11788 palabras distribuidas en 182 categorías.

En el GI las etiquetas de categorías atribuidas a las palabras son derivadas de las fuentes Harvard IV-4 Dictionary y Lasswell Value Dictionary, y también, son atribuidas nuevas categorías y un marcador de categorías para la reducción de ambigüedad. Entre las categorías, este recurso léxico posee palabras con la etiqueta Positiv con un total de 1915 palabras indicando orientación positiva, 2291 palabras etiquetadas con la etiqueta Negativ indicado orientación negativa, posee también un 217 palabras etiquetadas con Negate, que son términos conocidos de palabras negadoras o modificadoras. En la **Tabla 2.1** se muestra un ejemplo de registros y categorías del General Inquirer.

Tabla 2.1 Un fragmento de registros y categorías en el General Inquirer.

Entrada	**Fuente**	**Positiv**	**Negativ**	**Negate**
ADMIRATION	H4Lvd	Positiv		
CANNOT	H4Lvd			Negate
DISLIKE#1	H4Lvd		Negativ	Negate
DISLIKE#2	H4Lvd		Negativ	Negate
IMPOSSIBLE#1	H4Lvd			Negate
IMPOSSIBLE#2	H4Lvd			Negate
INEXPENSIVE	H4	Positiv		Negate
UNLIKE	H4Lvd			Negate
UNLIMITED	H4Lvd	Positiv		Negate
WORSE	H4Lvd		Negativ	

Las palabras que aparecen en más de una fila son porque tienen asociado más de un sentido. La fuente puede ser del diccionario Harvard (H4), Lasswell (Lvd) o ambos (H4Lvd). Las categorías que se atribuyen son: Positiv para indicar positividad, Negativ para indicar negatividad y Negate si la palabra representa una palabra negadora o modificadora. Además,

se adicionan la etiqueta OthTags que exhibe la clase gramatical y Defined que la exhibe la definición para el sentido de la palabra.

The Subjectivity Lexicon

Otro recurso disponible es el léxico llamado The Subjectivity Lexicon (Jain & Nemade 2010). Este diccionario contiene 8221 unidades léxicas con indicación o pista subjetiva (clue), algunas identificadas a partir de tareas desarrolladas manualmente y otras identificadas automáticamente, usando datos apuntados y no apuntados. Cada término es identificado con una intensidad de subjetividad (débilmente o fuertemente subjetivo), etiqueta POS y polaridad (positivo, negativo y neutro) (Vechtomova 2010). La base de datos está disponible en archivo texto (Vechtomova 2010), abajo listamos algunos registros ejemplificando el formato en que los datos están disponibles:

type=strongsubj len=1 word1=enflame pos1=verb stemmed1=y priorpolarity=negative
type=weaksubj len=1word1=engage pos1=verbstemmed1=y priorpolarity=neutral
type=strongsubj len=1word1=engaging pos1 =adj stemmed1=n priorpolarity=positive

El atributo type indica si la subjetividad es fuerte (strongsubj) o débil (weaksubj), len indica el tamaño de la pista subjetiva (clue) y, word1 indicada la palabra, pos1 indica la clase gramatical, el stemmed1 igual a "y" indica que el clue corresponde a todas las variantes de la palabra de acuerdo con la etiqueta POS y priorpolarity indica la polaridad (Jain & Nemade 2010).

Como podemos ver, están disponibles diversos recursos léxicos que pueden ser empleados en el contexto de la minería de opinión. Cada uno posee particularidades que difieren tanto en el número de términos analizados como en la cantidad de términos lexicales pertenecientes a cada base. Difieren también en cuanto a los tipos de clases gramaticales que componen cada recurso. Algunos léxicos extienden de otros léxicos como el SentiWordNet, Micro-WNop y el WordNet Affect. Y difieren tambiénen la forma de cómo los datos fueron etiquetados, donde se han empleado técnicas manualmente trabajadas o adoptado una combinación de reglas automatizadas como en el caso del SentiWordNet.

Reunimos en la **Tabla 2.2** los léxicos citados comparando la cantidad de términos y las clases gramaticales de cada base. El recurso Micro-WNop abarca las cuatro clases gramaticales principales, pero posee menor número de recursos lexicales, y el recurso HM lexicón abarca

solamente la clase gramatical de adjetivos. El léxico SentiWordNet, además de emplear las cuatro clases gramaticales, abarca el mayor número de términos.

Tabla 2.2 Comparación de los principales recursos léxicos para la minería de opinión.

Léxico	Términos	Clase gramatical
General Inquirer	11788 términos (+ 1915 / - 2291 / not 217)	sustantivo adjetivo verbo adverbio
HM léxicon	1336 términos	adjetivo
Micro-WNOp	1105 synsets	sustantivo adjetivo verbo adverbio
SentiWordNet	117374 synsets / 147306 términos	sustantivo adjetivo verbo adverbio
The Subjectivity Lexicon	8221 términos	sustantivo adjetivo verbo adverbio
WordNet Affect	2874 synsets / 4787 términos	sustantivo adjetivo verbo adverbio

Al investigar sobre posibles léxicos aplicables a la minería de opinión, adoptamos como criterio recursos léxicos que contengan la categoría polaridad (positivo y negativo) relacionada a los términos, pues el objetivo de este trabajo es clasificar opiniones en cuanto a su polaridad. También resulta de interés aquellos léxicos que tienen otras categorías relacionadas a la afectividad, porejemplo: General Inquirer y WordNet Affect, aplicables a la minería de opinión. Estos recursos son ricas fuentes de informaciones que pueden colaborar a una mejor identificación de la subjetividad, mejorando así la eficacia de métodos para la clasificación de la opinión. En este trabajo desarrollaremos métodos para la clasificación de la polaridad de la opinión utilizando un recurso léxico derivado del SentiWordNet. Mejoraremos este recurso léxico, que se destaca por poseer el mayor número de términos relacionados, además de abarcarlas cuatro clases gramaticales principales. Su versión más reciente es SWN 3.0, del año 2010 disponible en http://sentiwordnet.isti.cnr.it/.

2.2 Nuevos recursos léxicos

Para la creación de la nueva versión del SentiWordNet 3.0 se tuvieron en cuenta los problemas que tiene la versión original de esta herramienta en cuanto a formato y a las polaridades de los términos.

En su formato no todos los términos contienen una única palabra y algunos de ellos incluyen los signos _ y – para hacer referencia a frases y palabras compuestas, respectivamente. Este

análisis por frases o palabras compuestas es de gran utilidad para otro tipo de procesamiento del lenguaje natural, pero en la minería de opinión en la mayoría de los casos no aporta información adicional que la que brindan los términos por separado, y por el contrario, complejiza significativamente el procesamiento al tener que buscar parejas, ternas, y cuartetos de palabras. En lo que se refiere a las puntuaciones positivas y negativas de los términos, la mayoría de ellos habían sido puntuados con cero en ambas polaridades constituyendo este fenómeno un gran problema en la tarea de clasificación de la polaridad, objetivo esencial de este recurso. Estos errores se enmarcan en la Figura 2.1. Como se puede apreciar están enmarcados los términos adductive, adducting, adducente los cuales claramente tiene una polaridad negativa y sin embargo aparecen con puntuación 0 tanto positiva como negativa y tan bien pueden apreciar el caso de una frase como ready_to_hand.

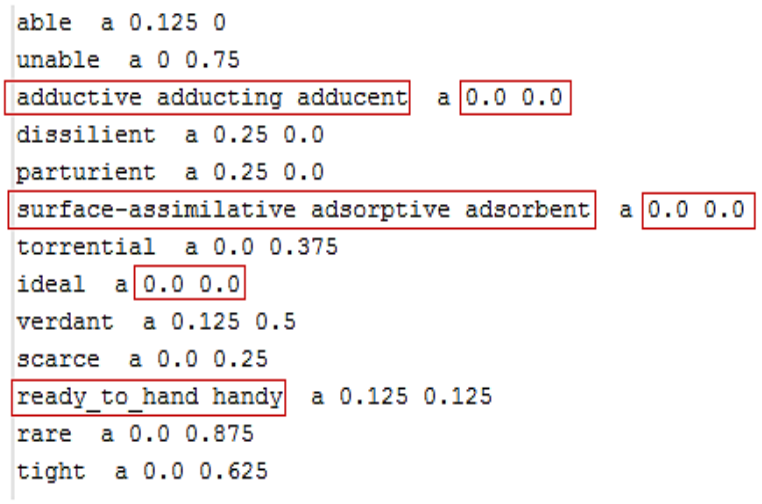


Figura 2.1. Fragmento de errores en las puntuaciones en SentiWordNet 3.0.

Para re-etiquetar la asignación de polaridades positiva y negativa a los términos en SentiWordNet se realizó un análisis de coincidencias entre los términos positivos y negativos del SentiWordNet (cuando se refiere a términos positivos o negativos del SentiWordNet, se escoge el que mayor puntuación tenga positiva o negativa, respectivamente) con dos listas de palabras positivas y negativas publicadas por Bing Liu[15]. Las listas de palabras positivas y

[15] http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html

negativas de Bing Liu cuentan con 2005 palabras positivas en cualquier contexto y con 4781 negativas, una comparación de términos entre el SWN y estas listas nos daría una proporción de cuán deficientes son las puntuaciones en el SWN. De este análisis se obtuvieron los resultados que se muestran en la **Tabla 2.3. Coincidencias positivas y negativas entre el SWN y las listas de palabras positivas y negativas de Bing Liu.Tabla 2.3**.

Tabla 2.3. Coincidencias positivas y negativas entre el SWN y las listas de palabras positivas y negativas de Bing Liu.

Total de términos del SWN 3.0	Coincidencias positivas	Coincidencias negativas	Total
117659	735	1175	1910

Estos resultados nos demuestran la gran inexactitud en las polaridades de los términos del SentiWordNet. Como parte de nuestra investigación pretendemos mejorar las asignaciones de puntuación positiva y negativa a los términos y proporcionalmente obtener mejores resultados en las coincidencias.

A continuación detallaremos las modificaciones realizadas en la herramienta SentiWordNet particularizando primero en las transformaciones en el formato y después en la asignación de polaridad.

2.2.1 Sentiwordnet 4.0

Como afirmábamos anteriormente el formato del SWN en algunos casos contiene más de una palabra por término, pero cuando se usa más de una palabra para identificar un término específico es porque dichas palabras son sinónimas, por lo que la solución a este problema sería sencillamente crear, a partir del término señalado, un término nuevo por cada una de las palabras y otorgarles a cada uno la misma etiqueta y polaridad. En la

Figura 2.2 se muestra el resultado final de esta modificación para *assimilatory assimilative assimilating.*

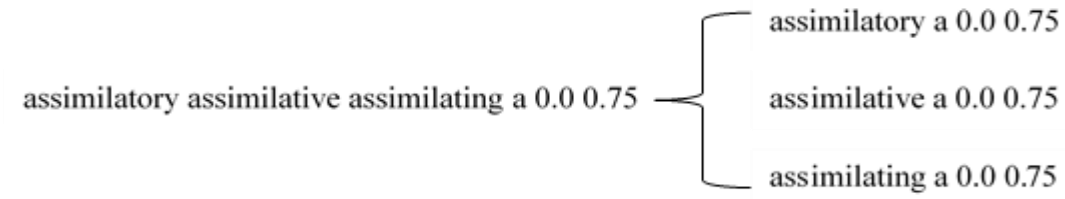


Figura 2.2. Resultado de modificar un término con más de una palabra.

Como ya comentamos, la presencia de frases y palabras complejas complejiza el análisis y no aporta beneficios, por lo que es necesario eliminar la presencia de los signos _ y – entre los términos. Primeramente se extrajeron los términos simples que conformaban las palabras compuestas o frases y se eliminaron aquellos que constituían stopwords, es decir, se eliminaron los términos que no ofrecen ningún significado en el idioma Inglés. Seguidamente, se buscaron las posibles etiquetas de cada uno de los términos, mediante la función *getPos*() de la biblioteca RitaWordNet[16], que dado un término como parámetro devuelve todas las posibles etiquetas que se ajustan al significado; creando así nuevos términos incluyendo los términos con sus respectivas etiquetas y con valores de polaridad positiva y negativa. La **Figura 2.3** muestra cómo se realizó el proceso anterior para la frase *wide_of_the_mark.*

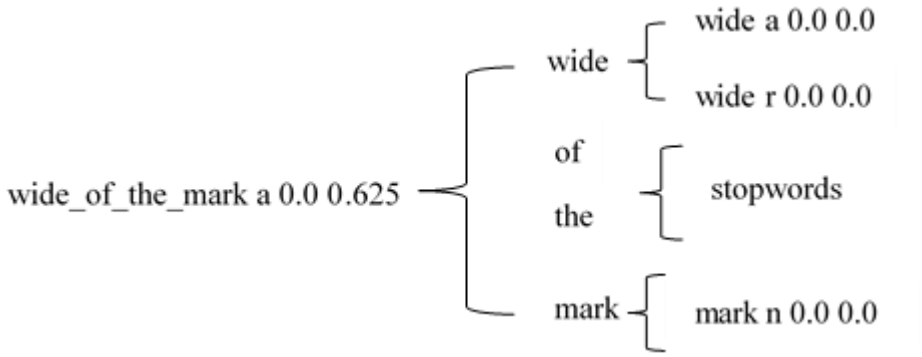


Figura 2.3 Proceso de eliminación de los signos _ y – para la frase wide_of_the_mark.

Con este proceso se crearon muchos términos que ya estaban incluidos dentro del SWN 3.0 por lo que se hizo necesario eliminar términos repetidos teniendo en cuenta las etiquetas que los identifican ya sea como un adjetivo, sustantivo, verbo o adverbio. Por ejemplo, el término *ideal* aparecía cinco veces en el SWN, dos como adjetivo y tres como sustantivo; y después de eliminar las repeticiones solamente quedaron dos apariciones, uno etiquetado como sustantivo y otro como adjetivo. Además de eliminar términos repetidos se excluyeron también aquellos términos que constituían nombres propios, pues estos no repercuten en la polaridad de la opinión.

[16] http://www.rednoise.org/rita/wordnet/documentation/

De esta manera queda conformado el formato final de la nueva versión del SentiWordNet 3.0, con 9030 términos etiquetados como verbos, 53599 etiquetados como sustantivos, 18076 como adjetivos y 3637 etiquetados como adverbios, para un total de 84342 términos.

Otra de las deficiencias encontradas en la herramienta SWN es que la mayoría de los términos que lo conforman tienen como valor positivo y negativo cero, de un total 84342 términos 80795 tienen como polaridad positiva y negativa cero; este fenómeno limita considerablemente a este recurso en su objetivo fundamental, la clasificación de la polaridad de una opinión, pues al no tener suficientes términos puntuados correctamente no puede catalogar las opiniones de manera exacta.

En aras de darle solución al problema anterior se realizó una búsqueda de los términos positivos y negativos del SentiWordNet en las listas de palabras positivas y negativas de Bing Liu, con el propósito de sumar uno a los valores positivos en caso de que el término positivo del SentiWordNet coincidiera con una palabra de la lista de positivas presentadas por Liu & Hu, y lo mismo para el caso de un término negativo del SentiWordNet.

Hasta ahora solamente han sido otorgadas polaridad positiva o negativa a 5037 términos del SentiWordNet, faltando 95.4% de los términos por puntuar. Con el objetivo de aumentar los términos puntuados del SentiWordNet, definimos cuatro etapas, por las que transitó la asignación de polaridad de los términos. En cada una de estas etapas fue de suma importancia la base de datos léxica WordNet, otra herramienta clave en el campo de la minería de opinión. A continuación detallaremos cada una de las etapas definidas, analizando también los resultados obtenidos.

Primera etapa: Para comenzar esta etapa primeramente se separaron los términos sin puntuar y los términos puntuados del SWN. Teniendo como punto de partida los términos del SentiWordNet que no habían sido puntuados se utilizó la biblioteca RitaWordNet para obtener de cada uno de estos términos sus sinónimos mediante la función *getAllSynonyms*, con el objetivo de buscar si alguno de estos sinónimos ya poseía polaridad en el SentiWordNet y asignar esta misma al término señalado. En la **Figura 2.4** se muestra el algoritmo creado para asignar la polaridad de cada término en esta primera etapa.

```
Algoritmo para asignar polaridad (Etapa 1)

Entrada: término no puntuado
polaridad positiva ⟶ 0
polaridad negativa ⟶ 0

PASO 1  Separar del término general del SWN la palabra y la etiqueta que lo identifica.
PASO 2  Obtener de la palabra sus sinónimos:
            sinónimos = getAllSynonyms(palabra)
PASO 3  Si la palabra no tiene sinónimos
            Pasar al PASO 4.
        sino
            Repetir con cada sinónimo
                Si todos los sinónimos fueron analizados
                    Pasar al PASO 4.
                sino
                    Buscar sinónimo en los términos puntuados del SWN
                    Si se encuentra el sinónimo
                        polaridad positiva ⟶ polaridad positiva del sinónimo
                        polaridad negativa ⟶ polaridad negativa del sinónimo
                    Pasar al PASO 4.
                    sino
                        Continuar con el siguiente sinónimo
Paso 4  Retornar "palabra  etiqueta polaridad positiva polaridad negativa".
```

Figura 2.4 Algoritmo de asignación de polaridad para un término (Etapa 1).

Como resultado de la asignación de la polaridad en esta etapa se obtuvieron los valores que se muestran en la **Tabla 2.4.**

Tabla 2.4 Resultados obtenidos en la asignación de polaridad en la Etapa 1.

Total de términos del SentiWordNet con nuevo formato	84342
Términos puntuados en el SentiWordNet original	5037
Términos puntuados en la primera etapa	51027
Total de términos puntuados	**56064**

Como se puede observar en la **Tabla 2.4**, los términos puntuados aumentaron a 56064, por lo que solamente falta el 34.5% de los términos del SentiWordNet por asignarle valores de polaridad de un 95.4% que no tenían valores.

Antes de comenzar con la segunda etapa se separaron los términos que todavía no habían sido puntuados del resto, y los términos que fueron puntuados en la primera etapa se incluyeron en el listado de términos puntuados.

Segunda etapa: En la segunda etapa se aplicó un proceso similar a la primera, pero utilizando la función *getAllAntonyms*() de RitaWordNet, con el objetivo de buscar los antónimos de cada término y si alguno de estos antónimos ya poseía polaridad en el SentiWordNet, asignar la polaridad inversa al término señalado; es decir, como estamos analizando los antónimos, las puntuaciones fueron cambiadas: el valor positivo del antónimo será la polaridad negativa del término no puntuado y el valor negativo del antónimo será la polaridad positiva del término no puntuado. En la **Figura 2.5** se muestra el algoritmo utilizado para asignar la polaridad de cada término en esta etapa.

```
Algoritmo para asignar polaridad (Etapa 2)

Entrada: término no puntuado
polaridad positiva ⟶ 0
polaridad negativa ⟶ 0

PASO 1  Separar del término general del SWN la palabra y la etiqueta que lo identifica.
PASO 2  Obtener de la palabra sus antónimos:
            antónimos = getAllAntonyms(palabra)
PASO 3  Si la palabra no tiene antónimos
            Pasar al PASO 4.
        sino
            Repetir con cada antónimo
                Si todos los antónimos fueron analizados
                    Pasar al PASO 4.
                sino
                    Buscar antónimo en los términos puntuados del SWN
                    Si se encuentra el antónimo
                        polaridad positiva ⟶ polaridad negativa del antónimo
                        polaridad negativa ⟶ polaridad positiva del antónimo
                    Pasar al PASO 4.
                    sino
                        Continuar con el siguiente antónimo
Paso 4  Retornar "palabra  etiqueta polaridad positiva polaridad negativa".
```

Figura 2.5 Algoritmo de asignación de polaridad para un término (Etapa 2).

```
Algoritmo para asignar polaridad (Etapa 3)

Entrada: término puntuado

PASO 1  Separar del término general del SWN la palabra, la polaridad positiva y
        la polaridad negativa.
PASO 2  Obtener de la palabra sus sinónimos:
            sinónimos = getAllSynonyms(palabra)
PASO 3  Si la palabra no tiene sinónimos
            Terminar
        sino
            Repetir con cada sinónimo
                Si todos los sinónimos fueron analizados
                    Terminar
                sino
                    Buscar sinónimo en los términos no puntuados del SWN
                    Si se encuentra el sinónimo
                        Separar del término general del SWN la palabra y la etiqueta que lo identifica.
                        Retornar
                        "palabra del sinónimo  etiqueta del sinónimo polaridad positiva polaridad negativa".
                    sino
                        Continuar con el siguiente sinónimo
```

Figura 2.6 Algoritmo de asignación de polaridad para un término (Etapa 3).

Los resultados de esta etapa se muestran en la **Tabla 2.5Tabla 2.5. Resultados obtenidos en la asignación de polaridad en la Etapa 2.**, donde se puede observar que aumentó la cantidad de términos puntuados en 5678 términos, lo que significa que ya han sido puntuado el 73.2% de los términos del SentiWordNet.

Tabla 2.5. Resultados obtenidos en la asignación de polaridad en la Etapa 2.

Total de términos del SentiWordNet con nuevo formato	84342
Términos puntuados en el SentiWordNet original	5037
Términos puntuados en la primera etapa	51027
Términos puntuados en la segunda etapa	5678
Total de términos puntuados	**61742**

Antes de comenzar con la tercera etapa se separaron los términos que todavía no habían sido puntuados, y los términos que fueron puntuados en la segunda etapa, se incluyeron en el listado de términos puntuados.

Tercera etapa: En esta etapa se utilizaron los términos que hasta el momento ya habían sido puntuados; y a cada uno de ellos se le aplicó la función *getAllSynonyms*() de RitaWordNet con el objetivo de buscar aquellos sinónimos que coincidieran con algún término no puntuado y así colocarle la polaridad del sinónimo coincidente al término no puntuado identificado. En la **Figura 2.6** se muestra el algoritmo que se utilizó en esta etapa.

En la **Tabla 2.6** se muestran los resultados obtenidos al transitar por la etapa 3. Es posible observar que se aumentó la cantidad de términos puntuados en 577.

Tabla 2.6 Resultados obtenidos en la asignación de polaridad en la Etapa 3.

Total de términos del SentiWordNet con nuevo formato	84342
Términos puntuados en el SentiWordNet original	5037
Términos puntuados en la primera etapa	51027
Términos puntuados en la segunda etapa	5678
Términos puntuados en la tercera etapa	5770
Total de términos puntuados	**67512**

Cuarta etapa: En esta etapa también se utilizaron los términos que hasta el momento ya habían sido puntuados. Se aplicó la función *getAllAntonyms*() de RitaWordNet con el propósito de buscar si algunos de los antónimos de los términos coinciden con algún término no puntuado y así colocarle la polaridad al término encontrado. En este caso como se están analizando los antónimos, las polaridades son alternadas como mismo se efectuó en la etapa 2. En la **Figura 2.7** se observa el algoritmo utilizado en esta etapa.

En la **Tabla 2.7** se muestran los resultados de las cuatro etapas. Es posible observar que finalizadas las cuatro etapas logramos aumentar la cantidad de términos con polaridad a 69051, faltando solamente 15991 términos por puntuar.

Tabla 2.7 Resultados obtenidos en la asignación de polaridad en la Etapa 4.

Total de términos del SentiWordNet con nuevo formato	84342
Términos puntuados en el SentiWordNet original	5037
Términos puntuados en la primera etapa	51027
Términos puntuados en la segunda etapa	5678
Términos puntuados en la tercera etapa	5770
Términos puntuados en la cuarta etapa	1539
Total de términos puntuados	**69051**

```
Algoritmo para asignar polaridad (Etapa 4)

Entrada: término puntuado

PASO 1  Separar del término general del SWN la palabra, la polaridad positiva y
        la polaridad negativa.
PASO 2  Obtener de la palabra sus antónimos:
            antónimos = getAllAntonyms(palabra)
PASO 3  Si la palabra no tiene antónimos
            Terminar
        sino
            Repetir con cada antónimo
                Si todos los antónimos fueron analizados
                    Terminar
                sino
                    Buscar antónimo en los términos no puntuados del SWN
                    Si se encuentra el antónimo
                        Separar del término general del SWN la palabra y la etiqueta que lo identifica.
                        Retornar
                        "palabra del antónimo  etiqueta del antónimo polaridad positiva polaridad negativa".
                    sino
                        Continuar con el siguiente antónimo
```

Figura 2.7 Algoritmo de asignación de polaridad para un término (Etapa 4).

Como resultado de todas las etapas, incrementamos el número de términos con polaridad a 69051 representando el 81.9% de los términos que incluimos en la nueva versión del SentiWordNet 3.0 a la cual denominamos SentiWordNet 4.0.

2.2.2 SpanishSentiwordNet

Para la creación del SpanishSentiWordNet fue de suma importancia la utilización del Índice Intralingüístico de WordNet para el español y el SentiWordNet 4.0. Los índices intralingüísticos (ILI) constituyen una base no estructurada de conceptos, que tiene como propósito proporcionar una cartografía eficaz entre diferentes lenguajes. Cada concepto se representa como un registro del ILI que representa una referencia a la fuente del synset asociado.

El formato que contiene este índice es primeramente el término en Español seguido por un carácter de espacio, luego la etiqueta *pos* que puede tomar los valores *n*, *a*, *v*, *r*; donde *n* representa a los sustantivos, *a* a los adjetivos, *v* a los verbos y *r* a los adverbios, a continuación se especifica un carácter de tabulación y seguido varios identificadores de las diferentes relaciones semánticas del término, después se especifica otro carácter de tabulación y finalmente las acepciones de la palabra en Inglés separadas por un carácter de espacio. A continuación se muestra un ejemplo con la palabra *agresor* para su mejor compresión:

agresor n 09195176 09158637 09848308 attacker assailant aggressor assaulter aggressor robber

Los índices intralingüísticos surgieron como una lista de synsets en la versión 1.5 de WordNet, pero se han ido adaptando para relacionar versiones de WordNet para distintos idiomas. Este recurso se utilizó en la creación del EuroWordNet[17], pues constituye la herramienta idónea para conectar versiones de WordNet.

Teniendo estas herramientas creamos un recurso léxico enfocado al idioma Español siguiendo el método propuesto en (Amores 2013) que plantea:

"Evaluar la polaridad de un término sumando las polaridades positivas y negativas de sus acepciones".

Para solucionar el problema de la asignación de los valores de polaridad necesitamos las acepciones en Inglés del término que nos proporciona el Índice Intralingüístico y posteriormente adquirir las puntuaciones de dichas acepciones proporcionadas por el SentiWordNet 4.0. Si ya contamos con las polaridades positiva y negativa de cada acepción en Inglés del término original en Español, es posible conformar el SpanishSentiWordNet aplicando el método propuesto en (Amores 2013), que suma las polaridades positivas y negativas de todas las acepciones, proporcionando así la puntuación final del término original en Español. Otros operadores de agregación pudieran ser utilizados para combinar los valores de polaridad de las acepciones de cada término, se mantiene el uso de la suma de las polaridades por los resultados reportados en (Amores 2013).

En la **Figura 2.8** se muestra un esquema que ilustra la conformación del SpanishSentiWordNet, según se describió anteriormente.

[17] http://www.illc.uva.nl/EuroWordNet/

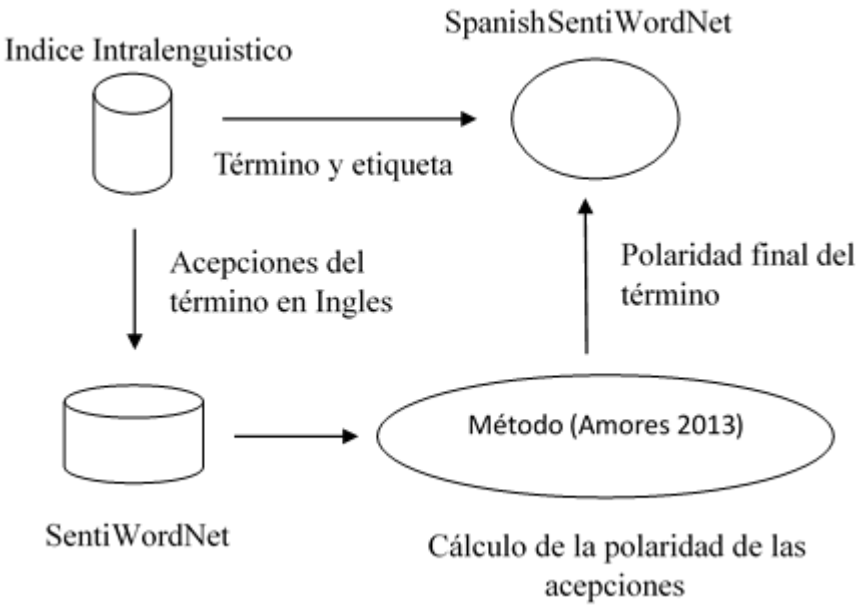


Figura 2.8 Proceso de creación del SpanishSentiWordNet.

La nueva herramienta SpanishSentiWordNet está conformada por 54289 términos en total y su estructura es la siguiente: primeramente el término en Español seguido por un carácter de tabulación, a continuación la etiqueta que lo identifica como adjetivo, adverbio, verbo o sustantivo, seguido de un espacio en blanco y por último las polaridades positivas y negativas en ese orden separadas por un carácter de espacio. En la **Figura 2.9** se muestra un fragmento de la estructura en que quedó conformado el SpanishSentiWordNet.

```
agresor    n 0.0 3.125
burocrata    n 0.875 0.0
hedor    n 0.0 5.375
inconsciente    a 1.0 7.375
liderar    v 2.375 0.0
coaccion    n 0.5 1.0
cuerpo_de_niño    n 0.0 0.0
palurdo    a 0.5 0.0
individualidad    n 1.75 0.875
polyporus    n 0.125 0.0
llave_maestra    n 1.25 0.0
interrumpido    a 0.25 1.125
manecilla    n 0.25 0.25
alabar    v 1.0 0.0
combatir    v 0.0 0.5
```

Figura 2.9 Estructura del SpanishSentiWordNet.

Como se observa en la **Figura 2.9** algunos de los términos contienen el signo _ para indicar frases. Al igual que sucede con el SentiWordNet 4.0, en el SpanishSentiWordNet, también es conveniente dejar solamente palabras simples, ya que el procesamiento de frases y palabras compuestas complejiza el procesamiento y no aporta en el cálculo de la polaridad de las opiniones. Por tanto, estas herramientas están orientadas a la evaluación de la polaridad de un único término, de ahí que se hizo necesario eliminar los signos _.

Para la eliminación del signo _ en los términos, primeramente se separaron las palabras que estaban unidas por _. Separadas las palabras se eliminaron las que constituían stopwords, es decir, se eliminaron las palabras que no ofrecen ningún significado en el idioma Español, y se crearon así nuevos términos con las palabras restantes, con la misma etiqueta del término que se separó anteriormente y con polaridad positiva y negativa igual a 0. A continuación se eliminaron los términos repetidos teniendo en cuenta su etiqueta, es decir se eliminaron adjetivos, sustantivos y verbos repetidos. Además, se eliminaron los términos que constituyen nombres propios pues estos no aportan a la polaridad de la opinión. Finalmente, se normalizaron sus valores de polariadad. Así, quedó conformado el SpanishSentiWordNet con un total de 43525 términos, de ellos 30289 etiquetados como sustantivos, 8664 adjetivos y 4572 como verbos.

2.2.3 Sentiwordnet 4.1

Durante la creación del SentiWordNet 4.0 asumimos que cuando se encontrara un sinónimo de un término sin polaridad, o viceversa, un sinónimo sin polaridad de término con valores de polaridad asignados, se le transferirían íntegramente los valores ya asignados a los que no los poseyeran. Esto no es del todo cierto, pues la relación de sinónimos en WordNet es un poco más amplia y dos supuestos sinónimos no tienen por qué tener los mismos valores de polaridad asignados. Tomemos como ejemplo la palabra “computer”, como se puede apreciar a continuación aparece en dos synsets, por tanto todos los términos de estos dos synsets son sinónimos.

synset#1 {***computer, computing machine, computing device, data processor, electronic computer, information processing system***} - (*a machine for performing calculations automatically*)

synset#2 {calculator, reckoner, figurer, estimator, computer} - (*an expert at calculation (or at operating calculating machines)*)

Como se pudo apreciar, los términos difieren en cuanto a sentidos diferentes, por lo tanto, no deberían tener la misma polaridad. Para solucionar esto recurrimos a una medida de distancia entre dos términos que es proporcionada por la biblioteca RitaWordNet, que se basa en las distintas relaciones que forman los términos en la red semántica de WordNet. Esta medida que está en el rango de [0, 1] toma su máximo valor cuando los términos comparados están dentro del mismo synset, dado este caso, transferimos la polaridad íntegra del término con las puntuaciones de polaridad al término comparado previamente sin polaridad asignada. En los demás casos, le asignamos al término sin valores de polaridad asignados previamente, la puntuación obtenida de la distancia, así, mientras menos relacionada esté la palabra que se le busca asignar un valor de polaridad con la palabra que se compara, menor será su valor, acercándose más así a la realidad. Otras relaciones que miden esta distancia son descritas a continuación, lo que garantizó incorporar nuevas puntuaciones a otros términos.

(hiperónimo) => {***machine***} - (*any mechanical or electrical device that transmits or modifies energy to perform or assist in the performance of human tasks*)

(hipónimo) => {***analog computer, analogue computer***} - (*a computer that represents information by variable quantities* (*e.g., positions or voltages*))

(merónimo) => {***busbar, bus***} - (*an electrical conductor that makes a common connection between several circuits; "the busbar in this computer can transmit data either way between any two components of the system"*)

A realizar esta nueva puntación, se verificaron los términos puntuados con otros recursos léxicos anotados manualmente, algunos descritos anteriormente, como Micro-WNop, General Inquirer, SentiSense, The Subjectivity Lexicon. En el caso de no coincidir los valores, eran modificados para que los términos tuviesen la misma clase que en los recursos léxicos anotados manualmente.

Así, quedó finalmente conformado el SentiWordNet 4.1, con un total de 58969 términos anotados, de ellos 35520 etiquetados como sustantivos, 13190 adjetivos, 8298 verbos y 1961 adverbios.

2.3 Esquemas para la detección de la polaridad basados en recursos léxicos

En (Amores, Arco & Artiles 2015a) se presenta un esquema general para detectar la polaridad de las opiniones de manera no supervisada siguiendo el esquema que se muestra en la **Figura 2.10¡Error! No se encuentra el origen de la referencia.**. Esta propuesta determina la polaridad de las oraciones teniendo en cuenta la polaridad de las palabras considerando su sentido correcto en un contexto determinado.

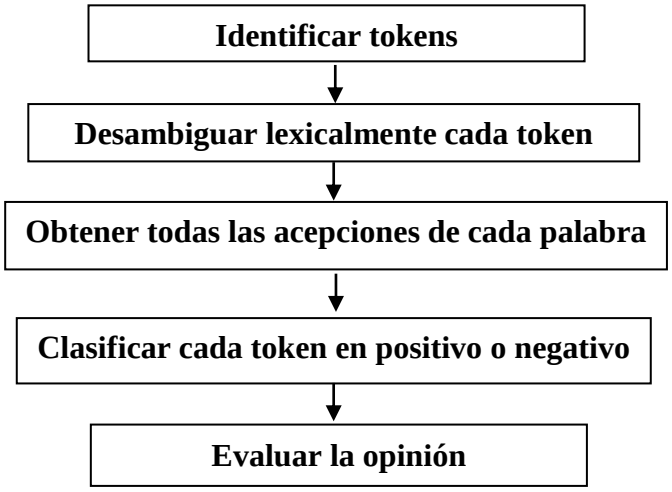


Figura 2.10 Esquema general para detectar la polaridad de las opiniones.

La Etapa 1 es la encargada de leer las opiniones que fueron especificadas en el XML de entrada y seleccionar los términos que aporten información útil, eliminando las palabras vacías. A continuación, en la Etapa 2, se parte de cada término que aporta información útil y se lematiza y se desambigua lexicalmente. Posteriormente, en la Etapa 3 se traducen los términos seleccionados en la Etapa 2, obteniendo todas las acepciones del término en Inglés, ya que, se sugiere utilizar el SentiWordNet para clasificar la polaridad de las palabras, y éste requiere que los términos estén en idioma Inglés. En la Etapa 4, se aplicó un método para el cálculo de la polaridad de los términos, considerando la polaridad de cada una de las acepciones del término. Se suman todos los valores de polaridad positivos de las acepciones, así como todos los valores de polaridad negativos, y se toma el mayor valor, asignándolo como la polaridad del término, y por tanto, contribuyendo a la polaridad de la opinión. Así, en la Etapa 5, al terminar de analizar todos los términos y sus acepciones, la opinión cuenta con

un valor positivo y otro negativo, los cuales son comparados, y se toma como polaridad de la opinión el mayor valor.

Este esquema y la herramienta que lo implementa PosNeg Opinion (Amores 2013; Amores, Arco & Artiles 2015b), tienen como desventaja que están basados en SentiWordNet 3.0, recurso que presenta algunos errores que mostramos anteriormente con los cuales no se pueden alcanzar grandes resultados en colecciones variadas y de una gran cantidad de opiniones. No obstante, este esquema mostró que para comentarios muy específicos sobre productos se logran buenos resultados debido al tratamiento de la negación y el uso de palabras modificadoras que logran subsanar algunos errores del recurso anteriormente mencionado. A pesar de que PosNeg Opinion trata con estos problemas característicos de las opiniones, lo hace de una manera muy rudimentaria y poco efectiva, lo que evidencia que aún es necesario continuar desarrollando propuestas que manejan de una manera más efectiva las características de las opiniones y que incorporen las ventajas de nuevos recursos léxicos.

2.3.1 Esquema general para la detección de la polaridad basado en Sentiwordnet 4.0 y en Spanish Sentiwordnet

Con el objetivo de mejorar la eficiencia y la efectividad de PosNeg Opinion es necesario transformar el esquema que esta aplicación sigue para el cálculo de la polaridad de las opiniones. Por eso, partimos del esquema propuesto en (Amores, Arco & Borroto 2015) y creamos un nuevo esquema para el cálculo de la polaridad de las opiniones. Para ello, remplazamos las etapas 3 y 4 del esquema original (Amores, Arco & Artiles 2015a) por el uso de los nuevos recursos desarrollados: SentiWordNet 4.0 y SpanishSentiWordNet, logrando de esta forma resolver identificar la polaridad de los términos en Español sin la necesidad de buscar cada una de las acepciones de los términos, además, trabajamos con valores más reales de la polaridad de los términos tanto para el Inglés como para el Español. Para facilitar la aplicación de estos recursos fue creada la biblioteca PolarityDetection (Borroto 2014).

El nuevo esquema propuesto se presenta en la **Figura 2.11**. Las etapas A, B y D coinciden con las etapas 1, 2 y 5 del esquema original, respectivamente. En la etapa C usamos la biblioteca PolarityDetection para procesar opiniones en Inglés y en Español usando SentiWordNet 4.0 y SpanishSentiWordNet.

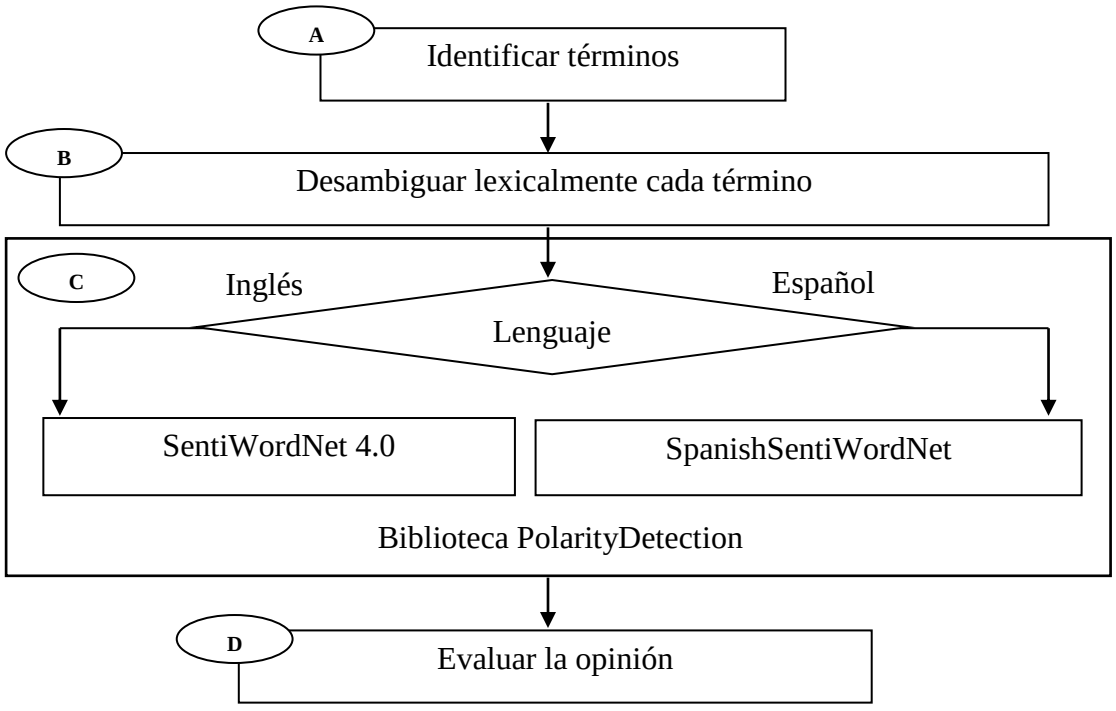


Figura 2.11 Esquema general para la detección de la polaridad usando la biblioteca PolarityDetection.

La aplicación PosNeg Opinion 2.0 sigue el esquema propuesto y en su implementación se utilizó la biblioteca PolarityDetection. Esta biblioteca nos permite sumar los valores de polaridad positivos y negativos de un término obtenidos de SentiWordNet 4.0 y SpanishSentiWordNet, respectivamente. También usa listas de palabras negadoras y modificadoras aunque continúa haciéndolo de una forma muy básica y pobre; sin embargo, con ellas es posible incrementar el valor de la polaridad de un término si éste está precedido por una palabra modificadora o intercambiar los valores de polaridad positiva y negativa si el término esta precedido de una palabra negadora. Desafortunadamente, con esta solución aún no se hace frente a otros problemas claves también presentes en las opiniones como la aparición de jergas o emoticonos. Este esquema es fácilmente extensible a otros idiomas, ya que solo necesitaríamos el índice intralingüístico del lenguaje específico al Inglés para obtener su respectivo lexicón.

2.3.2 Esquema general para la detección de la polaridad basado en Sentiwordnet 4.1, en Spanish Sentiwordnet y en recursos para el procesamiento del lenguaje natural

Con el objetivo de solventar los problemas que presenta el esquema anteriormente propuesto se desarrollaron algunos recursos y técnicas para tratar varios de los problemas presentes en las opiniones. También se diseñó un nuevo esquema el cual incluye estos nuevos recursos y técnicas. El recurso SentiWordNet 4.0 fue sustituido por el nuevo recurso desarrollado SentiWordNet 4.1 y también un nuevo SpanishSentiWordNet fue desarrollado a partir de este nuevo recurso siguiendo la metodología explicada en el epígrafe 2.2.2.

2.3.2.1 Nuevos recursos para manejar problemas presentes en las opiniones

Manejo de la negación:

La negación es un elemento clave en el análisis de opiniones, dado que son muy abundantes las opiniones negativas expresadas con términos positivos negados y viceversa (Jiménez Zafra et al. 2015). La oración "No me gusta la carcasa del teléfono" es un claro ejemplo de una opinión negativa con un término positivo (gusta) negado.

En el ámbito del análisis de sentimientos, la mayor parte de las investigaciones se han realizado para opiniones escritas en inglés. Sin embargo, la presencia cada vez mayor en Internet de otros idiomas pone de manifiesto la necesidad de desarrollar sistemas que traten lenguas diferentes al inglés. Así, han aparecido algunos trabajos de investigación que tratan con textos escritos en chino (Zhang et al. 2009), en árabe (Rushdi-Saleh et al. 2011) o en español (Martín-Valdivia et al. 2013).

Por otra parte, el tratamiento de la negación es un problema abierto dentro del PLN en general, y dentro de la minería de opinión en particular. Se trata de un fenómeno lingüístico que no ha sido suficientemente estudiado y que consideramos que requiere un análisis profundo. A pesar de las similitudes entre idiomas, la negación es un elemento lingüístico muy particular de cada lengua, por lo que para un tratamiento efectivo se debe realizar un estudio pormenorizado de los distintos elementos lingüísticos que intervienen en el proceso de negación de un enunciado. La investigación existente sobre la influencia de la negación es escasa, no estudia en profundidad estos aspectos lingüísticos y se centra casi en exclusividad en textos escritos en inglés, aunque existen algunas propuestas para otros idiomas como se planteó anteriormente.

Las primeras aproximaciones comenzaron en el año 2001 y sugieren métodos relativamente sencillos. Das y Chen (Das & Chen 2001) proponen añadir "NOT" a las palabras de la oración

que se encuentren cerca de términos negativos, como por ejemplo “no” o “don’t”. En (Pang et al. 2002) utilizan la técnica de Das y Chen asumiendo que las palabras afectadas por una palabra negativa (“not”, “isn’t”, “didn’t”, etc.) son todas aquellas que se encuentran entre dicha palabra y el primer signo de puntuación. Estos autores realizan experimentos empleando algoritmos de aprendizaje automático para la clasificación de las opiniones teniendo en cuenta el tratamiento de la negación (característica “NOT”) y sin tenerlo en cuenta, llegando a la conclusión de que con este método se produce una mejora insignificante. En (Polanyi & Zaenen 2006) dan un paso más allá, considerando además de la negación, intensificadores y atenuantes, introduciendo de esta manera el concepto de modificadores de valencia contextuales. Además, se trata del primer modelo computacional que asigna puntuaciones a expresiones polares, invirtiendo la polaridad de las expresiones negadas. Desafortunadamente este modelo no llegó a implementarse, por lo que solo se puede especular sobre su efectividad. Kennedy e Inkpen (Kennedy & Inkpen 2006) utilizan un modelo de negación muy similar al de (Polanyi & Zaenen 2006), definiendo como ámbito de una palabra negativa, intensificadora o atenuante aquella palabra inmediatamente posterior. En el caso de las palabras afectadas por la negación, el enfoque que siguen es el de invertir la polaridad y en el caso de las palabras que se encuentran en el ámbito de los intensificadores/atenuantes, lo que hacen es incrementar/disminuir el grado de positividad/negatividad según sea el caso. Para clasificar las opiniones utilizan dos métodos, en el primero de ellos clasifican un comentario en base al número de términos positivos y negativos que contiene y en el segundo emplean el algoritmo de aprendizaje automático SVM, llegando a la conclusión de que el modelado de la negación es un hecho importante. Por otro lado, en (Wilson et al. 2005) proponen utilizar una ventana fija de tamaño cuatro para determinar el ámbito de la negación; es decir, una palabra se ve afectada por la negación si existe una expresión de negación entre las cuatro palabras anteriores a ella. En este trabajo también emplean un método de aprendizaje no supervisado para clasificar las opiniones en el que utilizan características para modelar las palabras afectadas por la negación, por intensificadores/atenuantes y por otras expresiones polares.

Los enfoques presentados en el párrafo anterior son considerados los trabajos pioneros en el modelado de la negación en el análisis de sentimiento en inglés. Sin embargo, estas soluciones no son suficientemente precisas. Por ello, se ha seguido trabajando en este tema para tratar de ofrecer mejores soluciones. Entre los últimos trabajos presentados cabe destacar los de (Jia et

al. 2009; Taboada et al. 2011; Díaz 2014). En (Jia et al. 2009) proponen un sistema basado en reglas que utiliza información derivada de los árboles de dependencias de las oraciones de estudio, mejorando los enfoques existentes hasta ese momento. En el completo y detallado trabajo de (Taboada et al. 2011) se presenta una versión mejorada de su sistema SO-CAL (Taboada et al. 2008), para calcular la orientación semántica de una opinión, en el que marcan como negadas todas aquellas palabras que se encuentren después de una partícula negativa hasta llegar a un signo de puntuación, un conector o ciertas palabras pertenecientes a una determinada categoría gramatical (ej. "it" (pronombre)). Los autores introducen una nueva forma de tratar la negación que consiste en reducir el valor de polaridad de las palabras negadas en lugar de invertirlo. En (Díaz 2014) presenta un sistema para el tratamiento de la negación y de la especulación en textos médicos y opiniones, y el primer corpus de opiniones etiquetado a nivel de negación y especulación (Konstantinova et al. 2012). Por último, mencionar el trabajo de (Wiegand et al. 2010) en el que se realiza una excelente revisión del estado del arte de la negación en inglés hasta ese momento.

Como se puede ver, la investigación sobre este tema en inglés es bastante amplia si la comparamos con la existente en español. Con respecto al tratamiento de la negación en documentos en español, el primer trabajo que conocemos es el de (Brooke et al. 2009) en el que adaptan su primera versión del sistema SO-CAL para análisis de opiniones en inglés (Taboada et al. 2008), a un nuevo idioma, el español. Con respecto al tratamiento de la negación siguen el mismo enfoque que el empleado en su versión en inglés, teniendo en cuenta que en español los adjetivos pueden aparecer antes y después de los sustantivos. Finalmente, en (Vilares et al. 2013) tienen en cuenta la estructura sintáctica del texto para el tratamiento de la negación, de la intensificación y de las oraciones subordinadas. Los resultados de los autores muestran una mejora en el rendimiento con respecto a los sistemas puramente léxicos.

Nuestra propuesta para manejar la negación se basa en la unión de varios de los enfoques antes descritos. Primeramente tomamos las reglas presentadas en (Na et al. 2005) para la negación en el idioma inglés y las modificamos para poder aplicarlas en nuestra aplicación, ya que en estas reglas están contemplados algunos términos los cuales descartamos en nuestro pre-procesamiento porque son palabras que no aportan en el cálculo de la polaridad. El rango de la negación está definido por las reglas. Por ejemplo, dada la oración "I don't like this laptop" la

rama de la regla que se muestra en la **Figura 2.12** se aplica como se explicará a continuación. Cuando se detecta el verbo "do" seguido del término "not" se activa el árbol de reglas para este término. El siguiente término que aparece en la oración es un verbo, por tanto, se recorre el árbol buscando esta entrada. Si aparece la entrada, como es el caso de este ejemplo que se muestra en la **Figura 2.12**, la polaridad de este término también se niega. A continuación, el próximo término a analizar es un sustantivo, por tanto, se busca en la rama del árbol esta entrada, como no existe, se termina la negación.

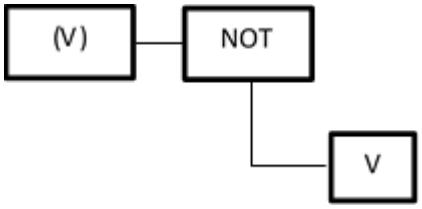


Figura 2.12 Rama de la regla de negación para el término "not".

Estas reglas están creadas para los términos "not", "never" y "no" del idioma Inglés. Para el idioma Español adaptamos las reglas descritas en (Vilares et al. 2013) para los términos "no", "nunca" y "sin". Las reglas solamente involucran los tres términos más utilizados para negar (nor, never y no); sin embargo, existen otros términos que expresan negación, de ahí la necesidad de utilizar otros enfoques para manejar la negación. Por ello, se crearon listas de palabras negadoras en ambos idiomas a partir de diversas listas ya publicadas. El rango de negación que consideramos para estas palabras se obtuvo a través de la experimentación con rangos de 1, 2 y 3 términos después de la palabra negadora, siendo el de mejores resultados el rango que considera negar los próximos dos términos. También se tuvo en cuenta el uso de sufijos y prefijos para ambos idiomas, un ejemplo para el inglés lo constituye el prefijo dis- (dislike) y para el español el prefijo in- (incomible). En todos los casos la negación se trata invirtiendo los valores de polaridad de cada término negado.

La aplicación de nuestra propuesta para manejar la negación considera el uso de uno u otro enfoque dependiendo del término que se presente en el procesamiento textual. Por ejemplo, si aparece la palabra "no" directamente se activa el enfoque basado en reglas, si el término a analizar presenta un prefijo que expresa negación entonces seguimos el enfoque basado en prefijos y si aparece otro término que expresa negación y que no es manejado desde las reglas, entonces se consulta la lista de palabras negadoras.

Manejo de palabras modificadoras:

En gramática, un intensificador es una palabra que hace hincapié en otra palabra o frase. También conocido como un refuerzo o un amplificador. Los adjetivos intensificadores modifican a los sustantivos; adverbios de intensificación modifican comúnmente verbos, adjetivos graduables, y otros adverbios. Algunos ejemplos en el idioma inglés los mostramos a continuación:

- "Oh, I am ***so*** not in the mood for this. I've just been shot!"
- "The woodwind has a ***slightly greater*** scope than the violin."
- "I'm Jewish, but I'm ***totally*** not."
- "Dude, I am ***so*** wicked psyched to slay my first dragon!"
- "A ***really*** good detective never gets married."
- "Maggy knew that something ***surpassingly*** important had happened to him."

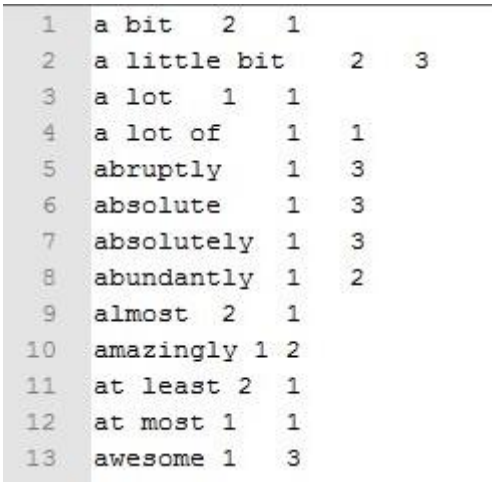
Hasta cierto punto, un intensificador actúa como una señal: anuncia que la siguiente palabra que se lleva a cabo debe ser entendida como insuficiente. Por ejemplo, en la frase “una noche completamente hermosa”, el autor está diciendo, a que fue algo más que bonito, que no tiene la palabra precisa para describirlo.

Los seres humanos somos de hecho exageradores natos, y este rasgo es uno de los principales impulsores de cambio de idioma. En ninguna parte es esto más evidente que en la renovación constante de las palabras de intensificación. Estas son las pequeñas palabras que fortalecen los adjetivos. Expresan un punto alto a lo largo de una escala. Por ejemplo, algo no sólo es bueno, sino muy bueno, o incluso, terriblemente bueno.

Para tratar estas palabras modificadoras, como resultado de esta investigación, se construyeron dos listas (una para cada idioma) de términos intensificadores a partir de la consultas manuales a diccionarios, análisis de listas de palabras modificadoras ya existentes y consultas a expertos lingüistas. Como estas palabras pueden influir tanto a los términos precedentes como a los posteriores se les agregó a cada uno una etiqueta para identificar donde ejercía éste su influencia. En la Figura 2.13 se muestra un fragmento de la lista construida para el idioma Inglés, donde es posible apreciar que el formato consiste en la frase o el término modificador seguido de un valor entero en el dominio {1,2}, donde uno significada que es una palabra intensificadora y dos que es una palabra minimizadora. A continuación aparece un segundo

valor entero en el dominio de {1,2,3} que señala a que término afecta siendo uno el término antecesor, dos el sucesor y tres para ambos casos. La forma de tratar con ellas fue agregarle uno al valor de polaridad predomínate del término predecesor o sucesor según sea la etiqueta que porte la palabra intensificadora.

Otras palabras modificadoras tratadas son las palabras minimizadoras, como por ejemplo, el término “less” en inglés o “poco” en español. Par el tratamiento de estas palabras se utilizó la misma propuesta que para las palabras intensificadoras solo que éstas en vez de aumentar el valor de polaridad lo reducen en un 50%.



```
1  a bit   2   1
2  a little bit    2   3
3  a lot   1   1
4  a lot of    1   1
5  abruptly    1   3
6  absolute    1   3
7  absolutely  1   3
8  abundantly  1   2
9  almost  2   1
10 amazingly 1 2
11 at least 2  1
12 at most 1   1
13 awesome 1   3
```

Figura 2.13 Fragmento de la lista de palabras modificadoras.

Manejo de jergas, abreviaturas y emoticonos:

Las jergas de internet son una variedad de lenguajes cotidianos y abreviaciones en la escritura utilizados por las diferentes comunidades en el ámbito de Internet (Yin 2006). Es difícil dar una definición estandarizada de la "jerga de Internet" debido a los constantes cambios realizados en su naturaleza y dada la rapidez de los adelantos tecnológicos. Sin embargo, y en forma general, se entiende como un tipo de jerga que han popularizado los usuarios de Internet y, en muchos casos, que se ha validado con el uso. Esos nuevos términos o formas de escritura a menudo se han originado con el propósito de abreviar pulsaciones de teclado, sobre todo por la dificultad y/o rapidez que a veces trae aparejado el uso de algunos dispositivos portátiles. Muchas personas utilizan las mismas abreviaturas en los mensajes de texto y en mensajería instantánea, así como en las redes sociales. De igual forma, los acrónimos, símbolos del teclado, y abreviaturas, también son de uso frecuente y más o menos normalizado, y por tanto,

integran lo que puede llamarse jerga de Internet. Otro elemento que encontramos dentro de las jergas de internet son los emoticonos (emoticon) que son una secuencia de caracteres ASCII que, en un principio, representaba una cara humana y expresaba una emoción. Posteriormente, fueron creándose otros emoticonos con significados muy diversos. Los emoticonos que expresan alegría u otras emociones positivas se clasifican normalmente como smileys (de "smile", "sonrisa" en inglés). Los emoticonos se emplean frecuentemente en mensajes de correo electrónico, en foros, SMS y en los chats mediante servicios de mensajería instantánea. Los emoticonos se han ido desarrollando a lo largo de los años, principalmente, para imitar las expresiones faciales y las emociones, para vencer las limitaciones de tener que comunicarse sólo en forma de texto y porque sirven como abreviaturas. Se han escrito libros sobre este tema, con listas interminables de emoticonos. En los foros de Internet, los emoticonos se suelen reemplazar automáticamente por las imágenes correspondientes. En algunos editores de textos (como por ejemplo Microsoft Word), la opción de "corrección automática" reconoce emoticonos básicos como :) y :(, cambiándolos por el carácter correspondiente.

Existen varias listas que incluyen jergas y emoticonos; sin embargo, cada una adolece de términos que presentan otras, sobre todo debido a la velocidad y dinámica con que surgen nuevos emoticonos y jergas. De ahí que para tratar con estos dos temas creamos dos listas, una para las jergas, y otra para los emoticonos, construidas a partir de varias listas previamente publicadas[18]. Cada emoticón o jerga tiene asignados sus correspondientes valores de polaridad positivo y negativo. Para obtener estos valores de polaridad de las jergas utilizamos su correspondiente significado en el idioma y sumamos los valores de polaridad que ofrece el SentiWordNet 4.1 de sus términos. Para los emoticonos es similar, solo que se le asigna a cada emoticón el valor de polaridad de la emoción que expresan. En ambos casos los valores de polaridad son agregados al valor final de toda la opinión.

A partir de los nuevos recursos y técnicas desarrollados se hizo necesario cambiar el esquema presentado en el epígrafe 2.3.2, dando como resultado un nuevo esquema conformado por seis etapas, el cual describimos a continuación.

[18]www.netlingo.com, www.internetslang.com, www.noslang.com/dictionary/, www.canal3000.com/moviles/emoticonos01.htm, https://es.wikipedia.org/wiki/Anexo:Emoticonos

2.3.2.2 Descripción de las fases del segundo esquema propuesto

Como podemos ver en la **Figura 2.14** las etapas 2, 4 y 6 coinciden con las etapas B, C y D del esquema presentado en la **Figura 2.11**, respectivamente. En la etapa 1 se agregó el proceso de identificación y tratamiento de las jergas y emoticonos, en la etapa 3 se realiza un pre-procesamiento del texto donde se ofrece, de manera opcional, la corrección de los errores ortográficos, si no se desea corregir los errores ortográficos, se pueden agregar los términos que comúnmente se escriben de manera incorrecta al recurso SentiWordNet 4.1 y asignarle los valores de polaridad de la palabra correctas correspondiente. Y en la etapa 5 después de asignados los valores de polaridad a todos los términos, se realiza el tratamiento de las negaciones.

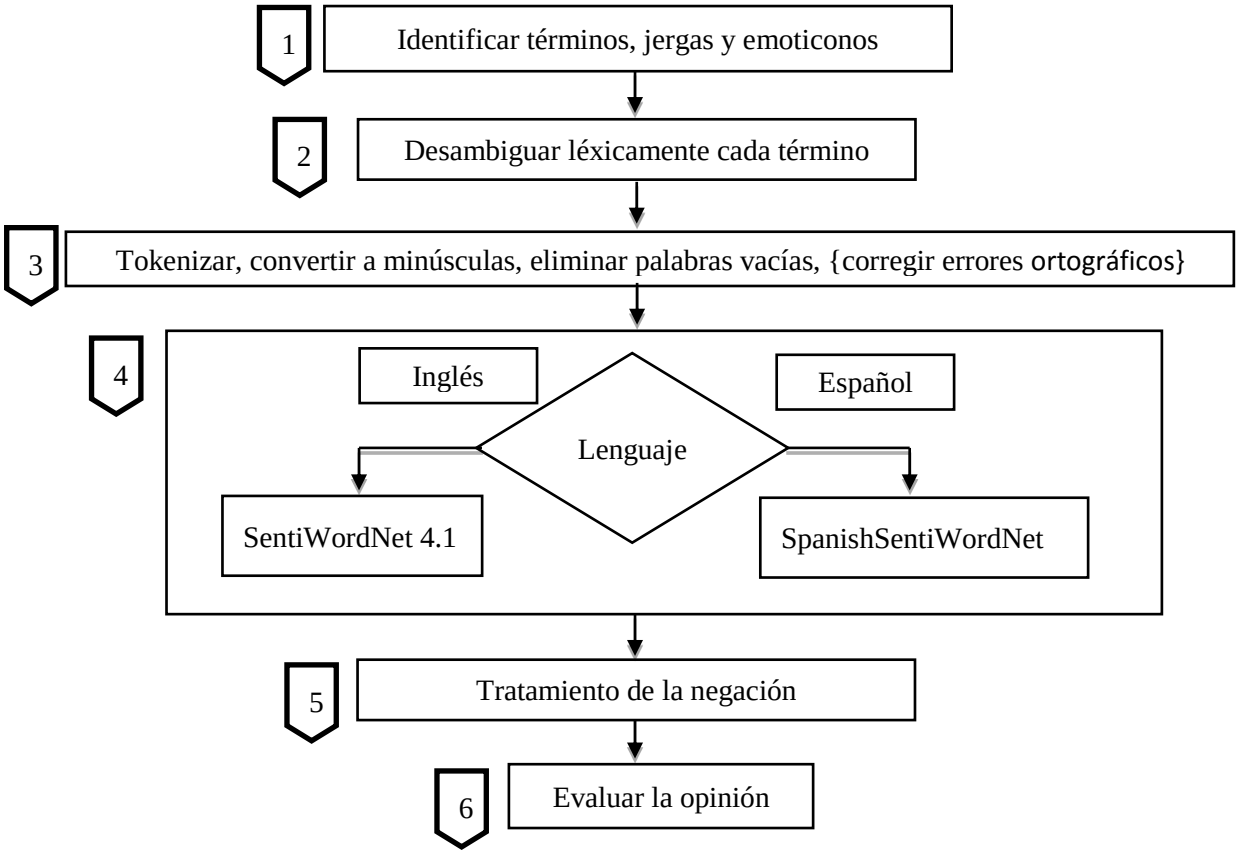


Figura 2.14 Esquema final propuesto.

En este esquema general primeramente se debe leer el texto de opiniones a analizar. Después de cargar el corpus textual es necesario localizar los emoticones y las jergas poara

preprocesarlos y determinar su influencia en el cálculo de la polaridad. Este es un paso que distingue la minería de opinión de otros enfoques de minería de textos donde en la etapa de preprocesamiento son eliminadas inicialmente los emoticonos y las jergas. Después de calcular los valores de polaridad asociados a las jergas y los emoticonos, estos deben ser eliminados, ya que no aparecerán directamente en los recursos léxicos a consultar. A continuación se realiza la etapa 2, donde se etiquetan los términos. Es importante realizar esta etapa antes de aplicar variantes del preprocesamiento textual, de forma tal que el etiquetador a utilizar puede hacerlo considerando los términos tal y como se presentaron originalmente en las oraciones a analizar. Concluida la etapa 2, estamos listos para aplicar los algoritmos de preprocesamiento previstos en la etapa 3 y por tanto realizar un proceso de tokenización, convertir a minúsculas, eliminar palabras vacías y corregir errores ortográficos si se desea seguir esta variante. Al terminar estas tres primeras etapas es que podemos asignar la polaridad a los términos resultantes. Como al concluir la etapa 4, los términos tienen los valores de polaridad asignados, es posible pasar a la etapa 5 para aplicar la estrategia híbrida definida para manejar la negación. Finalmente, en la etapa 6 se suma la polaridad de todos los términos y se otorgar un valor final de polaridad al texto analizado.

2.4 PosNeg Opinion 3.0

En este epígrafe se presenta la aplicación PosNeg Opinion 3.0 que se desarrolló como resultado de este trabajo siguiendo el esquema general propuesto en el epígrafe 2.3.2.2 para la detección no supervisada de la polaridad de las opiniones. A continuación detallaremos elementos del diseño, la implementación y la interfaz de esta aplicación.

2.3.1 Diseño e implementación de PosNeg Opinión

Siguiendo el esquema general propuesto, se implementó el software PosNeg Opinión 3.0. Como se muestra en la **Figura 2.15** la aplicación cuenta con nueve clases agrupadas en un solo paquete.

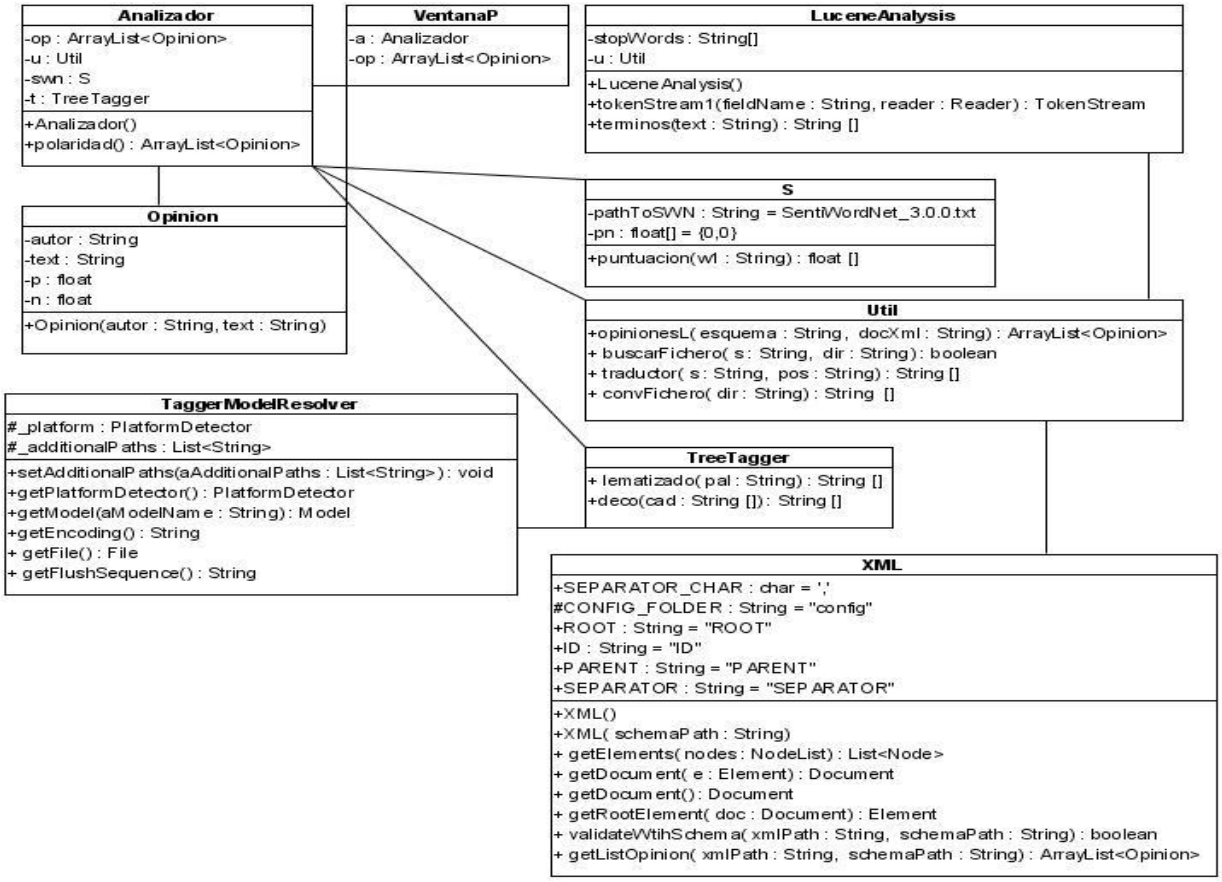


Figura 2.15 Diagrama de clases del sistema PosNeg Opinión.

La clase Analizadores es la manejadora pues usa a las restantes para analizar las opiniones dadas y brindar los resultados. Esta clase se instancia creando un objeto de la clase XML, como se observa en la **Figura 2.16.** El objeto XML se instancia con el esquema XML definido para la entrada de las opiniones y un archivo XML de donde son extraídas aquellas a analizar.

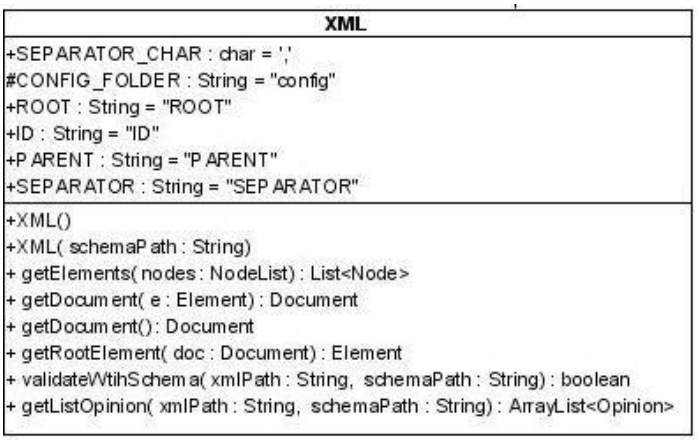


Figura 2.16 Clase XML.

La clase Analizador, como se observa en la **Figura 2.17**, cuenta con varios atributos como son una lista de opiniones, un objeto de la clase Útil, otro objeto de la clase S y uno de la clase TreeTagger, cuenta además con el único método polaridad() que retorna la lista de opiniones ya analizadas. Dentro de este método primeramente se crea un objeto de la clase LuceneAnalysis, como se observa en la **Figura 2.18**Figura 2.18, con el cual se eliminan las palabras vacías del contenido de la opinión.

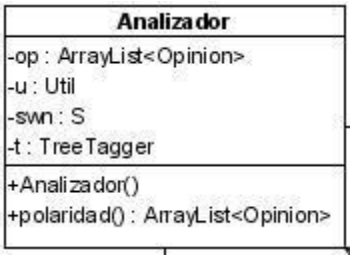


Figura 2.17 Clase Analizador.

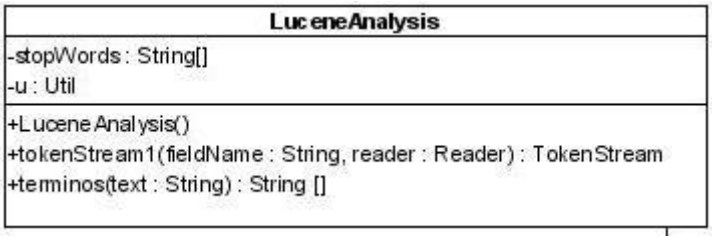


Figura 2.18 Clase LuceneAnalysis.

Seguidamente se utiliza el objeto de la clase TreeTagger, que se muestra en la **Figura 2.19**, para lematizar los términos mediante el método lematizado(String s) que devuelve un arreglo con el término en su raíz y una etiqueta de desambiguación, a la que se le aplica el método deco(String [] s) para convertirla a una etiqueta entendible por el índice intralingüístico del Wordnet y el SentiWordNet. A continuación se usa el método traductor(String s, String pos) de la clase Útil, como se muestra en la **Figura 2.20**Figura 2.20, para traducir el término mediante el índice antes mencionado.

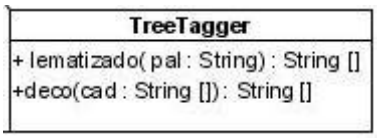


Figura 2.19 Clase TreeTagger.

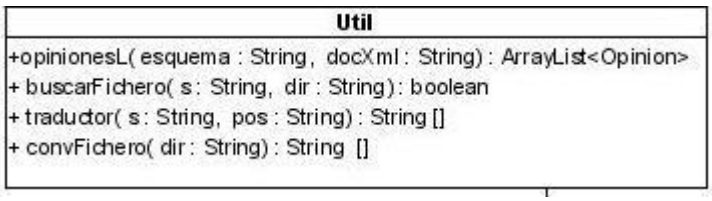


Figura 2.20 Clase Útil.

Para finalizar se utiliza el método buscarFichero(String dir, String p) con el cual se verifica si el término analizado se encuentra en una de las tres listas creadas y así bonificarlo según corresponda. Con el objeto de la clase S, que se muestra en **Figura 2.21**, se recogen los valores positivos y negativos del término en el SentiWordNet mediante el método puntuación(String hd).

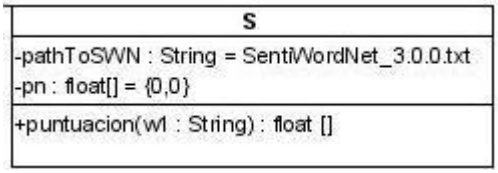


Figura 2.21 Clase S.

El sistema PosNeg Opinión se escribió utilizando el lenguaje de programación Java[19], ya que entre sus principales ventajas permite:

1. Usar la metodología de la programación orientada a objetos.
2. Ejecutar un mismo programa en múltiples sistemas operativos.
3. Ejecutar código en sistemas remotos de forma segura.
4. Usarse con facilidad y toma lo mejor de otros lenguajes orientados a objetos, como C++.

2.3.2 Interfaz visual de PosNeg Opinión

La interfaz visual de la aplicación, como se muestra en la **Figura 2.22**Figura 2.22, se puede dividir en tres partes: la entrada que es donde se inicializan todos los archivos necesarios para comenzar a analizar las opiniones, una segunda parte que son los resultados globales y una tercera donde se muestran una parte de los resultados, especificando aquellas opiniones más positivas y más negativas.

[19] http://java.sun.com/

Dentro del área de entrada, como se muestra en **Figura 2.23**, existe la opción de cargar el fichero XML con las opiniones que se desean procesar.

Figura 2.22 Interfaz visual de PosNeg Opinión.

Figura 2.23 Entrada de datos en PosNeg Opinión.

En el área de resultados globales, como se observa en la **Figura 2.24**, se muestran cuántas opiniones resultaron positivas y cuántas resultaron negativas; además, después se puede elegir mostrar el porciento de opiniones positivas o negativas mediante las etiquetas: Positiva o Negativo.

Figura 2.24 Área de resultados globales.

Por último en el área de muestra, como se observa en la **Figura 2.25**, se presentan las opiniones con mayores valores de positividad y negatividad para que puedan ser analizadas por el usuario.

Figura 2.25 Área de muestra de una parte de los resultados.

2.5 Conclusiones parciales

En este capítulo se analizaron los principales recursos léxicos encontrados en la literatura sobre minería de opinión. Se destaca el recurso SentiWordNet 3.0 por sus potencialidades para ofrecer la polaridad de las opiniones; sin embargo, este recurso tiene varias desventajas que limitan su uso, entre ellas: la mayoría de los términos tienen puntuación nula de polaridad tanto positiva como negativa, al ser un recurso generado automáticamente posee términos mal clasificados y tiene algunos elementos del formato que complejizan su aplicación.

Como resultado de esta investigación se crearon los recursos SentiWordNet 4.0 y 4.1, y el recurso SpanishSentiWordNet. Los dos primeros superan a SentiWordNet 3.0 porque logran

incrementar la cantidad de términos con polaridad asignada a 69051 representando el 81.9% del total de términos incluidos. Además, en la versión 4.1 se rectificó la asignación de algunos valores de polaridad y se explotaron mejor las relaciones entre términos que ofrece WordNet en el cálculo de la polaridad. SpanishSentiWordNet ofrece la polaridad de los términos en Español, facilitando de esta forma el procesamiento de opiniones en español ya que no es necesario buscar todas las acepciones de los términos en inglés ni interactuar con el índice intralingüístico.

Dos esquemas generales para la detección de la polaridad fueron propuestos como resultado de esta investigación. Ellos superan el esquema propuesto en (Amores, Arco & Artiles 2015a) ya que hacen uso de los recursos desarrollados y además incorporan nuevas variantes para manejar problemas presentes en las opiniones tales como jergas, emoticones y las negaciones.

El sistema PosNeg Opinion 3.0 implementa el esquema general final que se presenta en la Figura 2.14. Para ello, se diseñó e implementó la biblioteca PolarityDetection. Este sistema tiene un diseño extensible y generalizable. PosNeg Opinion 3.0 permite que el usuario analice un gran cúmulo de opiniones de manera sencilla y sin preocuparse por toda la teoría que hay detrás de ese análisis. Esta aplicación puede utilizarse como un módulo para una aplicación más general de minería de opinión pues resuelve una de las fases de este proceso y es fácilmente reutilizable, además, se puede comunicar con otras aplicaciones mediante ficheros XML.

Capítulo 3 Valoración de los recursos, esquemas y sistema desarrollados para detectar la polaridad de las opiniones

En este capítulo se describirán las principales colecciones existentes para validar aplicaciones que tributen a la minería de opiniones. También se presentarán las medidas más utilizadas para validar los sistemas que detectan la polaridad de las opiniones. Un análisis comparativo de los recursos léxicos creados respecto a aquellos publicados previamente, así como la evaluación de cada uno de los elementos novedosos que conforman la propuesta resultante de esta investigación serán presentados a continuación.

3.1 Descripción de las colecciones de opiniones textuales

Existen múltiples colecciones de opiniones textuales acerca de disímiles dominios. A continuación describiremos las colecciones más referenciadas en las publicaciones de minería de opinión.

Amazon:

Este corpus contiene comentarios sobre seis productos (cámaras, teléfonos móviles, TV, laptop, tablet y sistemas de video vigilancia) en Amazon.com, cada producto contiene los siguientes atributos: ID del producto (único), nombre, características del producto, precio y el URL de la imagen.

Tabla 3.1 Opiniones del corpus de Amazon.

Producto	Positivas	Negativas
Cámara	356 597	70 146
Teléfonos móviles	112 981	48 921
TV	183 666	32 706
Laptop	27 583	7 501
Tablet	80 952	19 353
Sistema de video vigilancia	38 079	11 623

Cada comentario también contiene además los siguientes atributos: ID del comentario (único), autor, titulo, contenido, puntuación global y fecha. La puntuación global es brindada por el usuario quien le da una puntuación de 0 a 5 estrellas, para este trabajo solo tomamos las puntuaciones de 0 a 2 como negativas, de 4 a 5 como positivas, dejando fuera las puntuaciones

de 3 estrellas pues éstas muestran indecisión de los usuarios al clasificar sus opiniones lo cual no asegura que sea una clasificación confiable. En la **Tabla 3.1** se muestra por cada producto la cantidad de opiniones positivas y negativas.

Mukherjee:

Para este corpus 8507 tweets fueron anotados manualmente, dentro de estos existen alrededor de 2000 diferentes entidades sobre unos 20 dominios. Estos fueron anotados en cuatro clases diferentes: positivo, negativo, objetivo-no-spam y objetivo-spam. Fue utilizada la API de Twitter Tweet Fetcher para tomar tweets automáticamente y de esta forma se logró recolectar 15 214 tweets basados en diferentes hashtag. Algunos de los hashtag utilizados para obtener los tweets positivos fueron #positive, #joy, #excited, #happy, etc. y para los negativos #negative, #sad, #depressed, #gloomy, #disappointed, etc. Para nuestro trabajo solo utilizaremos los tweets clasificados como positivos o negativos.

Opinion Spam_v1.4-2:

Este corpus contiene comentarios positivos y negativos de 20 hoteles de Chicago. La data está compuesta por 400 comentarios positivos extraídos de TripAdvisor, 400 comentarios positivos extraídos de Mechanical Turk, 400 comentarios negativos extraídos de Expedia, Hotels.com, Orbitz, Priceline, TripAdvisor, Yelp y 400 comentarios negativos Mechanical Turk. Cada conjunto de datos consiste en 20 comentarios por cada uno de los 20 hoteles más populares de Chicago.

SenSem:

El corpus SenSem (Sentence Semantics) para el español fue creado como parte de SenSem Databank. El trabajo sobre el Databank comenzó en 2004 y continua hasta hoy, todo el proyecto fue fundado por el gobierno español. La versión 1.1 de este corpus incluye textos de fuentes periodísticas y de literatura clásica conteniendo más de 30 000 sentencias (25 075 de noticias y 5 299 de literatura). La fuente principal de los textos de noticias fue el Periódico de Catalunya y un porcentaje menor pertenece a La Vanguardia. La mayoría de las sentencias sobre las obras literarias han sido extraídas desde internet.

Sentiment140:

Este corpus fue creado en la Universidad de Stanford por Go, Bhayani, y Huang. El conjunto de datos fue creado automáticamente recolectando de la red social Twitter tweets que

contengan emoticonos positivos (☺) y emoticonos negativos (☹), aquellos mensajes que contenían ambos no fueron tomados en cuenta. En total la colección cuenta con 800 000 tweets positivos y 800 000 tweets negativos. Algo interesante en esta colección de tweets es que la mayoría no contiene ningún hashtag.

STS Gold:

Este corpus contiene dos conjuntos textuales, uno de tweets con su polaridad anotada (Positivo o Negativo) y otro con 58 entidades y su polaridad anotada en cinco clases (Positivo, Negativo, Neutro, Mixto, Otro). El etiquetado de este corpus fue realizado manualmente por tres estudiantes de doctorado. Las sentencias agregadas a la versión final del corpus fueron aquellas en la cual los tres estudiantes coincidieron en su anotación de sentimiento, descartando así cualquier posible ruido en la colección. La versión final cuenta con 1 402 tweets negativos, 632 positivos y 77 neutrales. Para nuestro trabajo solo utilizaremos aquellos clasificados como negativos y positivos.

TASS 2013, 2014, 2015:

El Taller sobre Análisis de Sentimientos (TASS) es un evento enmarcado dentro de la XXIX edición del Congreso Anual de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN). El TASS-2013 es la segunda edición de esta competición y se proponen cuatro tareas. La primera consiste en determinar el análisis de sentimientos a nivel global de un tweet; la segunda, consiste en determinar el tópico de un tweet (política, economía, deportes,...); la tercera consiste en el análisis de sentimiento a nivel de entidad dentro de un tweet y la última tarea consiste en la identificación de las tendencias políticas (derechas, centro, izquierda, neutral) de un usuario basándose en sus tweets. Para nuestro trabajo utilizaremos el corpus propuesto para la primera tarea utilizando solo los tweets clasificados como positivos (2884) y los clasificados como negativos (2182). Esta colección no ha sido modificada desde el año 2013.

TripAdvisor:

Esta colección contiene comentarios de huéspedes de varios hoteles de todo el mundo. Además de la puntuación general de cada comentario, este corpus contiene puntuaciones para varios aspectos como habitaciones, locación, limpieza, recepción, servicios generales y servicios de negocios. Todos estos aspectos no están siempre en cada comentario. Para nuestro

trabajo solo utilizaremos la puntuación general que está dada por el usuario en un rango de cinco estrellas, de la cual tomaremos como negativos aquellos comentarios con una o dos estrellas y como positivos aquellos con cuatro y cinco estrellas, descartando los clasificados con tres estrellas ya que pueden introducir ruidos a la colección. La colección a utilizar en nuestra experimentación finalmente quedó compuesta por 181 747 comentarios positivos y 34 983 negativos.

LARA:

Este corpus está compuesto por dos colecciones: una es un subconjunto de comentarios sobre hoteles tomados del corpus de TripAdvisor y la otra son cometarios extraídos de amazon.com sobre reproductores de mp3. En la colección de los hoteles tomaron además de la puntuación general la puntuación otorgada a cada uno de los siete aspectos: generalidades, valor, habitaciones, ubicación, limpieza, registro y servicio. Para la construcción de la colección final se pre-procesaron todos los comentarios y eliminaron aquellos que tuvieran alguno de los aspectos ausentes y aquellos con menos de 50 palabras, también se eliminaron los signos de puntuación y se llevaron todas la palabras a minúsculas. La colección final quedó conformada por 10 911 comentarios positivos y 3 187 negativos

Large Movie Review:

Este corpus contiene comentarios sobre películas extraídos de IMDb y una clasificación binaria de polaridad para cada uno (positivo o negativo). En toda la colección no existen más de 30 comentarios de una misma película ya que los comentarios sobre una misma película tienden a tener una misma calificación. La anotación de la polaridad fue dada tomando en cuenta la puntuación en una escala de 10 ofrecida por el usuario tomando como positivos aquellos con una evaluación igual o superior a 7, como negativos aquellos con una puntuación igual o inferior a 4 y se descartaron aquellos con puntuación de 5 o 6 por considerase neutrales. La colección quedo compuesta por 12 500 comentarios positivos y 12 500 comentarios negativos.

SEMEVAL 2013 (Subtareas A y B)

La Tarea 2, llamada análisis de sentimiento en Twitter, de la competición SemEval 2013 incluye dos subtareas, la subtarea A que es a nivel de expresiones y la subtarea B que es a nivel de sms. Para ambas tareas, la anotación de las colecciones fue realzada sobre Mechanical

Turk. Cada oración fue anotada por cinco trabajadores de Mechanical Turk (Turkers), para garantizar la calidad cada Turker tenía que tener una aprobación de más del 95% de los demás trabajadores. Esta clasificación incluía la clase neutra que para nuestro trabajo no es relevante por lo cual la excluimos. La **Tabla 3.2** muestra la cantidad de oraciones y mensajes positivos y negativos.

Tabla 3.2 Opiniones de cada subtarea de la tarea 2 de la competición Semeval 2013

Subtarea	Positivos	Negativos
Subtarea A	1 572	601
Subtarea B	492	394

SFU Review Raw:

Este corpus fue principalmente generado de comentarios extraídos desde el sitio web Epinions, para la construcción de este corpus se enfocaron en la frecuencia de aparición de los adjetivos presentes en los cometarios. En esta colección los comentarios están divididos en ocho categorías (libros, carros, computadoras, implementos de cocina, hoteles, películas, música y teléfonos) con 25 comentarios positivos y 25 comentarios negativos cada una. La clasificación está basada en la etiqueta "recomendada" o "no recomendada" ofrecida por el usuario. De forma general la colección cuenta con 200 comentarios positivos y 200 comentarios negativos.

3.2 Formas de evaluación del cálculo de la polaridad

Han sido definidos varios términos estándar para medir el desempeño de un clasificador. A continuación describiremos los más utilizados.

- La Exactitud (Accuracy; Ac) es la proporción del número total de predicciones que fueron correctas

$$Ac = \frac{a + b}{a + b + c + d} \tag{3.1}$$

- La Razón de Verdaderos Positivos (True Positive Rate; TPR), a veces también denominada Recall, es la proporción de casos positivos que fueron correctamente identificados.

$$TPR = \frac{d}{c + d} \tag{3.2}$$

- La Razón de Falsos Positivos (False Positive Rate; FPR) es la proporción de casos negativos que han sido incorrectamente clasificados como positivos.

$$FPR = \frac{b}{a + b} \tag{3.3}$$

- La Razón de Verdaderos Negativos (True Negative Rate; TNR) es la proporción de casos negativos que han sido correctamente clasificados.

$$TNR = \frac{a}{a + b} \tag{3.4}$$

- La Razón de Falsos Negativos (False Negative Rate; FNR) es la proporción de casos positivos que fueron incorrectamente clasificados como negativos.

$$FNR = \frac{c}{c + d} \tag{3.5}$$

- La Precisión (Precision; P) es la proporción de casos predichos positivos que fueron correctos

$$P = \frac{d}{b + d} \tag{3.6}$$

Donde *a* es el número de predicciones correctas de que un caso es negativo, *b* es el número de predicciones incorrectas de que un caso es positivo, *c* es el número de predicciones incorrectas de que un caso es negativo y *d* es el número de predicciones correctas de que un caso es positivo.

Todas estas expresiones suelen referirse en términos de fracciones por 1, como han sido formuladas, o también, en términos de porcientos (multiplicadas por 100). La Exactitud o *Accuracy* ha sido muy usada para determinar la eficiencia de un clasificador pero ella puede no ser una medida adecuada del desempeño de un clasificador cuando el número de casos negativos es mucho mayor que el número de casos positivos. Suponga por ejemplo que hay 1000 casos en la muestra de los cuales 995 son negativos y 5 son positivos. Si el sistema clasifica a todos como negativos la exactitud sería del 99.5% aun cuando el clasificador se equivoque en todos los positivos. Para este caso es recomendado usar la medida-F (F-mesuare). El Valor-F (denominada también F-score o medida-F) en estadística es la medida

de precisión que tiene un test. Se emplea en la determinación de un valor único ponderado de la precisión y la exactitud. El valor F se considera como una media armónica que combina los valores de la precisión y de la exactitud. De tal forma que:

$$Medida - F = \frac{2 * P * Ac}{P + Ac} \tag{3.7}$$

3.3 Análisis comparativo de los recursos léxicos

El objetivo de este estudio consiste en evidenciar las potencialidades que ofrecen las versiones 4.0 y 4.1 de SentiWordNet respecto a la versión 3.0 para la detección de la polaridad de las opiniones. Para este análisis se tomó como caso de estudio el corpus de los comentarios de Amazon, y en la aplicación PosNeg Opinion 3.0 solo se utilizaron las versiones 3.0, 4.0 y 4.1 del recurso léxico SentiWordNet para evaluar la polaridad de los términos. De esta forma podremos evaluar la efectividad de cada uno de ellos por si solos. Los resultados de estas comparaciones se muestran en la **Tabla 3.3** y en la **Tabla 3.4**.

Tabla 3.3 Sentiwordnet 3.0 V.S. Sentiwordnet 4.0

Recurso	**Precisión**	**Recall**	**Exactitud**	**F1**
SentiWordNet 3.0	87 %	82%	75%	85%
SentiWordNet 4.0	**88%**	**89%**	**81%**	**88%**

Tabla 3.4 Sentiwordnet 4.0 V.S. Sentiwordnet 4.1

Recurso	**Precisión**	**Recall**	**Exactitud**	**F1**
SentiWordNet 4.0	88%	89%	81%	88%
SentiWordNet 4.1	**91%**	**89%**	**84%**	**90%**

Como se puede apreciar en la **Tabla 3.3** la propuesta de SentiWorNet 4.0 logra superar claramente a su versión anterior logrando altos resultados de precisión, recall, exactitud y F1. También en la **Tabla 3.4** podemos ver como los cambios realizados al SentiWorNet 4.0 para crear el SentiWorNet 4.1 producen una mejoría en todas las medidas expuestas.

3.4 Evaluación de los recursos desarrollados para el procesamiento del lenguaje natural en la detección de la polaridad

En este epígrafe se desea mostrar las potencialidades que ofrecen los recursos desarrollados para el manejo de jergas, emoticonos, negaciones y el uso de palabras modificadoras en el cálculo de la polaridad de las opiniones. Para realizar esta evaluación formaremos tres casos de estudios. Los dos primeros serán un subconjunto del corpus Amazon y el tercero será el corpus Sentiment140.

3.4.1 Impacto de las palabras intensificadoras y minimizadoras

Para verificar el impacto que tiene las palabras intensificadoras y minimizadoras en la clasificación de la polaridad tomamos del corpus Amazon aquellas opiniones que contuviesen dichas palabras quedando así una colección de 99 148 opiniones negativas y 645 678 opiniones positivas. En la **Tabla 3.5** se muestra como al realizar el análisis de dichas palabras se logran mejores medidas de precison, recall, exactitud y F1.

Tabla 3.5 Impacto de las palabras intensificadoras y minimizadoras.

Propuesta	**Precisión**	**Recall**	**Exactitud**	**F1**
Sin intensificadores y minimizadores	89%	83%	82%	86%
Con intensificadores y minimizadores	**92%**	**87%**	**83%**	**90%**

3.4.2 Impacto del manejo de las negaciones en la detección de la polaridad

Igualmente para mostrar el impacto que tiene la negación en la clasificación de la polaridad tomamos del corpus Amazon aquellas opiniones que contuviesen palabras negadoras quedando así una colección de 65 578 opiniones negativas y 225 758 opiniones positivas. En la **Tabla 3.6** se muestra como al realizar el análisis de la negación se logran mejores medidas de precisión, recall, exactitud y F1.

Tabla 3.6 Impacto del manejo de las negaciones en la detección de la polaridad.

Propuesta	**Precisión**	**Recall**	**Exactitud**	**F1**
Sin negación	88%	86%%	80%	87%
Con negación	**90%**	**88%**	**83%**	**89%**

También en la **Tabla 3.7** se puede notar que al realizar este análisis se logra una mayor proporción de casos positivos y negativos que fueron correctamente identificados (TPR y TNR) y una disminución de la proporción de casos negativos que han sido incorrectamente

clasificados como positivos (FPR), así como una disminución de casos positivos que han sido incorrectamente clasificados como negativos (FNR).

Tabla 3.7 Valores de TPR, FPR, TNR, FNR para el análisis de la negación.

Propuesta	**TPR**	**FPR**	**TNR**	**FNR**
Sin negación	86%	51%	49%	14%
Con negación	**88%**	**42%**	**58%**	**12%**

3.4.3 Impacto del manejo de los emoticones en la detección de la polaridad

Para verificar el impacto que tienen los emoticones en la clasificación de la polaridad tomamos el corpus Sentiment140 ya que parte de éste fue construido a partir de la búsqueda de diferentes emoticones. En la **Tabla 3.8** se muestra como al realizar el análisis de este corpus se obtienen mejores resultados de precisión, recall, exactitud y F1 al incluir el análisis de los emoticonos.

Tabla 3.8 Impacto del manejo de los emoticones en la detección de la polaridad.

Propuesta	**Precisión**	**Recall**	**Exactitud**	**F1**
Sin emoticonos	83%	90%	77%	86%
Con emoticonos	**88%**	**92%**	**83%**	**90%**

3.4.4 Impacto del manejo de las jergas en la detección de la polaridad

Para el análisis del impacto de las jergas también utilizamos el corpus Sentiment140 ya que este se encuentra construido sobre la red social Twitter donde es común la utilización de estos elementos. En la **Tabla 3.9** se muestra como realizar el análisis de las jergas ayuda a la obtención de mejores resultados de precisión, recall, exactitud y F1.

Tabla 3.9 Impacto del manejo de las jergas en la detección de la polaridad.

Propuesta	**Precisión**	**Recall**	**Exactitud**	**F1**
Sin jergas	83%	90%	77%	86%
Con jergas	**87%**	**90%**	**80%**	**88%**

3.5 Valoración de PosNeg Opinion 3.0

En este epígrafe pretendemos analizar cuáles son los resultados obtenidos por los diferentes enfoques en la literatura y mostrar los resultados obtenidos por nuestra aplicación en todas las colecciones descritas en el epígrafe 3.1. También compararemos nuestros resultados con otros enfoques que hayan realizado su experimentación sobre algunas de estas colecciones.

3.5.1 Descripción de los principales resultados obtenidos en la literatura para la detección de la polaridad

Tabla 3.10 Resultados de distintos enfoques encontrados en la literatura.

Enfoque	Técnica	Trabajo, Año	Corpus	Exactitud	Dominio
Aprendizaje automático (supervisado)	SVM	(Dang et al. 2010)	Comentarios	82,7%	Marketing
		(Khairnar & Kinikar 2013)	Comentarios	----------	General
		(Saleh et al. 2011)	Comentarios	91,5%	
		(Samsudin et al. 2013a)	Comentarios	82,9%	
	NB	(Zhang et al. 2011)	Comentarios	84,5%	Marketing
		(Go et al. 2009)	Tweets	82,7%	General
		(Mahendran et al. 2013)	Tweets	----------	
		(Samsudin et al. 2013a)	Comentarios	84,9%	
		(Smeureanu et al. 2012)	Comentarios	81,4%	
		(Samsudin et al. 2013b)	Mensajes	91,4%	
		(Kishore & Naidu 2013)	Comentarios	----------	
	ME	(Go et al. 2009)	Tweets	83%	
		(Smeureanu et al. 2012)	Comentarios	77,1%	
	KNN	(Samsudin et al. 2013a)	Comentarios	64,1%	
		(Samsudin et al. 2013b)	Mensajes	79,8%	
Basado en recurso léxicos	Basado en diccionarios	(Hamouda & Rohaim 2011)	Comentarios	67% - 68,6%	General
		(Fei et al. 2012)	Tweets	81,2%	
		(Dang et al. 2010)	Comentarios	82,7%	
	Basado en corpus	(Sobkowicz et al. 2012)	Comentarios	76,8%	
		(Riloff & Wiebe 2003)	Comentarios	71% - 85%	
		(Turney 2002)	Comentarios	74,39%	

La

Tabla 3.10, la cual fue construida a partir de la información de tres estudios (Othman et al. 2014; Kaur & Duhan 2015; Vasantharaj et al. 2015), dos de ellos realizados en el 2015 y el tercero en el 2014, muestra los resultados obtenidos por diferentes enfoques existentes en la literatura. Podemos apreciar que los mejores resultados son obtenidos por los enfoques supervisados.

En la **Tabla 3.11** podemos ver que el mayor valor de exactitud encontrado para los enfoques supervisados fue de 91.5%, la media fue de 75.3% y el menor valor encontrado fue de 64.1%. Para los enfoques no supervisados el mayor valor fue de 85%, la media fue de 77% y el menor valor fue de 67%.

Tabla 3.11 Máximo, mínimo y media de los valores de exactitud encontrados en cada enfoque.

Enfoques en la literatura	Máximo, Mínimo y Media de la Exactitud
Aprendizaje automatizado	Max: 91,5% Min: 64,1% Med: 75.3%
Basado en recursos léxicos	Max: 85% Min: 67% Med: 77%

3.5.2 Validación de PosNeg Opinion 3.0

En la **Tabla 3.12** se muestra el máximo, la media y el menor valor de exactitud y F1 alcanzados al analizar las colecciones anteriormente descritas con PosNeg Opinion 3.0. Se puede apreciar que estos valores superan los resultados encontrados en la literatura tanto de los enfoques supervisados, los cuales hasta ahora eran los que mejores resultados presentaban en la literatura, como los no supervisados, que es el enfoque que nosotros seguimos.

En las **tablas Tabla 3.13**,
Tabla 3.14, **Tabla 3.15**, **Tabla 3.16**,

Tabla 3.17, **¡Error! No se encuentra el origen de la referencia.**, **Tabla 3.18** se presenta la comparación de PosNeg Opinion 3.0 con trabajos que hayan presentado resultados sobre

algunas de las colecciones utilizadas en nuestra experimentación. En la totalidad de los casos, PosNeg Opinon 3.0 logra superar a las demás aplicaciones previamente publicadas.

Tabla 3.12 Comparación global de PosNeg Opinion 3.0

Corpus examinados por PosNeg Opinion 3.0	**Máximo, Mínimo y Media de la Exactitud**	**Máximo, Mínimo y Media de la F1**
Amazon reviews Mukherjee Opinion Spam v1.4-2 SenSem Sentiment140 STS-Gold Large Movie Review SemEval 2013 SMS SemEval 2013 Tweets SFU Review TASS 2013 LARA	Max: 90% Min: 75% Med: 85.3%	Max: 91% Min: 76% Med: 86.6%

Tabla 3.13 Comparación de PosNeg Opinion 3.0 sobre el corpus Mukherjee.

Propuesta	**Precisión**	**Recall**	**Exactitud**	**F1**
(Mukherjee, Malu, et al. 2012)	------	------	66.69%	------
(Mukherjee, Bhattacharyya, et al. 2012)	------	------	72.81%	------
PosNeg Opinion 3.0	74%	94%	**79%**	83%

Tabla 3.14 Comparación de PosNeg Opinion 3.0 sobre el corpus Opinion Spam_v1.4-2.

Propuesta	**Precisión**	**Recall**	**Exactitud**	**F1**
(Ott et al. 2013)	87.7%	89.3%	88.4%	88.5%
(Ott et al. 2013)	85.3%	87.0%	86.0%	86.1%
PosNeg Opinion 3.0	**99%**	**83%**	**90%**	**91%**

Tabla 3.15 Comparación de PosNeg Opinion 3.0 sobre el corpus Stanford.

Propuesta	**Precisión**	**Recall**	**Exactitud**	**F1**
(Kouloumpis et al. 2011)	------	------	75%	68%
(dos Santos & Gatti 2014)	------	------	86.4%	------
(Socher et al. 2013)	------	------	84.7%	------
(Socher et al. 2013)	------	------	83%	------
(Socher et al. 2013)	------	------	82.7%	------
(Socher et al. 2013)	------	------	82.2%	------

(Saif et al. 2012)	------	------	86.3%	------
PosNeg Opinion 3.0	**89%**	**93%**	**90%**	**91%**

Tabla 3.16 Comparación de PosNeg Opinion 3.0 sobre el corpus STS Gold.

Propuesta	**Precisión**	**Recall**	**Exactitud**	**F1**
(Aston et al. 2014)	------	------	87.5%	------
(Wilson et al. 2005)	------	------	57.47%	57.46%
(Baccianella et al. 2010)	------	------	56.64%	55.92%
(Thelwall et al. 2012)	------	------	81.32%	78.56%
(Saif et al. 2014)	------	------	80.33%	77.52%
(Saif et al. 2013)	------	------	85.69%	82.454%
PosNeg Opinion 3.0	87%	92%	**89%**	**90%**

Tabla 3.17 Comparación de PosNeg Opinion 3.0 sobre el corpus de la tarea 1 del TASS.

Competición	**Propuesta**	**Precisión**	**Recall**	**Exactitud**	**F1**
TASS 2014	ELiRF-UPV	------	------	71%	------
	Elhuyar	------	------	70%	------
	LyS	------	------	67%	------
	SINAIword2vec	------	------	61%	------
	JRC	------	------	61%	------
	SINAI-ESMA	------	------	61%	------
	IPN	------	------	55%	------
	PosNeg Opinion 3.0	70%	83%	**75%**	76%

Tabla 3.18 Comparación de PosNeg Opinion 3.0 sobre el corpus Large Movie Review

Propuesta	**Precisión**	**Recall**	**Exactitud**	**F1**
(Maas et al. 2011)	------	------	88.89%	------
PosNeg Opinion 3.0	88%	89%	**89%**	89%

3.6 Conclusiones parciales

En este capítulo se muestran los resultados experimentales obtenidos a partir de la aplicación de PosNeg Opinion sobre 12 colecciones disponibles en internet para evaluar aplicaciones que tributan a la minería de opinión. Se evidenció que el cálculo de la polaridad de las opiniones siguiendo el esquema general finalmente propuesto utilizando los recursos léxicos SentiWordNet 4.0 y 4.1 permitió obtener mejores valores de Precisión, Recall, Exactitud y F1

que cuando se utilizaba el SentiWordNet 3.0. También se mostró el impacto positivo que tiene el manejo de las palabras modificadoras, las negaciones, los emoticones y las jergas en el cálculo de la polaridad mediante el uso de los recursos diseñados como resultados de esta investigación a tales efectos. Se compararon los resultados obtenidos por PosNeg Opinion 3.0 siguiendo el esquema general finalmente propuesto con los mejores resultados reportados en la literatura al detectar la polaridad de las opiniones presentes en las 12 colecciones seleccionadas para la experimentación y para cada una de ellas, PosNeg Opinion 3.0 superó los resultados de referencia.

Conclusiones y recomendaciones

Como resultado de esta investigación se desarrolló el sistema PosNeg Opinion 3.0 que implementa el esquema general propuesto para la detección no supervisada de la polaridad de las opiniones a partir del empleo de nuevos recursos léxicos y del manejo de jergas, emoticones, palabras modificadoras y negaciones, obteniéndose valores promedios de exactitud y F1 del 85%; cumpliéndose de esta forma el objetivo general propuesto, ya que:

- Se identificaron y analizaron los principales recursos léxicos encontrados en la literatura sobre minería de opinión, destacándose el recurso SentiWordNet 3.0 por sus potencialidades para ofrecer la polaridad de las opiniones; sin embargo, este recurso tiene varias desventajas que limitan su uso, entre ellas: la mayoría de los términos tienen puntuación nula de polaridad tanto positiva como negativa, al ser un recurso generado automáticamente posee términos mal clasificados y tiene algunos elementos del formato que complejizan su aplicación. De ahí que se concluyera que este recurso debía ser rectificado para contribuir de manera efectiva al cálculo de la polaridad de las opiniones mediante el uso de sistemas no supervisados basados en recursos léxicos.
- Los recursos creados SentiWordNet 4.0 y 4.1, y SpanishSentiWordNet permiten elevar los valores de Exactitud, Recall, Precisión y F1 en la detección no supervisada de la polaridad de las opiniones en Inglés y en Español, respecto a las facilidades que ofrecía SentiWordNet 3.0 solo para el idioma Inglés. SentiWordNet en sus versiones 4.0 y 4.1 logra incrementar la cantidad de términos con polaridad asignada a 69051, representando el 81.9% del total de términos, superando significativamente al SentiWordNet 3.0 que solo tiene el 4,2% de los términos puntuados. Además, en la versión 4.1 se rectificó la asignación de algunos valores de polaridad y se explotaron mejor las relaciones entre términos que ofrece WordNet en el cálculo de la polaridad. SpanishSentiWordNet ofrece la polaridad de los términos en Español, facilitando de esta forma el procesamiento de opiniones en español ya que no es necesario buscar todas las acepciones de los términos en inglés ni interactuar con el índice intralingüístico.
- Dos esquemas generales para la detección de la polaridad fueron propuestos como resultado de esta investigación. Ellos superan el esquema propuesto en (Amores, Arco

& Artiles 2015a) ya que hacen uso de los recursos creados, y además, el esquema finalmente propuesto incorpora nuevas variantes para manejar problemas presentes en las opiniones tales como jergas, emoticones, palabras modificadoras y las negaciones.

- El sistema PosNeg Opinion aplicado sobre algunas de las 12 colecciones disponibles en internet para evaluar aplicaciones que tributan a la minería de opinión mostró el impacto positivo que tiene el manejo de las palabras modificadoras, las negaciones, los emoticones y las jergas en el cálculo de la polaridad mediante el uso de los recursos diseñados como resultado de esta investigación a tales efectos; lográndose aumentar los valores de las medidas de calidad cuando se utilizaron estos recursos respecto al procesamiento sin el uso de estos.
- Se diseñó e implementó la biblioteca PolarityDetection, permitiendo de esta forma un diseño extensible y generalizable de la herramienta PosNeg Opinion 3.0, la cual permite que el usuario analice un gran cúmulo de opiniones de manera sencilla y sin preocuparse por toda la teoría que hay detrás de ese análisis. Esta aplicación puede utilizarse como un módulo para una aplicación más general de minería de opinión pues resuelve una de las fases de este proceso y es fácilmente reutilizable, además, se puede comunicar con otras aplicaciones mediante ficheros XML.
- Se compararon los resultados obtenidos por PosNeg Opinion 3.0 siguiendo el esquema general finalmente propuesto con los mejores resultados reportados en la literatura al detectar la polaridad de las opiniones presentes en las 12 colecciones seleccionadas para la experimentación y para cada una de ellas, PosNeg Opinion 3.0 superó los resultados de referencia.

Derivadas del estudio realizado, así como de las conclusiones generales emanadas del mismo, se recomienda:

- Integrar PosNeg Opinion al marco de trabajo OpinionTopicDetection de forma tal que se logre calcular la polaridad de las opiniones por tópicos.
- Continuar actualizando las puntuaciones asociadas de los términos en el recurso SentiWordNet mediante la consulta a listas de términos con polaridades asignadas así como retroalimentándonos de los resultados de la experimentación.

Referencias bibliográficas

Abbasi, A. et al., 2013. Automatic seed word selection for unsupervised sentiment classification of Chinese text. *LREC*, 2(1), pp.443–447.

Abbasi, A. et al., 2011. Selecting attributes for sentiment classification using feature relation networks. *Knowledge and Data Engineering, IEEE Transactions on*, 23(3), pp.447–462.

Abbasi, A., Chen, H. & Salem, A., 2008. Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums. *ACM Transactions on Information Systems (TOIS)*, 26(3), p.12.

Abdul-Mageed, M. & Diab, M.T., 2014. SANA: A Large Scale Multi-Genre, Multi-Dialect Lexicon for Arabic Subjectivity and Sentiment Analysis. In *LREC*. pp. 1162–1169.

Abdul-Mageed, M., Diab, M.T. & Korayem, M., 2011. Subjectivity and sentiment analysis of modern standard arabic. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*. pp. 587–591.

Alm, E.C.O., 2008. *Affect in text and speech*, BiblioBazaar.

Amores, M., 2013. *Detección no supervisada de la polaridad de las opiniones en Español*. Universidad Central "Marta Abreu" de las Villas.

Amores, M., Arco, L. & Artiles, M., 2015a. PosNeg opinion: Una herramienta para gestionar comentarios de la Web. *Revista Cubana de Ciencias Inform{á}ticas*, 9(1), pp.12–20.

Amores, M., Arco, L. & Artiles, M., 2015b. PosNeg opinion: Una herramienta para gestionar comentarios de la Web PosNeg opinion: A tool for managing comments from the web. *Revista Cubana de Ciencias Informáticas*, 9(1). Available at: http://rcci.uci.cu.

Amores, M., Arco, L. & Borroto, C., 2015. SentiWordNet 4.0 and SpanishSenti-WordNet assisting Polarity Detection. In *Eureka Workshop*.

Angulakshmi, G. & ManickaChezian, R., 2014. An analysis on opinion mining: techniques and tools. *International Journal of Advanced Research in Computer Communication Engineering*, 3(7), pp.7483–7487.

Annett, M. & Kondrak, G., 2008. A comparison of sentiment analysis techniques: Polarizing movie blogs. In *Advances in artificial intelligence*. Springer, pp. 25–35.

Anwer, N., Rashid, A. & Hassan, S., 2010. Feature based opinion mining of online free format customer reviews using frequency distribution and bayesian statistics. In *Networked Computing and Advanced Information Management (NCM), 2010 Sixth International Conference on*. pp. 57–62.

Archak, N., Ghose, A. & Ipeirotis, P.G., 2007. Show me the money!: deriving the pricing power of product features by mining consumer reviews. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. pp. 56–65.

Aston, N. et al., 2014. Sentiment analysis on the social networks using stream algorithms. *Journal of Data Analysis and Information Processing*, 2014.

Baccianella, S., Esuli, A. & Sebastiani, F., 2010. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In *LREC*. pp. 2200–2204.

Bahrainian, S.-A. & Dengel, A., 2013. Sentiment analysis and summarization of twitter data. In *Computational Science and Engineering (CSE), 2013 IEEE 16th International Conference on*. pp. 227–234.

Balahur, A. & Turchi, M., 2014. Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis. *Computer Speech & Language*, 28(1), pp.56–75.

Barbosa, L. & Feng, J., 2010. Robust sentiment detection on twitter from biased and noisy data. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. pp. 36–44.

Becker, L. et al., 2013. Avaya: Sentiment analysis on twitter with self-training and polarity lexicon expansion. In *Second Joint Conference on Lexical and Computational Semantics (* SEM)*. pp. 333–340.

Beineke, P. et al., 2003. An exploration of sentiment summarization. In *Proceedings of AAAI*. pp. 12–15.

Benamara, F. et al., 2011. Towards Context-Based Subjectivity Analysis. In *IJCNLP*. pp. 1180–1188.

Bespalov, D. et al., 2011. Sentiment classification based on supervised latent n-gram analysis. In *Proceedings of the 20th ACM international conference on Information and knowledge management*. pp. 375–382.

Bethard, S. et al., 2004. Automatic extraction of opinion propositions and their holders. In *2004 AAAI Spring Symposium on Exploring Attitude and Affect in Text*.

Blair-Goldensohn, S. et al., 2008. Building a sentiment summarizer for local service reviews. In *WWW Workshop on NLP in the Information Explosion Era*. pp. 339–348.

Bloom, K., Garg, N. & Argamon, S., 2007. Extracting Appraisal Expressions. In *HLT-NAACL*. pp. 308–315.

Boiy, E. & Moens, M.-F., 2009. A machine learning approach to sentiment analysis in multilingual Web texts. *Information retrieval*, 12(5), pp.526–558.

Bollen, J., Mao, H. & Zeng, X., 2011. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), pp.1–8.

Borroto, C., 2014. *Creación y perfeccionamiento de herramientas para la minería de opinión en idioma Español*. Universidad Central "Marta Abreu" de las Villas.

Branavan, S.R.K. et al., 2009. Learning document-level semantic properties from free-text annotations. *Journal of Artificial Intelligence Research*, pp.569–603.

Brody, S. & Elhadad, N., 2010. An unsupervised aspect-sentiment model for online reviews. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. pp. 804–812.

Brooke, J., Tofiloski, M. & Taboada, M., 2009. Cross-Linguistic Sentiment Analysis: From English to Spanish. In *RANLP*. pp. 50–54.

Buche, A., Chandak, D. & Zadgaonkar, A., 2013. Opinion mining and analysis: a survey. *arXiv preprint arXiv:1307.3336*.

Cabral Cavalcanti, D., 2011. *Uma abordagem não supervisionada para classificação de opinião usando o recurso léxico SentiWordNet*. Universidade Federal de Pernambuco.

Cambria, E. et al., 2013. Knowledge-based approaches to concept-level sentiment analysis. *IEEE Intelligent Systems*, (2), pp.12–14.

Cardie, C. et al., 2003. Combining Low-Level and Summary Representations of Opinions for Multi-Perspective Question Answering. In *New directions in question answering*. pp. 20–27.

Carenini, G., Cheung, J.C.K. & Pauls, A., 2013. MULTI-DOCUMENT SUMMARIZATION

OF EVALUATIVE TEXT. *Computational Intelligence*, 29(4), pp.545–576.

Cerini, S. et al., 2007. Micro-WNOp: A gold standard for the evaluation of automatically compiled lexical resources for opinion mining. *Language resources and linguistic theory: Typology, second language acquisition, English linguistics*, pp.200–210.

Cernian, A., Sgarciu, V. & Martin, B., 2015. Sentiment analysis from product reviews using SentiWordNet as lexical resource. In *Electronics, Computers and Artificial Intelligence (ECAI), 2015 7th International Conference on*. p. WE–15.

Chandrakala, S. & Sindhu, C., 2012. Opinion Mining and sentiment classification a survey. *ICTACT journal on soft computing*, 3(1), pp.420–425.

Chaovalit, P. & Zhou, L., 2005. Movie review mining: A comparison between supervised and unsupervised classification approaches. In *System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on*. p. 112c–112c.

Chen, L.-S. & Chiu, H.-J., 2009. Developing a neural network based index for sentiment classification. *IAENG Hong Kong*.

Chesley, P. et al., 2006. Using verbs and adjectives to automatically classify blog sentiment. *Training*, 580(263), p.233.

Choi, Y. et al., 2005. Identifying sources of opinions with conditional random fields and extraction patterns. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*. pp. 355–362.

Choi, Y., Breck, E. & Cardie, C., 2006. Joint extraction of entities and relations for opinion recognition. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. pp. 431–439.

Choi, Y. & Cardie, C., 2010. Hierarchical sequential learning for extracting opinions and their attributes. In *Proceedings of the ACL 2010 conference short papers*. pp. 269–274.

Cruz, F.L. et al., 2011. Automatic expansion of feature-level opinion lexicons. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*. pp. 125–131.

Dang, Y., Zhang, Y. & Chen, H., 2010. A lexicon-enhanced method for sentiment classification: An experiment on online product reviews. *Intelligent Systems, IEEE*, 25(4), pp.46–53.

Das, D. & Martins, A.F.T., 2007. A survey on automatic text summarization. *Literature Survey for the Language and Statistics II course at CMU*, 4, pp.192–195.

Das, S. & Chen, M., 2001. Yahoo! for Amazon: Extracting market sentiment from stock message boards. In *Proceedings of the Asia Pacific finance association annual conference (APFA)*. p. 43.

Dave, K., Lawrence, S. & Pennock, D.M., 2003. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web*. pp. 519–528.

Deshpande, D.S., 2012. A survey on web data mining applications. *IJCA Proceedings on Emerging Trends in Computer Science and Information Technology (ETCSIT2012) etcsit1001*, (3), pp.28–32.

Díaz, N.P.C., 2014. *Detección de la Negación y la Especulación en Textos Médicos y de Opinión*. Universidad de Huelva.

Ding, X., Liu, B. & Yu, P.S., 2008. A holistic lexicon-based approach to opinion mining. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*. pp. 231–240.

Education, A.S., 2004. Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. *Bo Pang and Lillian Lee*.

Efron, M., 2004. Cultural orientation: Classifying subjective documents by cocitation analysis. In *AAAI Fall Symposium on Style and Meaning in Language, Art, and Music*. pp. 41–48.

Elarnaoty, M., AbdelRahman, S. & Fahmy, A., 2012. A machine learning approach for opinion holder extraction in Arabic language. *arXiv preprint arXiv:1206.1011*.

Esuli, A., 2008. Automatic generation of lexical resources for opinion mining: models, algorithms and applications. In *ACM SIGIR Forum*. pp. 105–106.

Esuli, A. & Sebastiani, F., 2006a. Determining Term Subjectivity and Term Orientation for Opinion Mining. In *EACL*. p. 2006.

Esuli, A. & Sebastiani, F., 2005. Determining the semantic orientation of terms through gloss classification. In *Proceedings of the 14th ACM international conference on Information and knowledge management*. pp. 617–624.

Esuli, A. & Sebastiani, F., 2009. Enhancing opinion extraction by automatically annotated

lexical resources. In *Human Language Technology. Challenges for Computer Science and Linguistics*. Springer, pp. 500–511.

Esuli, A. & Sebastiani, F., 2007. *SENTIWORDNET: A high-coverage lexical resource for opinion mining*,

Esuli, A. & Sebastiani, F., 2006b. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC*. pp. 417–422.

Fan, M. & Wu, G., 2012. Opinion Summarizationof CustomerComments. *Physics Procedia*, 24, pp.2220–2226.

Fei, G. et al., 2012. A Dictionary-Based Approach to Identifying Aspects Im-plied by Adjectives for Opinion Mining. In *24th International Conference on Computational Linguistics*. p. 309.

Fellbaum, C., 1998. *WordNet*, Wiley Online Library.

Ferguson, P. et al., 2009. Exploring the use of paragraph-level annotations for sentiment analysis of financial blogs. *WOMAS 2009 - Workshop on Opinion Mining and Sentiment Analysis*.

Gamon, M. et al., 2005. Pulse: Mining customer opinions from free text. In *Advances in Intelligent Data Analysis VI*. Springer, pp. 121–132.

Ganesan, K., Zhai, C. & Han, J., 2010. Opinosis: a graph-based approach to abstractive summarization of highly redundant opinions. In *Proceedings of the 23rd international conference on computational linguistics*. pp. 340–348.

Ganter, V. & Strube, M., 2009. Finding hedges by chasing weasels: Hedge detection using Wikipedia tags and shallow linguistic features. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*. pp. 173–176.

Garrido Merchán, E., 2015. *Expansión supervisada de léxicos polarizados adaptable al contexto*. ETSI_Informatica.

Ghiassi, M., Skinner, J. & Zimbra, D., 2013. Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with applications*, 40(16), pp.6266–6282.

Go, A., Bhayani, R. & Huang, L., 2009. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1, p.12.

Goldberg, A.B. & Zhu, X., 2006. Seeing stars when there aren't many stars: graph-based semi-supervised learning for sentiment categorization. In *Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing*. pp. 45–52.

Greene, S. & Resnik, P., 2009. More than words: Syntactic packaging and implicit sentiment. In *Proceedings of human language technologies: The 2009 annual conference of the north american chapter of the association for computational linguistics*. pp. 503–511.

Grefenstette, G. et al., 2004. Coupling niche browsers and affect analysis for an opinion mining application. In *Coupling approaches, coupling media and coupling languages for information retrieval*. pp. 186–194.

Griffiths, T.L. et al., 2004. Integrating Topics and Syntax. In *NIPS*. pp. 537–544.

Gryc, W. & Moilanen, K., 2014. Leveraging textual sentiment analysis with social network modelling. *From Text to Political Positions: Text Analysis Across Disciplines*, 55, p.47.

Hai, Z., Chang, K. & Kim, J., 2011. Implicit feature identification via co-occurrence association rule mining. In *Computational Linguistics and Intelligent Text Processing*. Springer, pp. 393–404.

Hamdan, H., Béchet, F. & Bellot, P., 2013. Experiments with DBpedia, WordNet and SentiWordNet as resources for sentiment analysis in micro-blogging. In *Second Joint Conference on Lexical and Computational Semantics (* SEM)*. pp. 455–459.

Hamouda, A. & Rohaim, M., 2011. Reviews classification using sentiwordnet lexicon. In *World Congress on Computer Science and Information Technology*.

Hannah, D. et al., 2007. University of Glasgow at TREC 2007: Experiments in Blog and Enterprise Tracks with Terrier. In *TREC*.

Hardisty, E.A., Boyd-Graber, J. & Resnik, P., 2010. Modeling perspective using adaptor grammars. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. pp. 284–292.

Hatzivassiloglou, V. & McKeown, K.R., 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th annual meeting of the association for computational linguistics and eighth conference of the european chapter of the association for computational linguistics*. pp. 174–181.

Hatzivassiloglou, V. & Wiebe, J.M., 2000. Effects of adjective orientation and gradability on

sentence subjectivity. In *Proceedings of the 18th conference on Computational linguistics-Volume 1*. pp. 299–305.

Hiroshi, K., Tetsuya, N. & Hideo, W., 2004. Deeper sentiment analysis using machine translation technology. In *Proceedings of the 20th international conference on Computational Linguistics*. p. 494.

Hogenboom, A. et al., 2013. Exploiting emoticons in sentiment analysis. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing*. pp. 703–710.

Hu, M. et al., 2010. Opinion Extraction, Summarization and Tracking in News and Blog Corpora. In *AAAI spring symposium: Computational approaches to analyzing weblogs*. Citeseer, pp. 261–377.

Hu, M. & Liu, B., 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. pp. 168–177.

Hung, C. & Lin, H.-K., 2013. Using objective words in SentiWordNet to improve word-of-mouth sentiment classification. *IEEE Intelligent Systems*, 28(2), pp.47–54.

Indurkhya, N. & Damerau, F.J., 2010. *Handbook of natural language processing* CRC Machin. Chapman & Halln, ed., CRC Press.

Jain, T.I. & Nemade, D., 2010. Recognizing contextual polarity in phrase-level sentiment analysis. *International Journal of Computer Applications*, 7(5), pp.12–21.

Jakob, N. & Gurevych, I., 2010. Extracting opinion targets in a single-and cross-domain setting with conditional random fields. In *Proceedings of the 2010 conference on empirical methods in natural language processing*. pp. 1035–1045.

Jia, L., Yu, C. & Meng, W., 2009. The effect of negation on sentiment analysis and retrieval effectiveness. In *Proceedings of the 18th ACM conference on Information and knowledge management*. pp. 1827–1830.

Jiménez Zafra, S.M. et al., 2015. Tratamiento de la Negación en el Análisis de Opiniones en Espanol. *Procesamiento del Lenguaje Natural, Revista n^{o} 54*, pp.37–44.

Jin, X. et al., 2007. Sensitive webpage classification for content advertising. In *Proceedings of the 1st international workshop on Data mining and audience intelligence for advertising*. pp. 28–33.

Jo, Y. & Oh, A.H., 2011. Aspect and sentiment unification model for online review analysis. In *Proceedings of the fourth ACM international conference on Web search and data mining*. pp. 815–824.

Kaur, A. & Duhan, N., 2015. A Survey on Sentiment Analysis and Opinion Mining. *International Journal of Innovations & Advancement in Computer Science IJIACS*, 4.

Kennedy, A. & Inkpen, D., 2006. Sentiment classification of movie reviews using contextual valence shifters. *Computational intelligence*, 22(2), pp.110–125.

Kessler, J.S. & Nicolov, N., 2009. Targeting Sentiment Expressions through Supervised Ranking of Linguistic Configurations. In *ICWSM*.

Khairnar, J. & Kinikar, M., 2013. Machine learning algorithms for opinion mining and sentiment classification. *International Journal of Scientific and Research Publications*, 3(6), pp.1–6.

Khoo, C.S.G., Johnkhan, S.B. & Na, J.-C., 2015. Evaluation of a General-Purpose Sentiment Lexicon on A Product Review Corpus. In *Digital Libraries: Providing Quality Information*. Springer, pp. 82–93.

Kim, H.D. & Zhai, C., 2009. Generating comparative summaries of contradictory opinions in text. In *Proceedings of the 18th ACM conference on Information and knowledge management*. pp. 385–394.

Kim, S.-M. & Hovy, E., 2004. Determining the sentiment of opinions. In *Proceedings of the 20th international conference on Computational Linguistics*. p. 1367.

Kim, S.-M. & Hovy, E., 2006a. Extracting opinions, opinion holders, and topics expressed in online news media text. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text*. pp. 1–8.

Kim, S.-M. & Hovy, E., 2006b. Identifying and analyzing judgment opinions. In *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*. pp. 200–207.

Kim, S.-M. & Hovy, E., 2005. Identifying opinion holders for question answering in opinion texts. In *Proceedings of AAAI-05 Workshop on Question Answering in Restricted Domains*. pp. 1367–1373.

Kim, Y., Kim, S. & Myaeng, S.-H., 2008. Extracting Topic-related Opinions and their Targets

in NTCIR-7. In *NTCIR*.

Kishore, P.Y.S. & Naidu, K.B., 2013. Sentiment Analysis Using Semi-Supervised Naive Bayes Classifier. *IJITR*, 1(5), pp.478–482.

Kobayashi, N. et al., 2006. Opinion mining on the web by extracting subject-aspect-evaluation relasions. *Proceedings of AAAI-CAAW*.

Konstantinova, N. et al., 2012. A review corpus annotated for negation, speculation and their scope. In *LREC*. pp. 3190–3195.

Kouloumpis, E., Wilson, T. & Moore, J.D., 2011. Twitter sentiment analysis: The good the bad and the omg! *Icwsm*, 11, pp.538–541.

Lerman, K., Blair-Goldensohn, S. & McDonald, R., 2009. Sentiment summarization: evaluating and learning user preferences. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*. pp. 514–522.

Lerman, K. & McDonald, R., 2009. Contrastive summarization: an experiment with consumer reviews. In *Proceedings of human language technologies: The 2009 annual conference of the North American chapter of the association for computational linguistics, companion volume: Short papers*. pp. 113–116.

Li, F., Han, C., et al., 2010. Structure-aware review mining and summarization. In *Proceedings of the 23rd international conference on computational linguistics*. pp. 653–661.

Li, F., Huang, M. & Zhu, X., 2010. Sentiment Analysis with Global Topics and Local Dependency. In *AAAI*. pp. 1371–1376.

Lin, C. & He, Y., 2009. Joint sentiment/topic model for sentiment analysis. In *Proceedings of the 18th ACM conference on Information and knowledge management*. pp. 375–384.

Lin, D., 1999. MINIPAR: a minimalist parser. In *Maryland linguistics colloquium*.

Lin, W.-H. et al., 2006. Which side are you on?: identifying perspectives at the document and sentence levels. In *Proceedings of the Tenth Conference on Computational Natural Language Learning*. pp. 109–116.

Liu, B., 2011. Opinion mining and sentiment analysis. In *Web Data Mining*. Springer, pp. 459–526.

Liu, B., 2010a. Opinion mining and sentiment analysis: NLP meets social sciences. *STSC, Hawaii*.

Liu, B. et al., 2005. Opinion Observer: analyzing and comparing opinions on theWeb. In *Proceedings of 14th international Conference onWorldWideWeb*. pp. 342–351.

Liu, B., 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), pp.1–167.

Liu, B., 2010b. Sentiment Analysis and Subjectivity. *Handbook of natural language processing*, 2, pp.627–666.

Liu, C.-L. et al., 2012. Movie rating and review summarization in mobile environment. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 42(3), pp.397–407.

Liu, J. & Seneff, S., 2009. Review sentiment scoring via a parse-and-paraphrase paradigm. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*. pp. 161–169.

Long, C., Zhang, J. & Zhut, X., 2010. A review selection approach for accurate feature rating estimation. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. pp. 766–774.

Lu, Y. et al., 2010. Exploiting structured ontology to organize scattered online opinions. In *Proceedings of the 23rd International Conference on Computational Linguistics*. pp. 734–742.

Lu, Y. & Zhai, C., 2008. Opinion integration through semi-supervised topic modeling. In *Proceedings of the 17th international conference on World Wide Web*. pp. 121–130.

Lu, Y., Zhai, C. & Sundaresan, N., 2009. Rated aspect summarization of short comments. In *Proceedings of the 18th international conference on World wide web*. pp. 131–140.

Maas, A.L. et al., 2011. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. pp. 142–150.

Mahendran, A. et al., 2013. Opinion mining for text classification. *International Journal of Scientific Engineering and Technology*, 2(6), pp.589–594.

Martín-Valdivia, M.-T. et al., 2013. Sentiment polarity detection in Spanish reviews

combining supervised and unsupervised approaches. *Expert Systems with Applications*, 40(10), pp.3934–3942.

Martín-Wanton, T. et al., 2010. Opinion Polarity Detection-Using Word Sense Disambiguation to Determine the Polarity of Opinions. In *ICAART (1)*. pp. 483–486.

Medlock, B. & Briscoe, T., 2007. Weakly supervised learning for hedge classification in scientific literature. In *ACL*. pp. 992–999.

Mei, Q. et al., 2007. Topic sentiment mixture: modeling facets and opinions in weblogs. In *Proceedings of the 16th international conference on World Wide Web*. pp. 171–180.

Mejova, Y., 2009. Sentiment analysis: an overview. *Comprehensive exam paper, available on http://www. cs. uiowa. edu/~ ymejova/publications/CompsYelenaMejova. pdf [2010-02-03]*.

Milidiú, R.L., 2006. *Aprendizado de Máquina para o Problema de Sentiment Classification*. PUC-Rio.

Miller, G.A. et al., 1990. Introduction to wordnet: An on-line lexical database*. *International journal of lexicography*, 3(4), pp.235–244.

Miller, G.A., 1995. WordNet: a lexical database for English. *Communications of the ACM*, 38(11), pp.39–41.

Miranda, C.N.H., Luna, J.A.G. & Salcedo, D., 2016. Minería de Opiniones basado en la adaptación al español de ANEW sobre opiniones acerca de hoteles. *Procesamiento del Lenguaje Natural*, 56, pp.25–32.

Mishne, G. & others, 2006. Multiple ranking strategies for opinion retrieval in blogs. In *Online Proceedings of TREC*.

Mishra, N. & Jha, C.K., 2012. Classification of Opinion Mining Techniques. *International Journal of Computer Applications*, 56(13).

Moghaddam, S. & Ester, M., 2011. ILDA: interdependent LDA model for learning latent aspects and their ratings from online product reviews. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. pp. 665–674.

Moghaddam, S. & Ester, M., 2010. Opinion digger: an unsupervised opinion miner from unstructured product reviews. In *Proceedings of the 19th ACM international conference*

on Information and knowledge management. pp. 1825–1828.

Moghaddam, S. & Ester, M., 2013. The flda model for aspect-based opinion mining: Addressing the cold start problem. In *Proceedings of the 22nd international conference on World Wide Web*. pp. 909–918.

Mooney, R.J. & Bunescu, R., 2005. Mining knowledge from text using information extraction. *ACM SIGKDD explorations newsletter*, 7(1), pp.3–10.

Moraes, R., Valiati, J.F. & Neto, W.P.G., 2013. Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications*, 40(2), pp.621–633.

Mukherjee, A. & Liu, B., 2012. Aspect extraction through semi-supervised modeling. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*. pp. 339–348.

Mukherjee, S., Malu, A., et al., 2012. TwiSent: a multistage system for analyzing sentiment in twitter. In *Proceedings of the 21st ACM international conference on Information and knowledge management*. pp. 2531–2534.

Mukherjee, S. & Bhattacharyya, P., 2013. Sentiment Analysis: A Literature Survey. *arXiv preprint arXiv:1304.4520*.

Mukherjee, S., Bhattacharyya, P. & others, 2012. Sentiment Analysis in Twitter with Lightweight Discourse Analysis. In *COLING*. pp. 1847–1864.

Mukras, R., 2009. *Representation and learning schemes for sentiment analysis.* The Robert Gordon University.

Mukund, S. & Srihari, R.K., 2010. A vector space model for subjectivity classification in Urdu aided by co-training. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. pp. 860–868.

Muñiz-Cuza, C.E. & Ortega-Bueno, R., 2016. Método para la Clasificación de Polaridad basado en Aspectos de Productos. *Revista Cubana de Ciencias Inform{á}ticas*, 10(1), pp.141–151.

Murakami, A. & Raymond, R., 2010. Support or oppose?: classifying positions in online debates from reply activities and opinion expressions. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. pp. 869–875.

Na, J.-C., Khoo, C. & Wu, P.H.J., 2005. Use of negation phrases in automatic sentiment classification of product reviews. *Library Collections, Acquisitions, and Technical Services*, 29(2), pp.180–191.

Neviarouskaya, A., Prendinger, H. & Ishizuka, M., 2010. Recognition of affect, judgment, and appreciation in text. In *Proceedings of the 23rd international conference on computational linguistics*. pp. 806–814.

Neviarouskaya, A., Prendinger, H. & Ishizuka, M., 2011. SentiFul: A lexicon for sentiment analysis. *Affective Computing, IEEE Transactions on*, 2(1), pp.22–36.

NG, J.P., 2010. Processing Sentiments and Opinions in Text.

Nishikawa, H. et al., 2010a. Opinion summarization with integer linear programming formulation for sentence extraction and ordering. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. pp. 910–918.

Nishikawa, H. et al., 2010b. Optimizing informativeness and readability for sentiment summarization. In *Proceedings of the ACL 2010 Conference Short Papers*. pp. 325–330.

Oard, D. et al., 2006. *TREC-2006 at Maryland: Blog, enterprise, legal and QA tracks*,

Ohana, B. & Tierney, B., 2009. Sentiment classification of reviews using SentiWordNet. In *9th. IT & T Conference*. p. 13.

Othman, M. et al., 2014. Opinion mining and sentimental analysis approaches: a survey. *Life Science Journal*, 11(4), pp.321–326.

Ott, M., Cardie, C. & Hancock, J.T., 2013. Negative Deceptive Opinion Spam. In *HLT-NAACL*. pp. 497–501.

Padró, L. & Stanilovsky, E., 2012. Freeling 3.0: Towards wider multilinguality. In *LREC2012*.

Pang, B. & Lee, L., 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*. pp. 115–124.

Pang, B., Lee, L. & Vaithyanathan, S., 2002. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*. pp. 79–86.

Park, S., Lee, K. & Song, J., 2011. Contrasting opposing views of news articles on contentious

issues. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. pp. 340–349.

Pasqualotti, P.R., 2008. *Reconhecimento de expressões de emoções na interação mediada por computador*. Universidade do Vale do Rio do Sinos.

Paul, M.J., Zhai, C. & Girju, R., 2010. Summarizing contrastive viewpoints in opinionated text. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. pp. 66–76.

Peng, T.-C. & Shih, C.-C., 2010. An Unsupervised Snippet-based Sentiment Classification Method for Chinese Unknown Phrases without using Reference Word Pairs. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*. pp. 243–248.

Poirier, D. et al., 2011. Automating opinion analysis in film reviews: the case of statistic versus linguistic approach. In *Affective Computing and Sentiment Analysis*. Springer, pp. 125–140.

Polanyi, L. & Zaenen, A., 2006. Contextual valence shifters. In *Computing attitude and affect in text: Theory and applications*. Springer, pp. 1–10.

Pontiki, M. et al., 2014. Semeval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014)*. pp. 27–35.

Popescu, A.-M., Nguyen, B. & Etzioni, O., 2005. OPINE: Extracting product features and opinions from reviews. In *Proceedings of HLT/EMNLP on interactive demonstrations*. pp. 32–33.

Poria, S. et al., 2014. Sentic patterns: Dependency-based rules for concept-level sentiment analysis. *Knowledge-Based Systems*, 69, pp.45–63.

Qiu, G. et al., 2011. Opinion word expansion and target extraction through double propagation. *Computational linguistics*, 37(1), pp.9–27.

Qu, L., Ifrim, G. & Weikum, G., 2010. The bag-of-opinions method for review rating prediction from sparse text patterns. In *Proceedings of the 23rd International Conference on Computational Linguistics*. pp. 913–921.

Raaijmakers, S. & Kraaij, W., 2010. Classifier calibration for multi-domain sentiment

classification. In *ICWSM*.

Ranade, S. et al., 2013. Online debate summarization using topic directed sentiment analysis. In *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining*. p. 7.

Rashid, A. et al., 2013. A survey paper: areas, techniques and challenges of opinion mining. *IJCSI International Journal of Computer Science Issues*, 10(2), pp.18–31.

Raut, V.B. & Londhe, D.D., 2014. Survey on Opinion Mining and Summarization of User Reviews on Web. *IJCSIT) International Journal of Computer Science and Information Technologies*, 5(2), pp.1026–1030.

Remus, R., Quasthoff, U. & Heyer, G., 2010. SentiWS-A Publicly Available German-language Resource for Sentiment Analysis. In *LREC*.

Riloff, E., Patwardhan, S. & Wiebe, J., 2006. Feature subsumption for opinion analysis. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. pp. 440–448.

Riloff, E. & Wiebe, J., 2003. Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*. pp. 105–112.

Rothfels, J. & Tibshirani, J., 2010. Unsupervised sentiment classification of English movie reviews using automatic selection of positive and negative sentiment items. *CS224N-Final Project*.

Rushdi-Saleh, M. et al., 2011. OCA: Opinion corpus for Arabic. *Journal of the American Society for Information Science and Technology*, 62(10), pp.2045–2054.

Saif, H. et al., 2013. Evaluation datasets for Twitter sentiment analysis: a survey and a new dataset, the STS-Gold. *1st Interantional Workshop on Emotion and Sentiment in Social and Expressive Media: Approaches and Perspectives from AI (ESSEM 2013)*.

Saif, H. et al., 2014. Senticircles for contextual and conceptual semantic sentiment analysis of twitter. In *The Semantic Web: Trends and Challenges*. Springer, pp. 83–98.

Saif, H., He, Y. & Alani, H., 2012. Alleviating data sparsity for twitter sentiment analysis. In *: 2nd Workshop on Making Sense of Microposts (#MSM2012): Big things come in small packages at the 21st International Conference on theWorld Wide Web (WWW'12)*.

Saleh, M.R. et al., 2011. Experiments with SVM to classify opinions in different domains. *Expert Systems with Applications*, 38(12), pp.14799–14804.

Samsudin, N. et al., 2013a. Immune based feature selection for opinion mining. In *Proceedings of the World Congress on Engineering*. pp. 3–5.

Samsudin, N. et al., 2013b. Mining Opinion in Online Messages. *IJACSA) International Journal of Advanced Computer Science and Applications*, 4(8).

dos Santos, C.N. & Gatti, M., 2014. Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts. In *COLING*. pp. 69–78.

Sarawagi, S., 2008. Information extraction. *Foundations and trends in databases*, 1(3), pp.261–377.

Sauper, C., Haghighi, A. & Barzilay, R., 2011. Content models with attitude. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. pp. 350–358.

Scaffidi, C. et al., 2007. Red Opal: product-feature scoring from reviews. In *Proceedings of the 8th ACM conference on Electronic commerce*. pp. 182–191.

Schapire, R.E. & Singer, Y., 2000. BoosTexter: A boosting-based system for text categorization. *Machine learning*, 39(2), pp.135–168.

Schrauwen, S., 2010. Machine learning approaches to sentiment analysis using the Dutch Netlog Corpus. *Computational Linguistics and Psycholinguistics Research Center*.

Seerat, B. & Azam, F., 2012. Opinion mining: Issues and challenges (a survey). *International Journal of Computer Applications*, 49(9).

Seki, K. & Uehara, K., 2009. Adaptive subjective triggers for opinionated document retrieval. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining*. pp. 25–33.

Seki, Y., 2008. A Multilingual Polarity Classification Method using Multi-label Classification Technique Based on Corpus Analysis. In *NTCIR*.

Seki, Y. et al., 2006. Opinion-focused summarization and its analysis at DUC 2006. In *Proceedings of the Document Understanding Conference (DUC)*. pp. 122–130.

Selvam, B. & Abirami, S., 2009. A Survey on Opinion Mining Framework. *International*

Journal of Advanced Research in Computer and Communication Engineering, 2(9), pp.3544–3549.

Sharma, N.R. & Chitre, V.D., 2014. Opinion mining, analysis and its challenges. *International Journal of Innovations & Advancement in Computer Science*, 3(1), pp.59–65.

Shelke, N.M., Deshpande, S. & Thakre, V., 2012. Survey of techniques for opinion mining. *International Journal of Computer Applications*, 57(13).

Sindhu, C. & Ch, S., 2013. A Survey on Opinion Mining and Sentiment Polarity Classification. *International Journal of Emerging Technology and Advanced Engineering*.

Singh, V.K., Piryani, R., Uddin, A. & Waila, P., 2013. Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification. In *Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), 2013 International Multi-Conference on*. pp. 712–717.

Singh, V.K., Piryani, R., Uddin, A., Waila, P., et al., 2013. Sentiment analysis of textual reviews; Evaluating machine learning, unsupervised and SentiWordNet approaches. In *Knowledge and Smart Technology (KST), 2013 5th International Conference on*. pp. 122–127.

Singh, V.K., Adhikari, R. & Mahata, D., 2010. A clustering and opinion mining approach to socio-political analysis of the blogosphere. In *Computational Intelligence and Computing Research (ICCIC), 2010 IEEE International Conference on*. pp. 1–4.

Smeureanu, I., Bucur, C. & others, 2012. Applying Supervised Opinion Mining Techniques on Online User Reviews. *Informatica Economica*, 16(2), pp.81–91.

Sobkowicz, P., Kaschesky, M. & Bouchard, G., 2012. Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web. *Government Information Quarterly*, 29(4), pp.470–479.

Socher, R. et al., 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the conference on empirical methods in natural language processing (EMNLP)*. p. 1642.

Somasundaran, S. & Wiebe, J., 2009. Recognizing stances in online debates. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International*

Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1. pp. 226–234.

Steinberger, J., 2013. *Multilingual Summarisation and Sentiment Analysis*. Citeseer.

Strapparava, C., Valitutti, A. & others, 2004. WordNet Affect: an Affective Extension of WordNet. In *LREC*. pp. 1083–1086.

Su, Q. et al., 2008. Hidden sentiment association in chinese web opinion mining. In *Proceedings of the 17th international conference on World Wide Web*. pp. 959–968.

Subasic, P. & Huettner, A., 2001. Affect analysis of text using fuzzy semantic typing. *Fuzzy Systems, IEEE Transactions on*, 9(4), pp.483–496.

Taboada, M. et al., 2011. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2), pp.267–307.

Taboada, M., Gillies, M.A. & McFetridge, P., 2006. Sentiment classification techniques for tracking literary reputation. In *LREC workshop: towards computational models of literary analysis*. pp. 36–43.

Taboada, M., Voll, K. & Brooke, J., 2008. Extracting sentiment as a function of discourse structure and topicality. *Simon Fraser Univeristy School of Computing Science Technical Report*.

Tan, S. et al., 2009. Adapting naive bayes to domain adaptation for sentiment analysis. In *Advances in Information Retrieval*. Springer, pp. 337–349.

Tanawongsuwan, P., 2010. Product review sentiment classification using parts of speech. In *Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on*. pp. 424–427.

Tata, S. & Di Eugenio, B., 2010. Generating fine-grained reviews of songs from album reviews. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. pp. 1376–1385.

Tatemura, J., 2000. Virtual reviewers for collaborative exploration of movie reviews. In *Proceedings of the 5th international conference on Intelligent user interfaces*. pp. 272–275.

Terveen, L. et al., 1997. PHOAKS: A system for sharing recommendations. *Communications of the ACM*, 40(3), pp.59–62.

Thelwall, M., Buckley, K. & Paltoglou, G., 2012. Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1), pp.163–173.

Titov, I. & McDonald, R., 2008. Modeling online reviews with multi-grain topic models. In *Proceedings of the 17th international conference on World Wide Web*. pp. 111–120.

Turney, P.D., 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*. pp. 417–424.

Valitutti, A., Strapparava, C. & Stock, O., 2004. Developing Affective Lexical Resources. *PsychNology Journal*, 2(1), pp.61–83.

Vasantharaj, S. et al., 2015. A Survey on Sentiment Analysis Applied in Opinion Mining. *Journal of Network Communications and Emerging Technologies (JNCET) www. jncet. org*, 1(1).

Vechtomova, O., 2010. Facet-based opinion retrieval from blogs. *Information processing & management*, 46(1), pp.71–88.

Veloso, A. & Meira Jr, W., 2007. Efficient on-demand Opinion Mining. In *SBBD*. pp. 332–346.

Vilares, D., Alonso, M.A. & Gómez-Rodríguez, C., 2013. Clasificación de polaridad en textos con opiniones en espanol mediante análisis sintáctico de dependencias. *Procesamiento del lenguaje natural*, 50, pp.13–20.

Wang, D. & Liu, Y., 2011. A pilot study of opinion summarization in conversations. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. pp. 331–339.

Wei, Z., Yu, S. & Meng, W., 2007. Opinion retrieval from blogs. In *CIKM'07: Proceedings of the sixteenth ACM Conference on Information and Knowledge Management*.

Weichselbraun, A., Gindl, S. & Scharl, A., 2010. A context-dependent supervised learning approach to sentiment detection in large textual databases. *Journal of Information and Data Management*, 1(3), p.329.

Wiebe, J., 2000. Learning subjective adjectives from corpora. In *AAAI/IAAI*. pp. 735–740.

Wiebe, J. et al., 2004. Learning subjective language. *Computational linguistics*, 30(3), pp.277–

308.

Wiebe, J. & Riloff, E., 2005. Creating subjective and objective sentence classifiers from unannotated texts. In *Computational Linguistics and Intelligent Text Processing*. Springer, pp. 486–497.

Wiebe, J.M., Bruce, R.F. & O'Hara, T.P., 1999. Development and use of a gold-standard data set for subjectivity classifications. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*. pp. 246–253.

Wiegand, M. et al., 2010. A survey on the role of negation in sentiment analysis. In *Proceedings of the workshop on negation and speculation in natural language processing*. pp. 60–68.

Wilson, T., Wiebe, J. & Hoffmann, P., 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*. pp. 347–354.

Wilson, T., Wiebe, J. & Hwa, R., 2004. Just how mad are you? Finding strong and weak opinion clauses. In *aaai*. pp. 761–769.

Wu, Y. et al., 2009. Phrase dependency parsing for opinion mining. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3*. pp. 1533–1541.

Xianghua, F. et al., 2013. Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon. *Knowledge-Based Systems*, 37, pp.186–195.

Xu, Z., 2010. *A sentiment analysis model integrating multiple algorithms and diverse features*. The Ohio State University.

Yang, K. et al., 2006. WIDIT in TREC 2006 Blog Track. In *TREC*.

Yang, K., Yu, N. & Zhang, H., 2007. WIDIT in TREC 2007 Blog Track: Combining Lexicon-Based Methods to Detect Opinionated Blogs. In *TREC*.

Yatani, K. et al., 2011. Analysis of adjective-noun word pair extraction methods for online review summarization. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*. p. 2771.

Yessenov, K. & Misailovic, S., 2009. Sentiment analysis of movie review comments.

Methodology, pp.1–17.

Yi, J. et al., 2003. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *Data Mining, 2003. ICDM 2003. Third IEEE International Conference on*. pp. 427–434.

Yin, Y., 2006. World Wide Web and the Formation of the Chinese and English "Internet Slang Union." *Computer-Assisted Foreign Language Education*, 1.

Yu, H. & Hatzivassiloglou, V., 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*. pp. 129–136.

Yu, J. et al., 2011. Aspect ranking: identifying important product aspects from online consumer reviews. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. pp. 1496–1505.

Yu, L.-C. et al., 2013. Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news. *Knowledge-Based Systems*, 41, pp.89–97.

Zhang, C. et al., 2009. Sentiment analysis of Chinese documents: From sentence to document level. *Journal of the American Society for Information Science and Technology*, 60(12), pp.2474–2487.

Zhang, C. et al., 2008. Sentiment classification for chinese reviews using machine learning methods based on string kernel. In *Convergence and Hybrid Information Technology, 2008. ICCIT'08. Third International Conference on*. pp. 909–914.

Zhang, L. & Liu, B., 2011. Identifying noun product features that imply opinions. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*. pp. 575–580.

Zhang, M. & Ye, X., 2008. A generation model to unify topic relevance and lexicon-based sentiment for opinion retrieval. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. pp. 411–418.

Zhang, W. & Yu, C., 2006. UIC at TREC 2006 Blog Track: a Notebook Paper. In *In Proc. of TREC 2006*.

Zhang, Z. et al., 2011. Sentiment classification of Internet restaurant reviews written in

Cantonese. *Expert Systems with Applications*, 38(6), pp.7674–7682.

Zhao, W.X. et al., 2010. Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. pp. 56–65.

Zhu, J. et al., 2009. Multi-aspect opinion polling from textual reviews. In *Proceedings of the 18th ACM conference on Information and knowledge management*. pp. 1799–1802.

Zhu, X., Kiritchenko, S. & Mohammad, S.M., 2014. Nrc-canada-2014: Recent improvements in the sentiment analysis of tweets. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. pp. 443–447.

Zhuang, L., Jing, F. & Zhu, X.-Y., 2006. Movie review mining and summarization. In *Proceedings of the 15th ACM international conference on Information and knowledge management*. pp. 43–50.

Zuva, K. & Zuva, T., 2012. Evaluation of information retrieval systems. *International Journal of Computer Science & Information Technology (IJCSIT)*, 4(3).