

# Generación y enriquecimiento automático de recursos léxicos para el análisis de sentimientos

Gerardo Real-Flores<sup>1</sup>, Betzabet García-Mendoza<sup>1</sup>, Ester Calderón-Casanova<sup>1</sup>,  
Gabriela Ramírez-de-la-Rosa<sup>2</sup>, Esaú Villatoro-Tello<sup>2</sup>

<sup>1</sup> Maestría en Diseño, Información y Comunicación (MADIC),  
División de Ciencias de la Comunicación y Diseño,  
Universidad Autónoma Metropolitana (UAM) Unidad Cuajimalpa, México

<sup>2</sup> Departamento de Tecnologías de la Información,  
Universidad Autónoma Metropolitana (UAM) Unidad Cuajimalpa, México

gereflo@aol.com, {betza07gm, estercalderon7}@gmail.com,  
{gramirez, evillatoro}@correo.cua.uam.mx

**Resumen.** La generación de contenidos por parte de los usuarios de internet en diversas plataformas, genera oportunidades y retos a los interesados en el análisis de sentimientos ya que se pueden clasificar grandes cantidades de información y conocer la opinión de los usuarios respecto a diversos temas. Una de las desventajas del análisis de sentimientos es que los dominios y por lo tanto las entidades a las que se les presenta una opinión cambian constantemente y de forma muy rápida. Por ello, en el enfoque de clasificación basada en recursos léxicos, la obtención de diccionarios adaptables a diferentes dominio, idiomas, entidades, es algo muy costoso. En este artículo proponemos un esquema de generación de recursos léxicos a la medida del dominio que se desea analizar, esto reduce el costo en tiempo y esfuerzo de la creación de un recurso léxico genérico. Adicionalmente, se presenta un método de enriquecimiento de recursos léxicos existentes usando información del dominio en análisis. Los resultados obtenidos son alentadores, donde se obtienen mejoras de hasta el doble en F-score que al usar un recurso genérico como ANEW para la clasificación de opiniones de revisiones de películas.

**Palabras clave:** Representación distribucional, generación de recurso léxico, análisis de sentimientos, clasificación de sentimientos.

## Automatically Generating and Enriching Lexical Resources for Sentiment Analysis

**Abstract.** The big amount of content generated from Internet users creates opportunities and challenges to researchers interested in the problem of sentiment analysis. All the information available regarding user's opinions on different topics represent a great opportunity for automatic sentiment classification systems. However, user's interest change very

rapidly over time, making automatic classification systems obsolete to new domains. In addition, tuning lexical resources and training data for creating new classification models is a very expensive and complicated task. In order to overcome such limitations, we propose a method for automatically generate a lexical resource from scratch, which in turn is useful in the training process of new classification models. Additionally, our proposed method allows to enrich an existing lexical resource. In general, our proposed method reduces time and effort costs associated to the creation of lexical resources. Basically, our proposed method analyses the semantic information contained in the data to be classified for building the lexical resource. We evaluate our proposal in two well known datasets for sentiment analysis, obtaining encouraging results in contrast to the use of generic lexical resources such as ANEW.

**Keywords:** Distributional terms representation, lexical resources generation, sentiment analysis, sentiment classification.

## 1. Introducción

En los últimos años la enorme cantidad de contenidos generados por los usuarios de sitios Web, *blogs*, *wikis*, redes sociales, entre otros, ha dado oportunidad a empresas, gobierno e incluso personajes públicos de conocer la opinión de los usuarios respecto a diversos temas y desde distintos enfoques. Una estrategia para aprovechar estas opiniones expuestas en dichos medios digitales es el análisis de sentimientos, una tarea particular del Procesamiento del Lenguaje Natural (PLN). El análisis de sentimientos se enfoca en analizar las opiniones, sentimientos, actitudes y emociones del lenguaje escrito [10]. El resultado de este análisis deriva en la clasificación de opiniones positiva, negativa o en algunos casos neutra.

Usualmente el análisis de sentimientos (AS) se aborda como una tarea de clasificación de textos, donde existen clases predefinidas cuyo objetivo principal es determinar a cuál de estas clases pertenece un documento nuevo. En el caso de AS, las clases comúnmente usadas son positiva y negativa. Una particularidad del análisis de sentimientos es que es una tarea de clasificación no temática, pues usualmente una opinión se genera sobre un mismo tema, producto, persona y/o entidad.

Históricamente, el AS se ha abordado desde tres diferentes enfoques: i) basados en estrategias de aprendizaje supervisado, ii) basados en recursos léxicos, e iii) híbridos, una combinación de los dos enfoques previos. Por un lado, los métodos de aprendizaje supervisado dependen de la existencia de una gran cantidad de documentos etiquetados además de la definición de una representación para estos documentos. Específicamente, para el AS, al ser una tarea de clasificación no temática, la representación utilizada para describir a estas opiniones tiene gran relevancia. Uno de los modelos de representación de documento más usado, debido a su simplicidad, es la bolsa de palabras (BoW) [11].

Por otro lado, en el segundo enfoque, aquellos basados en recursos léxicos, si bien no requieren un conjunto de documentos etiquetados y la definición de una representación particular, sí necesitan la compilación del recurso léxico que contenga un conjunto de términos o palabras asociadas a un tipo de sentimiento. La construcción de estos recursos son realizados por especialistas, por ejemplo lingüistas, lo que hace la generación de estos recursos una tarea costosa [12]. Un ejemplo de un recurso léxico de este tipo es ANEW (*Affective Norms for English Words*), el cual proporciona un conjunto de valoraciones emocionales normativas para 1030 palabras en el idioma inglés [1], cabe mencionar que también existe una adaptación de este recurso para idioma español con 1034 palabras [13].

Los recursos léxicos de este tipo son genéricos y se usan igual para diferentes dominios, sin embargo esto no garantiza que una palabra que sea positiva para un dominio, dígame revisiones de películas, lo sea también para un dominio distinto, dígame revisiones de libros. En este sentido, en este trabajo se exploran dos ideas para generar recursos léxicos adaptables al dominio en análisis: i) generar de manera automática una lista de palabras positivas y negativas a partir de un conjunto de documentos etiquetados y, ii) extender un recurso léxico inicial agregando información con carga positiva o negativa semánticamente similar a las palabras del recurso léxico inicial obtenidas de un conjunto de documentos del dominio de estudio.

El resto de este artículo está organizado como sigue: en la Sección 2 se hace un repaso de los métodos empleados en el análisis de sentimientos, luego en la Sección 3 se describe el esquema de generación automática de recursos léxicos. En la Sección 4 brevemente se mencionan dos de los enfoques de clasificación usados en el análisis de sentimientos. La configuración experimental y detalles de la evaluación realizada se muestra en la Sección 5. Finalmente, las conclusiones y trabajo a futuro se presenta en la Sección 6.

## 2. Trabajo relacionado

La principal tarea en el análisis de sentimientos es clasificar un texto dado un documento, esto significa determinar si la opinión expresada es positiva, negativa o neutral. Las técnicas usuales para clasificar sentimientos pueden dividirse en tres enfoques: basado en técnicas de aprendizaje supervisado, basado en recursos léxicos y un enfoque combinado de las dos anteriores. En este trabajo nos enfocaremos los métodos que usan recursos léxicos ya que son los que se vinculan con el enfoque de nuestra propuesta.

El enfoque basado en recursos léxicos se basa en una colección de palabras previamente recopiladas y etiquetadas en términos de sentimientos. Un ejemplo es ANEW que se encuentra disponible en inglés [1] y una adaptación al español [13]. La fortaleza de este tipo de recurso es que ha sido desarrollado y cuidadosamente evaluado por especialistas en el área de lingüística, prueba de su eficiencia se presenta en [4] donde reportan resultados de precisión del 90% con la versión adaptada al idioma español. Adicionalmente, en este trabajo se creó un recurso

léxico propio para clasificar opiniones de servicios turísticos en español, con este nuevo recurso obtienen como máximo 80 % en términos de precisión.

En este mismo sentido, existen otros esfuerzos que han planteado la generación semi o automática de recursos léxicos. Un ejemplo de esto es el trabajo realizado por [8], donde mediante de recolección de datos de Internet, un filtrado cuidadoso de los datos y etiquetado POS logra extraer un conjunto de palabras para clasificar opiniones de películas en inglés. Este recurso llamado Senti-CS plantea ser una mejora al recurso SentiWordNet [3]. Sus resultados están reportados en Exactitud y llegan a un 86.1 %. A pesar de los resultados prometedores de este conjunto de métodos su construcción requiere procesos en su mayoría costosos. Más aún, la mayoría de estos esfuerzos van enfocados a la construcción del recursos en inglés.

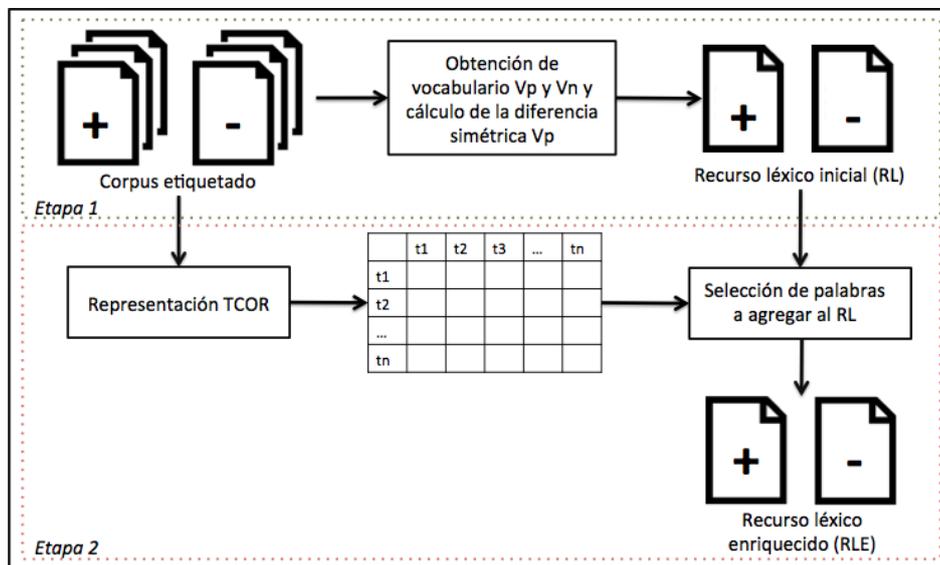
Un enfoque similar al que se propone en este artículo es descrito en [7]. Los autores proponen la construcción de un recurso léxico a partir de una colección masiva de documentos HTML de donde se extraen oraciones con polaridad (positiva o negativa) usando pistas estructurales. Al resultado lo denominan como corpus de oraciones polares, de ahí se determinan las oraciones polares por el mayor número de veces que aparece en oraciones positivas o negativas para finalmente seleccionar las oraciones con (mayor) polaridad y agregarlas al recurso léxico. Reportan resultados 92 % de exactitud sustentando la solidez de su propuesta.

En sentido opuesto a los trabajos mencionados, nuestra propuesta es la construcción de recursos léxicos a bajo costo. Usando solo información obtenida del dominio al cual pertenecen las opiniones que se desean clasificar. Más aún, nuestro esfuerzo también va enfocado al mejoramiento de recursos léxicos existentes usando información del dominio a analizar.

### **3. Método propuesto**

Como se ha establecido previamente, el objetivo de este trabajo es generar un recurso léxico particular a un dominio para el análisis de sentimientos. Este recurso léxico consiste en una lista de palabras asociada a una polaridad, i.e, positiva y negativa.

El método propuesto cuenta con dos etapas, en la primera se obtiene una lista inicial de palabras positivas y negativas, i.e, un recurso léxico inicial, el objetivo aquí es explorar la generación de un recurso léxico nuevo usando información de documentos etiquetados como positivos o negativos. En la segunda etapa se busca enriquecer un recurso léxico inicial utilizando información semánticamente relacionada al conjunto de palabras positivas y negativas iniciales. En la figura 1 se presenta la arquitectura general del método propuesto y en las siguientes secciones se explican cada una de las etapas y subetapas que lo conforman.



**Fig. 1.** Método propuesto para la generación automática de un recurso léxico a partir de un conjunto de datos etiquetados.

### 3.1. Etapa 1. Generación del recurso léxico inicial

La propuesta de este artículo es la generación automática de un recurso léxico a partir de un corpus. Para generar el recurso léxico inicial (RL) se realizaron los siguientes pasos:

1. Considerando que se cuenta con un conjunto de documentos etiquetados, se obtuvo el vocabulario de cada clase: positiva y negativa. Siendo  $V_p$  el vocabulario positivo y  $V_n$  el vocabulario negativo.
2. Se eliminan aquellas palabras que comparten ambos vocabularios, pues se considera que estas palabras no son discriminatorias de la polaridad, de modo que se conserva únicamente la diferencia simétrica de  $V_p$  y  $V_n$  ( $V_p \Delta V_n$ ). Esto se define formalmente como: dados dos conjuntos A y B, su diferencia simétrica  $A \Delta B$  es un conjunto que contiene los elementos de A y los de B excepto los que son comunes a ambos.
3. Las palabras en la diferencia simétrica será el recurso léxico inicial  $RL$ . Debido a que en esta propuesta no se consideró la obtención de orientación semántica de las palabras, como en el trabajo de Turney [14], o algún proceso que permita dar un valor numérico de que tan positiva o negativa es una palabra, se decidió asignar un peso igual a uno para las palabras positivas y un peso de menos uno para las palabras negativas.

### 3.2. Etapa 2. Enriquecimiento del léxico inicial mediante una representación distribucional

Como puede verse, la etapa inicial de la construcción del léxico sólo conserva las palabras que aparecen en los documentos de una única clase. Esta forma simple de obtener un primer RL puede ocasionar que los términos que ahí aparecen no sean suficientes para la tarea o que debido a que se eligen del conjunto previamente etiquetado, pueda generar sobre-ajuste en la etapa de la clasificación. Por ello, nuestro enfoque considera una segunda etapa donde el objetivo es desambiguar palabras que son comunes a ambas clases, pero que aún así tienen una carga hacia alguno de los polos. En otras palabras, la idea detrás de esta fase es enriquecer un RL inicial con palabras semánticamente relacionadas a las palabras positivas y negativas previamente encontradas.

Para generar un recurso léxico expandido (RLE), primero se necesita un RL inicial y un conjunto de documentos del dominio de análisis. El primer proceso, como puede verse en la Figura 1 usa el conjunto de documentos del dominio para que éstos sean representados mediante el modelo de espacio vectorial usando una representación TCOR.

La idea de esta representación distribucional es que la semántica de un término  $t_i$  puede revelarse a partir de otros términos con los que co-ocurre dentro de un documento de la colección [2,9]. De esta forma, cada término en el conjunto de términos únicos (vocabulario total) de la colección de documentos  $T$ ,  $t_i \in T$ , es representado con un vector de pesos  $t_i = \langle w_1, w_2, \dots, w_n \rangle$  donde  $w_j$  representa la contribución del término  $j$  a la descripción semántica de  $t_i$  como lo indica la siguiente fórmula:

$$w_k = \text{tff}(t_k, t_i) \cdot \log \frac{|T|}{T_k}, \quad (1)$$

donde  $T_k$  es el número de diferentes términos en el diccionario  $T$  que co-ocurren con  $t_i$  en al menos un documento y

$$\text{tff}(t_k, t_i) = \begin{cases} 1 + \log(\#(t_k, t_i)) & \text{if } (\#(t_k, t_i)) > 0, \\ 0 & \text{en otro caso,} \end{cases} \quad (2)$$

donde  $(\#(t_k, t_i))$  denota el número de documentos en los que el término  $t_i$  co-ocurre con el término  $t_k$ . El vector de pesos luego es normalizado.

Luego, en el segundo proceso de la Etapa 2 se obtienen las palabras o términos  $t_k$  con mayor similitud semántica a cada término  $t_j \in RL^{pos}$  y a  $t_l \in RL^{neg}$ . Para ello se crea una matriz de similitud de términos, dado que cada término en  $T$  está representado con un vector de pesos. De esta matriz de similitud se ordenan todos los términos de mayor a menor similitud con respecto a cada clase. Es decir, para la clase positiva se obtienen todos los  $t_k$  más similares a cada  $t_j \in RL^{pos}$ . En un paso posterior se conjuntan las listas de cada  $t_j$ , se ordenan y de esta lista general se obtiene el top-3% de términos.

Al final se obtiene un conjunto enriquecido  $RLE^{pos} = RL^{pos} \cup E^{pos}$  donde  $E^{pos}$  es el conjunto de términos que resultaron ser más similares a la clase

positiva. De forma equivalente se realiza el proceso para la clase negativa. Siendo el recurso léxico extendido representado por  $RLE = RLE^{pos} \cup RLE^{neg}$ .

## 4. Métodos de clasificación de sentimientos

El objetivo de este apartado es presentar los métodos de clasificación que se usarán en los experimentos. Los enfoques que abordaremos para evaluar los recursos léxicos generados son: clasificación basada en recursos léxicos y clasificación basada en aprendizaje supervisado.

### 4.1. Enfoque de clasificación basada en recursos léxicos

La clasificación basada en recursos léxicos parte de la premisa que existe un conjunto de términos asociados a una clase, es decir, diccionarios de pares  $\langle termino, v \rangle$  donde si  $v > 0$  el término tienen carga positiva, y si  $v < 0$  tiene carga negativa. Por lo tanto cuando se quiere clasificar un nuevo documento primero se determina la polaridad de ese documento de acuerdo con la siguiente fórmula:

$$pol = \sum_{w \in T \wedge w \in L} valor(w, L), \quad (3)$$

donde  $T$  es el conjunto de palabras en el documento a analizar,  $L$  es el conjunto de palabras en el recurso léxico usado y  $valor(w, L)$  es el valor numérico asociado a  $w$  en el recurso léxico  $L$ . Luego la clase es asignada de acuerdo a la expresión 4.

$$clase = \begin{cases} positiva & \text{if } (pol > 0), \\ negativa & \text{en otro caso.} \end{cases} \quad (4)$$

### 4.2. Enfoques de clasificación híbrida

Los métodos de clasificación híbridos combinan conocimiento previo mediante el uso de los recursos léxicos y algoritmos de aprendizaje para generar un clasificador capaz de determinar la clase de un nuevo documento.

En este sentido, una forma de aprovechar la información del recurso léxico es a través de la representación de los documentos. Es decir, se usan como atributos solo aquellos términos que aparecen en el recurso léxico ( $L$ ). De tal manera que se genera un vector por cada documento en la colección de la forma:  $d_i = \langle w_1, w_2, \dots, w_m \rangle$ , donde  $w_1$  determina la relevancia del término  $j$  en el documento  $i$ , y  $m = |L|$  para  $L$  el conjunto de términos en el recurso léxico usado.

El pesado que se utilizó para cada vector fue TF-IDF, el cual se define como:

$$w_{ki} = TF(t_k) \times IDF(t_k), \quad (5)$$

donde  $TF(t_k)$  es la frecuencia del término  $t_k$  en el documento  $d_i$ . IDF es la frecuencia inversa del término  $t_k$  dentro del documento  $d_i$  y es calculado a partir de la siguiente ecuación:

$$IDF(t_k) = \log \frac{|D|}{|\{d_i \in D : t_k \in d_i\}|}, \quad (6)$$

donde  $D$  representa la colección de documentos que está siendo indexada.

## 5. Evaluación experimental

### 5.1. Colecciones de datos

Para la evaluación del método propuesto se usaron dos colecciones de datos de un solo dominio: revisiones de películas. Con el objetivo de comprobar la independencia del idioma en el enfoque propuesto una de las colección usadas está en español mientras que la otra en inglés. En la Tabla 1 se muestran algunas estadísticas del corpora usado.

**Tabla 1.** Estadísticas de corpus empleados.

	Español	Inglés
Num. documentos	2,622	2,000
Vocabulario total	29,160	24,992
Promedio de palabras por docs.	255.6 $\pm$ 15.7	351.4 $\pm$ 15.2
Similitud promedio entre docs.	0.05 $\pm$ 0.02	0.05 $\pm$ 0.02

El corpus en español consiste en críticas de películas extraídas del sitio web *www.muchocine.net* [6], tiene originalmente un total de 3,878 críticas y aproximadamente 2 millones de palabras, con un promedio de 546 palabras por crítica. Cada crítica tiene un atributo de calificación de la película en la escala del 1 al 5. En nuestro procesamiento del corpus se conserva el mismo número de críticas pero el total de palabras descendió a 670,225 igual que el promedio de palabras por crítica que descendió a 255 palabras en promedio, como se puede ver en la Tabla 1. Otro aspecto a notar es que es un corpus relativamente balanceado, en la clase de los positivos tiene 1,349 documentos y en la negativa 1,273. Finalmente el vocabulario global consta de 29,160 palabras.

En relación al corpus en inglés [10] se caracteriza por estar totalmente balanceado: 1000 documentos por la clase negativa y 1000 de la positiva. Después del pre-procesamiento del corpus se cuenta con 702,836 palabras en total y un promedio de 351 palabras por documento. Este corpus es uno ampliamente usado para validar métodos de análisis de sentimientos.

## 5.2. Pre-procesamiento del corpora

Antes de comenzar los experimentos se realizaron algunos procesos que nos permitieron estandarizar los documentos y la organización de los mismos, los cuales se describen a continuación. Primero se organizaron los documentos de acuerdo a las dos clases para cada corpus: críticas positivas y críticas negativas. Para esta tarea fue requerido dividir el corpus en español en las dos clases mencionadas, de acuerdo a la puntuación que contenía cada crítica (valores de 1 a 5). Se decidió que los valores de 1 y 2 se tomaran como una crítica negativa, el valor de 3 como neutra, y los valores de 4 y 5 como positiva. Debido a que el corpus en inglés no contaba con críticas neutras, aquellas críticas en español que fueron evaluadas con una puntuación igual a 3, fueron excluidas de nuestro corpus.

Finalmente, el pre-procesamiento involucró las siguientes tareas:

- El texto fue convertido a minúsculas, para evitar repetición de términos, tal como el caso de *Excelente* y *excelente*.
- Se remplazaron las secuencias de espacios en blanco por un solo espacio en blanco.
- Se eliminaron signos de puntuación, tales como signos de admiración e interrogación, comas, puntos, comillas, paréntesis, etc.
- Se eliminaron los números.
- Se eliminaron palabras vacías por medio de una lista de palabras en español y una en inglés.

## 5.3. Diseño experimental

Para validar la pertinencia del recurso léxico generado automáticamente a partir de un conjunto de datos de un dominio específico se diseñaron 2 experimentos con el corpora previamente descrito. Todos los experimentos se realizaron con tres particiones distintas de 70%-30%. Se utilizaron 70% de los documentos para entrenar el modelo de clasificación o para generar el recurso léxico, el 30% restante fue usado para la evaluación. Cabe notar que cada partición contiene conjuntos de prueba disjuntos.

- Experimento 1: Clasificación de opiniones basada en recursos léxicos. Este experimento consiste en utilizar el recurso léxico (RL) generado en la etapa 1 del método propuesto en un enfoque de *clasificación basado en léxico* (ver Sección 4.1). La idea de este experimento es comparar el RL generado con un recurso léxico genérico, particularmente con ANEW. Además, se evaluará que el recurso léxico extendido RLE generado en la etapa 2 del método propuesto mejora la clasificación de cuando únicamente se utiliza el RL inicial. La idea de esta variación es determinar si al enriquecer el RL inicial usando información semántica mediante la representación TCOR, es posible mejorar la clasificación de opiniones.

**Tabla 2.** Número de palabras en los recursos léxicos utilizados, divididos por clases: positiva y negativa.

Recurso léxico	Español		Inglés	
	Positivas	Negativas	Positivas	Negativas
ANEW	511	523	581	449
ANEW-E	7052 $\pm$ 199	8747 $\pm$ 56	7140 $\pm$ 95	7799 $\pm$ 110
RL	16333 $\pm$ 437	12115 $\pm$ 117	9887 $\pm$ 88	8188 $\pm$ 125
RLE	19674 $\pm$ 559	15922 $\pm$ 291	12115 $\pm$ 135	10839 $\pm$ 152

- Experimento 2: Clasificación de opiniones basado en un enfoque híbrido. Este conjunto de experimentos consisten en utilizar un recurso léxico para representar los documentos y mediante un enfoque de clasificación supervisada, generar un clasificador que sea capaz de distinguir opiniones positiva de las negativas.

### Experimento 1: Clasificación de opiniones basada en recursos léxicos.

Como se mencionó antes, el objetivo de este conjunto de experimentos es evaluar, por un lado, que tanto mejora la clasificación de sentimientos usando un recurso léxico hecho a la medida en contraste con un recurso léxico genérico (por ejemplo, ANEW). Por otro lado, se desea evaluar en qué medida es posible mejorar la clasificación si se enriquece un recurso léxico existente, ya sea genérico o a la medida. El enriquecimiento de los dos recursos léxicos se realizó como se describe en la Sección 3 donde el porcentaje de términos agregados más similares al léxico inicial es de 3%.

En la Tabla 2 se muestran algunas estadísticas tanto del recurso ANEW para español e inglés como de los generados por el método propuesto en ambos idiomas. Como puede verse en la tabla, el número de palabras del recurso generado RL es mucho mayor que las palabras contenidas en ANEW para ambos idiomas, esto sugiere que para revisiones de películas existen muchos más términos asociados a las clases positiva y negativa que un recurso genérico con ANEW.

Cabe mencionar que al extender RL en RLE, el incremento de términos por clase es de entre el 20% y 30%, un número relativamente pequeño si comparamos el incremento de los términos de ANEW en ANEW-E (extendido), pues incrementa entre 12 y 16 veces el número de términos.

En el caso del recurso ANEW, por cada término se tiene asociado un valor de felicidad entre 0 y 10. Para este experimento se considera que un puntaje de felicidad entre 6 y 10 es considerado positivo, mientras que los términos asociados a un puntaje de felicidad de 0 a 5 son considerados negativos. Por lo tanto, en los pares  $\langle \text{termino}, \text{valor} \rangle$  para este recurso,  $\text{valor}$  es un número entero y  $0 \leq \text{valor} \leq 10$ . Para el caso de RL,  $\text{valor} = 1$  o  $\text{valor} = -1$ . La asignación de clase en estos métodos es dada como se indicó en las Ecuaciones 3 y 4 descritas en la Sección 4.1.

**Tabla 3.** Resultados de clasificar opiniones usando recursos léxicos.

Recurso léxico	<i>Español</i>			<i>Inglés</i>		
	<i>Precisión</i>	<i>Recuerdo</i>	<i>F-score</i>	<i>Precisión</i>	<i>Recuerdo</i>	<i>F-score</i>
ANEW	0.544	0.521	0.456	0.475	0.497	0.352
ANEW-E	0.629	0.583	<b>0.523</b>	0.640	0.634	<b>0.629</b>
RL	0.616	0.590	0.571	0.625	0.621	0.618
RLE	0.654	0.612	<b>0.587</b>	0.662	0.659	<b>0.658</b>

La Tabla 3 muestra los resultados de clasificación de usar cada uno de los 4 recursos. En esta tabla es posible observar que al usar un recurso léxico a medida del conjunto de documentos a clasificar, el resultado de clasificación mejora. Particularmente, para el caso del corpus en español la mejora del valor de F-score va de 0.45 a 0.57; para el caso del corpus en inglés la mejora es aún mayor, de 0.35 a 0.61 (comparando los renglones correspondientes a los recursos ANEW y RL).

Por otro lado, el recurso léxico extendido también mejora los resultados de clasificación al compararlo con el recurso léxico inicial. Es decir, si se usa ANEW como recurso léxico inicial el valor de F-score mejora de 0.45 a 0.52 y de 0.35 a 0.62, para español e inglés respectivamente. Esta mejora se conserva también cuando el recurso léxico inicial es RL, aumentado el valor del F-score de 0.57 a 0.58 y de 0.61 a 0.65, para español e inglés respectivamente.

De estos resultados podemos inferir, por un lado, que RL al ser hecho a la medida, contiene términos relevantes para este dominio en particular y que al extenderlo ya no se aporta información relevante para mejorar en mayor medida la clasificación. Por otro lado, al ser ANEW un recurso léxico genérico, el desempeño en este dominio es pobre, pero al enriquecerlo con información del dominio la clasificación mejora, esto se puede deber a que el número de términos que se agregaron al recurso tienen relación directa con el dominio en análisis.

De manera global se puede decir que utilizar un recurso léxico a medida extendido (es decir, RLE) es mejor que usar un recurso genérico. Pero que si se desea usar un recurso genérico como ANEW, es mejor utilizar su versión enriquecida (es decir, ANEW-E). Estos resultados dan pauta para suponer que el enfoque propuesto puede funcionar ya sea para otros recursos léxicos iniciales diferentes a ANEW como para otros dominios. Considerando que ANEW es un recurso léxico creado por expertos lingüistas, consideramos que nuestra propuesta es relevante para la clasificación de sentimientos, pues en base a intuiciones básicas se consiguen mejores resultados a los que obtiene el recurso léxico ANEW.

**Experimento 2: Clasificación de opiniones basado en un enfoque híbrido.** Este conjunto de experimentos tiene como objetivo evaluar, en un enfoque de clasificación de opiniones híbrido, el desempeño de la clasificación al usar como vocabulario para la representación de los documentos el conjunto de términos

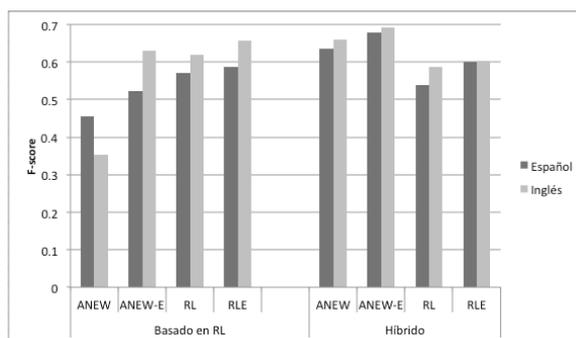
**Tabla 4.** Resultados de clasificar opiniones usando un enfoque híbrido.

Recurso léxico + SVM	Español			Inglés		
	Precisión	Recuerdo	F-score	Precisión	Recuerdo	F-score
ANEW	0.635	0.645	0.635	0.662	0.661	0.661
ANEW-E	0.679	0.679	<b>0.678</b>	0.693	0.692	<b>0.692</b>
RL	0.671	0.594	0.538	0.636	0.618	0.587
RLE	0.628	0.613	<b>0.600</b>	0.634	0.587	<b>0.604</b>

de un recurso léxico. Nuevamente, se utilizaron los recursos léxicos descritos en la Tabla 2, y como algoritmo de aprendizaje se usó SVM con los parámetros por defecto de la implementación de la herramienta Weka [5]. El resto de los detalles fueron descritos previamente en la Sección 4.2.

En la Tabla 4 se muestran los resultados de clasificación para precisión, recuerdo y F-score. De esos resultados se puede decir que también en un enfoque híbrido, los recursos léxicos enriquecidos (ANEW-E y RLE) son mejores para representar a los documentos que los recursos iniciales (ANEW, RL).

En la Figura 2 se observa un resumen de los resultados obtenidos en ambos experimentos en términos del F-score. En esta figura es posible apreciar mejor que el desempeño de nuestra propuesta funciona mejor para enfoques basados en diccionarios que en enfoques híbridos. De forma que si se cuenta con medios para generar conjuntos de datos etiquetados es mejor usar ese enfoque que uno basado en diccionarios. Sin embargo, si no se cuenta con suficiente documentos etiquetados para generar un clasificador confiable, nuestros resultados sugieren que es mejor generar un recurso léxico a la medida y luego extenderlo, esto bajo un esquema de clasificación basado en recursos léxicos.



**Fig. 2.** Valores de F-score para la clasificación de críticas de películas usando dos enfoques: clasificación basada en recursos léxicos (Basada en RL), e híbrida.

## 6. Conclusiones y trabajo a futuro

En este artículo se presentó un esquema para la generación automática de recursos léxicos con polaridad positiva y negativa para el análisis de sentimientos. La ventaja de este esquema es que se generan diccionarios a medida del dominio de los textos analizados. Por otro lado, el esquema propuesto plantea un método de enriquecimiento de un léxico dado, usando un conjunto de documentos que no necesitan estar etiquetados. Nuestro método de enriquecimiento hace uso de una representación distribucional, i.e., TCOR para encontrar términos semánticamente relacionados a aquellos en el diccionario inicial.

Los resultados obtenidos de evaluar los recursos léxicos generados muestran que es mejor hacer un recurso a medida que usar un recurso genérico. También muestra que si ya se tiene un recurso genérico, al enriquecerlo con nuestro método, los resultados pueden mejorar. Adicionalmente, al usar los recursos generados a medida en un enfoque de clasificación híbrida, es decir representar los documentos con los términos en el recurso léxico usado, no mejora el desempeño de clasificación cuando los documentos se representan con un recurso léxico genérico.

Se requieren más experimentos para concluir de manera contundente la importancia y utilidad de generar recursos léxico a bajo costo. Una de las pruebas que consideramos realizar a corto plazo es el uso de otros recursos genéricos como SentiWordNet. También se planea probar la creación del recurso léxico con un corpus de dominio diferente como revisiones de libros, hoteles u otro tipo de entidades. Como trabajo futuro adicional se contempla utilizar la orientación semántica de los términos incluidos en el recurso léxico, es decir, incluir valores que nos indiquen que tan positiva o negativa es una palabra; creemos que incorporar esta característica puede mejorar el esquema propuesto.

**Agradecimientos.** El trabajo de los primeros tres autores fue realizado con el apoyo de CONACyT (número de becas 458842, 458835 y 616127). Agradecemos también el apoyo otorgado a través de la Coordinación de la Maestría en Diseño, Información y Comunicación (MADIC) de la UAM Cuajimalpa, así como al CONACyT-SNI.

## Referencias

1. Bradley, M.M., Lang, P.J.: Affective norms for English words (ANEW): Stimuli, instruction manual, and affective ratings. Tech. rep., Center for Research in Psychophysiology, University of Florida, Gainesville, Florida (1999)
2. Cabrera, J.M., Escalante, H.J., Montes-y Gómez, M.: Distributional Term Representations for Short-Text Categorization, pp. 335–346. Springer Berlin Heidelberg, Berlin, Heidelberg (2013), [http://dx.doi.org/10.1007/978-3-642-37256-8\\_28](http://dx.doi.org/10.1007/978-3-642-37256-8_28)
3. Cruz, F.L., Troyano, J.A., Enriquez, F., Ortega, J.: Clasificación de documentos basada en la opinión: experimentos con un corpus de críticas de cine en español. *Procesamiento del Lenguaje Natural* 41(0) (2008), <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/2551>

4. Esuli, A., Sebastiani, F.: Sentiwordnet: A publicly available lexical resource for opinion mining. In: In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC'06. pp. 417–422 (2006)
5. García, A., Gaines, S., Linaza, M.T.: A lexicon based sentiment analysis retrieval system for tourism domain. *e-Review of Tourism Research (eRTR)* pp. 35–38 (May 2012)
6. Garner, S.R.: Weka: The waikato environment for knowledge analysis. In: In Proc. of the New Zealand Computer Science Research Students Conference. pp. 57–64 (1995)
7. Kaji, N., Kitsuregawa, M.: Building lexicon for sentiment analysis from massive collection of html documents. In: EMNLP-CoNLL. pp. 1075–1083 (2007)
8. Khan, F.H., Qamar, U., Bashir, S.: Senti-cs: Building a lexical resource for sentiment analysis using subjective feature selection and normalized chi-square-based feature weight generation. *Expert Systems* 33(5), 489–500 (2016), <http://dx.doi.org/10.1111/exsy.12161>, eXSY-Nov-14-253.R1
9. Lavelli, A., Sebastiani, F., Zanolini, R.: Distributional term representations: An experimental comparison. In: Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management. pp. 615–624. CIKM '04, ACM, New York, NY, USA (2004), <http://doi.acm.org/10.1145/1031171.1031284>
10. Liu, B.: Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies* 5(1), 1–167 (2012), <http://dx.doi.org/10.2200/S00416ED1V01Y201204HLT016>
11. Medhat, W., Hassan, A., Korashy, H.: Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal* 5(4), 1093 – 1113 (2014), <http://www.sciencedirect.com/science/article/pii/S2090447914000550>
12. Pang, B., Lee, L., Vaithyanathan, S.: 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. EMNLP '02, Association for Computational Linguistics, Stroudsburg, PA, USA (2002), <https://doi.org/10.3115/1118693.1118704>
13. Redondo, J., Fraga, I., Padrón, I., Comesaña, M.: The spanish adaptation of anew (affective norms for english words). *Behavior Research Methods* 39(3), 600–605 (2007), <http://dx.doi.org/10.3758/BF03193031>
14. Turney, P.D.: 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. ACL '02, Association for Computational Linguistics, Stroudsburg, PA, USA (2002), <http://dx.doi.org/10.3115/1073083.1073153>