

# Anotación automática de datos acerca de la reputación de los políticos en redes sociales

Rocío Abascal-Mena<sup>1</sup>, Jean-Valère Cossu<sup>2</sup>, Alejandro Molina<sup>3</sup>, Juan-Manuel Torres-Moreno<sup>2</sup>

<sup>1</sup> Universidad Autónoma Metropolitana, Unidad Cuajimalpa, México, D.F.

<sup>2</sup> Université d'Avignon et des Pays de Vaucluse, Laboratoire Informatique d'Avignon, Avignon, Francia

<sup>3</sup> Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO), México, D.F.

rocio.abascal@gmail.com, alejandro.molina@etd.univ-avignon.fr, {jean-valere.cossu, juan-manuel.torres}@univ-avignon.fr

**Resumen.** Recientemente, la minería de opinión en Twitter, por medio de la Recuperación de Información (RI) y del Procesamiento de Lenguaje Natural (PLN) atrajo interés en la investigación política. Sin embargo, la obtención de datos anotados de dominio específico sigue siendo un paso manual muy costoso. Además, la cantidad y calidad de estas anotaciones pueden ser críticas en relación al rendimiento de los métodos de PLN basados en técnicas de Aprendizaje de Máquina (AM). Una solución alternativa es utilizar conjuntos de datos cruzados por idioma (cross-language) y cruzados por dominio (cross-domain) para simular los datos de entrenamiento. Este artículo describe algunos enfoques de AM para la anotación automática de tuits en español que tratan acerca de la reputación de los políticos. Nuestra principal conclusión es que un clasificador estadístico sin entrenamiento específico del dominio político puede simular una anotación tan confiable como la provista por anotadores humanos. Este tipo de clasificadores, también puede resultar equivalente al uso de recursos específicos tales como lexicones de polaridad o datos de dominio traducidos manualmente.

**Palabras clave:** Procesamiento de lenguaje natural, aprendizaje de máquina, minería de opinión, análisis político automático.

## 1. Introducción

Los medios de comunicación modernos están cambiando nuestra visión sobre la sociedad en muchos aspectos. Las Ciencias Humanas, Sociales y Políticas deben evolucionar para tener todas las herramientas metodológicas con el fin de comprender las tendencias sociales y políticas tan rápido como lo requiere la sociedad moderna. Particularmente, Twitter<sup>1</sup> se ha utilizado no solo para hacer públicas las opiniones sobre

<sup>1</sup> <http://www.twitter.com>

diferentes eventos o personas, sino también para implicarse y participar en movimientos sociales.

Así, el uso de las redes sociales durante la campaña presidencial de 2012 en México adquirió mucha importancia y se convirtió el instrumento principal para ejercer la opinión pública, especialmente entre la gente joven. El movimiento juvenil *yosoy132* nació durante esa campaña, por medio las redes sociales, permitiendo a jóvenes de diversas las universidades y condición social, compartir un tema tendencia (trending topic) común. En efecto, los estudios conducidos por el *Instituto Nacional de Estadística, Geografía e Informática* (INEGI) afirman que el 40.3 % de los usuarios de Tecnologías de la Comunicación (TIC) en México son jóvenes que se comunican vía las redes sociales y los teléfonos móviles, y que permanecen conectados la mayor parte del día. Este porcentaje equivale a 15.3 millones de personas entre los 18 y 34 años de edad que son votantes potenciales [30]. La participación juvenil usando Twitter en México se ha incrementado creando significativas comunidades sociales y políticas alrededor de temas electorales. Por otra parte, los jóvenes mexicanos no son los únicos que envían tuits de manera masiva. Twitter, recientemente, ha captado una gran atención por parte de los principales candidatos que promueven su discurso en línea [27].

Como consecuencia, Twitter ofrece la oportunidad de recolectar, en tiempo real, grandes cantidades de datos, directamente de los usuarios; por lo que luego los tuits pueden ser analizados con el fin de realizar un seguimiento de las reacciones generadas en los distintos eventos. Debido a que Twitter ofrece la posibilidad de extraer tuits, a partir de los cuales se puede obtener un corpus, ha habido una gran cantidad de investigación lingüística aplicada en los tuits.

El uso de datos públicos en línea, para realizar estudios de sentimientos, reduce significativamente los costos, los esfuerzos y el tiempo necesario para administrar encuestas y cuestionarios de gran escala [5]. El Análisis de Sentimientos (AS) es un área útil en el estudio de la comunicación en línea debido a que le permite a los investigadores contar con la capacidad de medir automáticamente la emoción en textos en línea [31]. Los estudios políticos en español desde un punto de vista del Aprendizaje Automático o el Aprendizaje de Máquinas (AM) son todavía escasos [36]. Sin embargo, esto podría cambiar utilizando métodos de AM para simular las anotaciones humanas y ayudar a los expertos (en trabajos como [28]) a etiquetar una gran colección de datos.

Estudios usuales en el dominio asumen que se requiere un gran esfuerzo para la adquisición de tuits y su posterior etiquetado manual. Además, se necesita un proceso de validación para corregir los errores introducidos en el etiquetado manual. También es importante considerar que los acontecimientos políticos siempre ocurrirán mucho más rápido que nuestra capacidad para obtener los datos anotados manualmente en varios idiomas. En este contexto, proponemos un enfoque que puede proporcionar una pre-anotación confiable a partir de datos fuera del dominio que necesitan supervisión superficial antes de la validación con el fin de obtener un corpus fiable que se pueda usar para estudios políticos más complejos como la detección de la tendencia política de un usuario o el monitoreo de la reputación de un político.

El resto del artículo está organizado de la siguiente manera: la Sección 2 da una visión general de los trabajos relacionados y establece, además, la motivación de nuestro trabajo. En la Sección 3, ofrecemos detalles acerca de los conjuntos de datos utilizados.

En la Sección 4 proponemos nuestros enfoques mientras que la Sección 5 está dedicada a la evaluación. Finalmente, la Sección 6 presenta algunas conclusiones sobre nuestro trabajo y abre ciertas perspectivas de trabajo a futuro.

## **2. Estado del arte**

### **2.1. Minería de tuits y análisis de sentimientos en contextos políticos**

El análisis de datos políticos ha sido abordado en trabajos anteriores pero generalmente en inglés (véase [20] y [37]). En [24], se utiliza un léxico subjetivo proveniente del Opinion Finder con el fin de determinar puntuaciones positivas y negativas para cada uno de los conjuntos de datos correspondientes a un tuit. En este caso, los números crudos de tuits positivos y negativos acerca de un tema dado son usados para calcular una puntuación de confianza (la relación entre el número de tuits positivos y negativos). Los autores indican que con una simple inspección manual de los tuits se pueden encontrar ejemplos que han sido clasificados incorrectamente. Sin embargo, los autores utilizan este método para medir la “confianza de los consumidores” (la aprobación presidencial en este caso) durante las elecciones presidenciales de 2008 en Estados Unidos.

Un enfoque diferente se empleó en [5] para analizar las preferencias políticas mediante el estudio del humor contenido en los tuits. Un instrumento psicométrico llamado Perfil de Estados de Ánimo (Profile of Mood States, POMS) se utilizó para destilar seis diferentes atributos emocionales: tensión, depresión, cólera, vigor, fatiga y confusión. Originalmente, POMS proporciona una lista de adjetivos para la cual un paciente tiene que indicar el nivel de aprobación. Cada adjetivo se relaciona con un estado de ánimo y, por lo tanto, la lista puede ser explotada como base para un “anализador de ánimo” a partir de datos textuales.

En [34] se presenta un trabajo en dos partes: en la primera se introduce el concepto de LIWC (Linguistic Inquiry and Word Count) que se utiliza para realizar un análisis superficial de los tuits relacionados con los diferentes partidos políticos que compitieron en la elección federal de Alemania en 2009. En la segunda parte, los autores afirman que el conteo de tuits con referencias a una de los partidos, refleja con precisión los resultados de las elecciones. También, se establece que el Error Absoluto Medio (Mean Absolute Error, MAE) de la “predicción” basada en los datos de Twitter estuvo muy cerca de las encuestas reales que se llevaron a cabo .

No obstante, también los métodos puramente estadísticos han tenido impacto en el tema. Un número creciente de análisis empíricos de sentimientos y humor basados en las colecciones de Twitter han sido utilizados junto con sofisticados algoritmos de pre-procesamiento de texto, usando clasificadores basados en léxico, Máquinas de Soporte Vectorial (SVM) y métodos Bayesianos. La idea principal, de estos métodos, es la de entrenar un clasificador con las palabras clave de los tuits para determinar el estado de ánimo (ver [39,21]). También, estos métodos se han propuesto para la explotación de los tuits con el fin de detectar los cambios del estado de ánimo de la gente durante todo el día [21,19]. En [7], los autores midieron los cambios en el estado de ánimo de la población de Estados Unidos, durante tres años, a partir de tuits que proporcionaban indicadores relevantes acerca de la política.

Estudios más cualitativos proponen nuevas revelaciones sobre el comportamiento humano y como resultado se muestra que hay una enorme necesidad en el desarrollo de herramientas de minería de opinión acerca los medios sociales [22,8,11,14,18]. Sin embargo, la gran mayoría de estos trabajos utilizan corpus anotados en inglés para la experimentación, y hasta donde sabemos, no hay estudios en español ni en francés sobre el análisis de sentimientos políticos. Pero, ¿cómo hacer frente a la falta de conjuntos de datos para entrenar? En este artículo, se explora experimentalmente el uso de datos no específicos del dominio de la política combinados con datos en otro idioma, lo que nos permite transmitir la experiencia aprendida de un clasificador de dominio en francés a otro en español.

## 2.2. Procesamiento de idiomas cruzados

La traducción de la información expresada en diferentes idiomas en una representación común es, en general, muy compleja. Sistemas de Recuperación de Información en Varios Idiomas (CLIR, por sus siglas en inglés) ayudan a recuperar documentos en diferentes idiomas al plantear una consulta en un lenguaje fuente. Luego, la consulta se hace corresponder a una representación común con el fin de recuperar los documentos pertinentes en un idioma destino. La traducción de documentos, o incluso frases, en el lenguaje de consulta requiere de enormes recursos. Por lo general: (1) textos paralelos, (2) sistemas de traducción automática y (3) diccionarios bilingües en formatos legibles por una máquina.

Recientemente, el estudio del procesamiento de varios idiomas ya ha sido abordado en el Cross Language Evaluation Forum (CLEF). En [15] se describen los experimentos y resultados del uso del conjunto de datos de CLEF 2007 en un sistema de recuperación de información de una lengua hindú al inglés. Mediante el uso de una simple traducción de una palabra a otra y el aprendizaje de una tabla de alineación de palabras, obtuvieron 73 % de rendimiento de un sistema monolingüe. Especialmente, el resultado más importante de este trabajo es el descubrimiento de considerar las 4 traducciones de palabras más probables, sin umbral en la probabilidad de la traducción, lo que dio mejores resultados que la traducción de una palabra a otra.

En [6] se presenta un sistema de apoyo para la recuperación de información en varios idiomas, llamado Mulinex, el cual recupera documentos de la Web mediante el empleo de un diccionario basado en la traducción de consultas. Mulinex soporta francés, alemán e inglés mediante el uso de grandes cantidades de datos para la traducción y diferentes algoritmos de categorización de documentos: clasificadores de n-gramas para las entrada con ruido, el algoritmo del k-vecinos más cercanos para los documentos normales y clasificadores basados en patrones para documentos muy cortos. Además de la funcionalidad del cruzamiento de idiomas, Mulinex proporciona la traducción automática de documentos y sus resúmenes. Mulinex usa un asistente de consulta que ofrece la oportunidad de realizar desambiguación de la traducción de consultas interactivas. La traducción de los términos de la consulta son nuevamente traducidos en el lenguaje de consulta inicial. Sin embargo, este enfoque tiene algunas limitaciones claras debido a la falta de uso de sinónimos en el diccionario y porque los homónimos significativos en el idioma de destino pueden dar lugar a la confusión de las traducciones.

En el trabajo [32] se traduce del tamil al inglés utilizando traducción automática estadística. Los autores describen un sistema de Recuperación de Información (RI) en el dominio de la Agricultura para los agricultores de Tamil Nadu el cual les ayuda a especificar sus necesidades de información en tamil y recuperar los documentos en inglés. El sistema está diseñado con aprendizaje dinámico por lo que cualquier palabra nueva que se encuentra en el proceso de traducción se actualiza en el diccionario bilingüe.

En [38], en lugar de utilizar representaciones de documentos existentes, con información adicional en un entorno de agrupación de múltiples vistas, los autores utilizan un enfoque alternativo para codificar la información adicional como restricciones. Los resultados muestran que con los datos reales este enfoque es eficaz en la mejora de la agrupación únicamente a través del uso de los documentos originales.

Un trabajo interesante que utiliza un modelo de mezcla en varios idiomas para la clasificación de sentimientos se presenta en [23] en el cual se utiliza la alineación para reducir el sesgo hacia la lengua inicial de partida en el aprendizaje de transferencia. El modelo propuesto puede aprender sentimientos nunca antes vistos a partir de grandes volúmenes de datos sin etiquetar, que no están cubiertos por el limitado vocabulario en la traducción automática de los datos etiquetados. Con este método se utilizan datos en paralelo sin etiquetar; independientemente de si se utilizan los datos etiquetados en el idioma de destino o no.

### **2.3. Anotaciones automáticas de tuits**

Recientemente, varias investigaciones dentro del proyecto Limosine<sup>2</sup> [1,3] llevan a considerar la anotación automática para el análisis de la reputación en línea (e-reputation) de entidades corporativas, principalmente en inglés. Pero la reputación en línea de figuras públicas también interesa a los investigadores del francés y el español. En el marco de los proyectos Imagiweb<sup>3</sup> [35] y TASS<sup>4</sup> [36], respectivamente para ambos idiomas. En dichos proyectos, se ha observado que la interpretación humana de tuits, es propensa a tener discrepancias. Sigue siendo difícil obtener una anotación fiable y verdadera ya que hechos y opiniones tienen que ser considerados independientemente de si el contenido es de opinión o no. A menudo es difícil saber todas las implicaciones que un mensaje puede tener en la reputación en línea de una entidad determinada. Por último, cabe mencionar que el contexto político hace que la tarea sea aún más difícil. En este trabajo se investiga qué tanto las técnicas de Aprendizaje de Máquinas (Machine Learning) sin datos de entrenamiento correctos pueden ejecutar un trabajo semejante al que hacen los anotadores humanos.

---

<sup>2</sup> <http://www.limosine-project.eu>

<sup>3</sup> <http://mediamining.univ-lyon2.fr/velcin/imagiweb/>

<sup>4</sup> *Taller de Análisis de Sentimientos en la SEPLN* Ver: <http://www.daedalus.es/TASS2013/corpus.php>

### 3. Acerca de los conjuntos de datos

#### 3.1. Datos de AMLO durante la última elección presidencial en México

El corpus analizado es el mismo que el utilizado por [16]. Consiste en 800 tuits que contienen el hashtag #AMLO y que fueron extraídos entre el 9 y el 11 de junio de 2012. AMLO es el acrónimo de Andrés Manuel López Obrador, quien fue candidato a las elecciones presidenciales en México. Estos tuits han sido anotados manualmente de acuerdo a la polaridad de la reputación desde el punto de vista del autor<sup>5</sup>. Los desacuerdos en la anotación se han resuelto con el uso de un anotador adicional. El conjunto de datos utilizado es pequeño debido a que la anotación de un gran volumen de tuits especializados es un proceso que requiere mucho tiempo y dinero. Además, el número de anotadores para esta tarea era muy limitado. Sin embargo, estudios semejantes [27,28] acerca de Twitter y los políticos mexicanos han sido conducidos con la misma cantidad de datos.

Las clases del conjunto de datos de AMLO están bien equilibradas con solo una ligera diferencia con los tuits negativos con respecto a las clases de la colección de datos traducidos en francés como se muestra en la Tabla 1

Tabla 1: Distribución de las clases en los datos para español de México.

|          | Distribución de la clase<br>(Español) | Distribución de la clase<br>(Francés) |
|----------|---------------------------------------|---------------------------------------|
| Negativo | 0.41                                  | 0.37                                  |
| Neutral  | 0.29                                  | 0.30                                  |
| Positivo | 0.30                                  | 0.33                                  |

#### 3.2. ImagiWeb: datos de los candidatos durante la última elección presidencial en Francia

La colección utilizada en francés proviene del proyecto ImagiWeb, incluyendo todas las evaluaciones hechas por los anotadores. Esta colección recientemente se ha puesto a disposición del público<sup>6</sup> y fue utilizada en [16,35]. Consiste en 3,184 tuits anotados manualmente<sup>7</sup> para ambos candidatos presidenciales: François Hollande y Nicolas Sarkozy. Los tuits se extrajeron entre marzo y diciembre de 2012 y se refieren a los dos candidatos principales, que coincide casi el mismo período que el conjunto de política en español y el de RepLab, que se describirá más adelante.

La Tabla 2 muestra que la tendencia principal es negativa con un número muy limitado de tuits neutros. Según [35] la razón principal es que la política en Francia desata pasiones entre las personas. Para un análisis razonable solamente consideramos 3 niveles de polaridad sobre los 6 disponibles en el conjunto de datos.

<sup>5</sup> Es decir, si el autor tiene una opinión Positiva, Negativa o Neutral acerca de AMLO.

<sup>6</sup> <http://mediamining.univ-lyon2.fr/velcin/imagiweb/dataset.html>

<sup>7</sup> La anotación fue hecha por treinta personas (con educación superior), en cuanto a la polaridad (están disponibles más detalles estadísticos sobre el proceso de anotación en [35])

Tabla 2: Distribución de clases en la colección política francesa.

| Distribución de la clase |      |
|--------------------------|------|
| Negativo                 | 0.60 |
| Neutral                  | 0.12 |
| Positivo                 | 0.28 |

### 3.3. TASS: datos de la Política Española

Una parte de la evaluación de TASS 2013 [36] abarca el análisis de los sentimientos sobre mensajes políticos. El corpus proporcionado es una selección de 2,500 tuits (2,150 están todavía disponibles en línea), extraídos de Twitter durante la campaña electoral de las elecciones generales de 2011 en España. Los tuits mencionan a cualquiera de los cuatro principales partidos políticos a nivel nacional: *Partido Popular (PP)*, *Partido Socialista Obrero Español (PSOE)*, *Izquierda Unida (IU)* y *Unión, Progreso y Democracia (UPyD)*. Los tuits fueron anotados manualmente de acuerdo a la polaridad global y la polaridad a nivel de entidad (3 niveles más la etiqueta de ninguno). Esta polaridad a nivel de la entidad es similar a la anotación de polaridad para la reputación de RepLab y a la definición de polaridad en el conjunto de datos Imagiweb. Más detalles sobre el procedimiento conjunto de datos y la anotación se pueden encontrar en [36].

Tabla 3: Distribución de la clase de la colección política española de TASS 2013.

| Distribución de la clase |      |
|--------------------------|------|
| Negativo                 | 0.27 |
| Neutral                  | 0.38 |
| Positivo                 | 0.26 |
| Ninguno                  | 0.09 |

La Tabla 3 muestra que la tendencia principal es neutral con una ligera diferencia entre los valores positivos y negativos. Hemos eliminado de nuestros experimentos los tuits marcados como sin polaridad (la etiqueta ninguno).

### 3.4. RepLab: reputación de entidades

También usamos las anotaciones “polaridad de la reputación” española (23,100 tuits que representan cerca del 20% de la colección) de la colección de datos de RepLab 2013 [3]. En RepLab el objetivo de la anotación fue decidir si el contenido del tuit tenía consecuencias positivas o negativas para la reputación de la empresa cuando el contenido de las palabras contenían sentimientos explícitos o solo informaban acerca de los hechos. Las anotaciones manuales son: positivo, negativo y neutro. Como se muestra en la Tabla 4, la tendencia principal en el conjunto de RepLab es positiva.

Tabla 4: Distribución de la clase en el subconjunto español de la colección de RepLab'2013.

| Distribución de la clase |      |
|--------------------------|------|
| Negativo                 | 0.24 |
| Neutral                  | 0.28 |
| Positivo                 | 0.48 |

## 4. Protocolo Experimental

En este trabajo, nos concentramos en cómo mejorar la portabilidad de los datos anotados. Nos enfocamos en determinar las características discriminatorias, que no dependen de las entidades, del dominio o del idioma y que nos sirven para entrenar modelos estadísticos. La elección de dichas variables está motivada por muchas de las limitaciones identificadas en los enfoques de análisis de sentimientos basados en léxico. En primer lugar, debido a que requieren el desarrollo de lexicones de sentimiento y anotación específicos del lenguaje, que son caros, ya que dependen de la mano de obra humana. En segundo lugar, debido a la brevedad, el ruido y la poca edición del texto proveniente de las actualizaciones de las redes sociales donde se limita la cobertura de los lexicones y que resultan menos eficaces que en los textos estándar [12,25]. En tercer lugar, y lo más importante, porque en la opinión política, la polaridad de la reputación rara vez se codifica en palabras que devengan sentimientos; también están integradas en otras palabras y en contextos cortos. Por ejemplo, menciones de asuntos o de organizaciones financieras o escándalos están altamente correlacionados con opiniones negativas, así como las menciones de intención de voto al lado opuesto.

A continuación, se describe a través de nuestros enfoques el problema de detectar la polaridad de reputación utilizando varios métodos.

### 4.1. Enfoque léxico

Los enfoques léxicos comienzan con una lista de palabras positivas y negativas, que están ya pre-codificadas. En nuestra colección de datos primero se analizó mediante un enfoque léxico combinado con un análisis lingüístico los tuits sociales y políticos, durante un período de tiempo, con el fin de detectar los sentimientos. Comenzamos con un léxico en español y uno traducido del inglés con el fin de contar para cada tuit y para cada corpus el número de palabras positivas y negativas contenidas en cada tuit. Todo el proceso se realiza de forma automática mediante el uso de R<sup>8</sup>. Las palabras contenidas en un tuit se clasificaron en positivas o negativas a partir del uso del léxico sin tomar en cuenta el sarcasmo que transforma la polaridad de un enunciado aparentemente positivo o negativo en su opuesto [13].

### 4.2. Preprocesamiento de los datos

Ignoramos todos los tuits duplicados (elegimos considerar solo el primero de acuerdo a la fecha). Cada idioma es tratado por igual. El texto se escribe con minúsculas y

<sup>8</sup> <http://www.r-project.org/>

es limpiado a partir de la eliminación de enlaces de hipertexto, stop-words y signos de puntuación. Los hashtags no se eliminaron.

### 4.3. Aprendizaje de máquina

Se utilizó el mismo método de aprendizaje de máquina, descrito en [1] y [36] para proponer las anotaciones. El método de bootstrapping [10], ha mostrado que un pequeño conjunto anotado acoplado a la máquina de aprendizaje podría ser competitivo en comparación con los anotadores a la hora de ejecutar tareas de minería de texto. La anotación se abordó como un problema de clasificación que consistió en determinar la polaridad de cada tuit. La elección de nuestros clasificadores está motivada por su buen desempeño en muchas tareas de clasificación en investigaciones previas sobre la detección de polaridad y el análisis de los sentimientos [17,3].

Las características utilizadas por nuestras propuestas son palabras, bi-gramas y tri-gramas. Estas componen la representación de tuits codificados.

El primer modelo explorado fue frecuencia de término – frecuencia inversa de documento (TF-IDF) combinado con los criterios de pureza de Gini [33]. Este último trabajo reporta mejoras en el valor de TF-IDF sobre el uso de n-gramas ( $n \leq 3$ ). Con este modelo, estimamos la similitud de un tuit dado comparándolo con cada clase de polaridad y clasificándolo de acuerdo al índice de similitud .

La pureza de una palabra  $i$  está definida con el criterio de Gini como sigue (1):

$$gini_i = \sum_{c \in C} \mathbb{P}^2(i|c) = \sum_{c \in C} \left( \frac{DF_i(c)}{DF_T(i)} \right)^2 \quad (1)$$

donde  $C$  es un conjunto de clases,  $DF_T(i)$  es el número de tuits del conjunto entrenado  $T$  conteniendo la palabra  $i$  y  $DF_c(i)$  es el número de tuits del conjunto de datos entrenados con la clase  $c$  conteniendo la palabra  $i$ . El factor es utilizado para asignar un valor de peso a la contribución  $\omega_{i,d}$  de cada término  $i$  en el documento  $d$ , como en la ecuación (2):

$$\omega_{i,d} = TF_{i,d} \times \log\left(\frac{N}{DF_c(i)}\right) \times gini_i \quad (2)$$

Donde  $N$  es el número de tuits en conjunto entrenado y  $\omega_{i,c}$  la contribución de cada término  $i$  en la clase  $c$  (3).

$$\omega_{i,c} = DF_{i,c} \times \log\left(\frac{N}{DF_c(i)}\right) \times gini_i \quad (3)$$

**Distancia coseno.** Esta distancia se calcula para comparar las similitudes entre la bolsa de palabras de tuits y cada clase de la bolsa de palabras como se indica en la ecuación (4).

$$\cos(d, c) = \frac{\sum_{i \in d \cap c} \omega_{i,d} \times \omega_{i,c}}{\sqrt{\sum_{i \in d} \omega_{i,d}^2 \times \sum_{i \in c} \omega_{i,c}^2}} \quad (4)$$

**Máquinas de Soporte Vectorial** Los algoritmos de soporte vectorial han demostrado su capacidad para manejar grandes espacios vectoriales y determinar aquellas dimensiones que son relevantes para clasificación [17]. Nosotros elegimos entrenar Máquinas Lineales Multiclase<sup>9</sup> [9] con el objetivo de obtener la polaridad en un solo paso. La representación vectorial de la "bolsa-de-términos-peso" de cada tuit  $d$  está dada por la ecuación (5).

$$\omega_i = DF_{\mathbb{T}}(i) \times \log\left(\frac{N}{DF_{\mathbb{T}}(i)}\right) \times gini_i \quad (5)$$

**Baseline** El algoritmo de base (baseline) se calcula etiquetando cada tuit  $d_1$  de prueba con el tuit más similar  $d_2$  en el conjunto de entrenamiento (según el índice de Jaccard). Esta similitud se calcula como se indica en la ecuación 6.

$$sim(d_1, d_2) = \frac{\sum_{i \in d_1 \cap d_2} \omega_{i,d}}{\sum_{i \in d_1 \cup d_2} \omega_{i,d}} \quad (6)$$

#### 4.4. Información general del procesamiento

Las figuras 1, 2, 3 y 4 describen nuestra evaluación experimental. A continuación se explica cada evaluación.

#### 4.5. Traducción del léxico y los documentos

Como primer paso en la experimentación elegimos traducir manualmente una muestra de 220 tuits de nuestro conjunto de datos no etiquetados de México. Esto con el fin de realizar una clasificación automática utilizando el conjunto anotado francés como conjunto de entrenamiento, como se muestra en la figura 1. La muestra elegida para la traducción presenta un mejor equilibrio de clases (como se muestra en la tabla 1). El principal objetivo de este experimento es verificar la aplicabilidad de los mismos modelos a otro conjunto una prueba con otras entidades como se hizo durante RepLab 2012 [2]. Asimismo, separamos los dos candidatos del conjunto de Imagiweb en un conjunto separado de entrenamiento para evaluar si un candidato puede ser más similar a AMLO que otro.

Elegimos traducir, del inglés al español, un recurso léxico muy conocido utilizado en muchas tareas de Análisis de Sentimientos (AS) descrito en [4]. Hemos utilizado los datos de Bing Liu que se compone de alrededor de 6,800 palabras en inglés. Sin embargo, después de hacer una traducción automática utilizando el traductor de Google hemos recuperado solo 2,284 palabras positivas y 1,644 palabras negativas. En este caso, dejamos, de forma manual, solo las palabras que en español expresan un sentimiento. También se compararon los resultados de este enfoque con una clasificación utilizando un recurso léxico en español y construido específicamente para AS en Twitter y para el análisis de la política (Lexicon ElhPolar [29]). Luego, se evaluaron ambos

<sup>9</sup> Multi-Class Support Vector Machine [http://www.cs.cornell.edu/people/tj/svm\\_light/svm\\_multiclass.html](http://www.cs.cornell.edu/people/tj/svm_light/svm_multiclass.html)

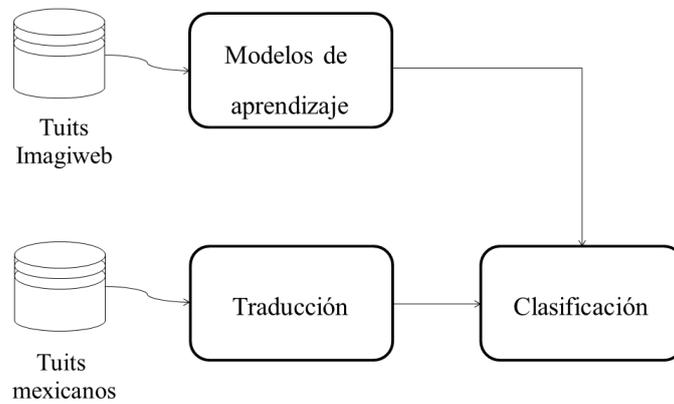


Figura 1: Proceso de clasificación usando los documentos traducidos

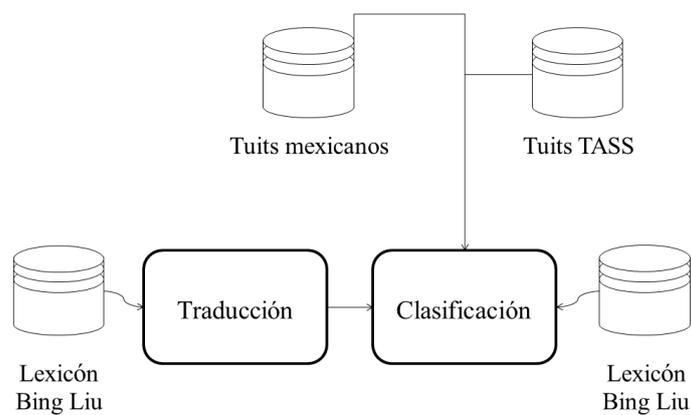


Figura 2: Clasificación usando el enfoque léxico

recursos léxicos usando el conjunto de datos TASS'2013 para comparar estos enfoques con respecto al estado del arte (véase la figura 2 y la Tabla 8).

A continuación, experimentamos en la clasificación utilizando el conjunto de reputación RepLab 2013 como conjunto de entrenamiento. Nos cuestionamos aquí acerca del

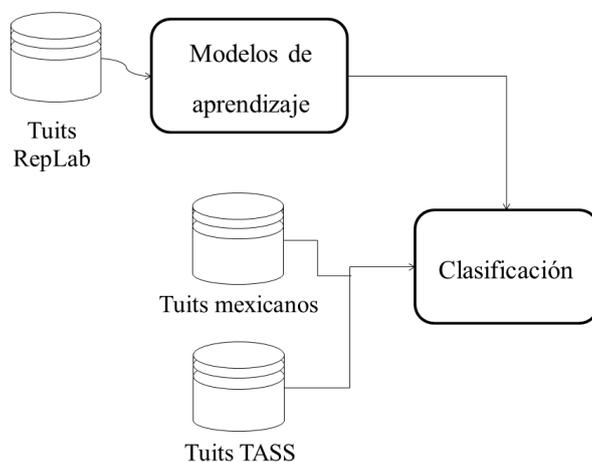


Figura 3: Proceso de clasificación usando el conjunto de datos RepLab'13 como conjunto de entrenamiento

desempeño de un mismo idioma en un conjunto masivamente etiquetado compartiendo el mismo, pequeño y ruidoso, vocabulario específico en textos provenientes de la red social. Se realizó la clasificación sobre los tuits mexicanos y el conjunto de datos políticos de TASS'2013 como se muestra en la figura 2.

En un último experimento usamos el conjunto de datos de TASS, como el mismo lenguaje, en el conjunto de entrenamiento para anotar automáticamente los tuits mexicanos (veáse figura 4).

## 5. Evaluación y resultados

### 5.1. Métricas

Para cada clasificador evaluado reportamos la precisión global y si la clase no está bien balanceada en cada conjunto de datos proponemos reportar el cálculo promedio de F-Score (Macro Averaged F-Score, anotada como F-Score en las tablas) así como la media de cada clase basándonos en la precisión y la exhaustividad con el fin de dar la misma importancia a cada clase. Por ejemplo, tomando en cuenta que la clase negativa representa el 60 % del conjunto de Imagiweb devolviendo todos los tuits como negativos daríamos una exactitud (accuracy) y un F-Score global del 60 % y una Macro F-Score Promedio F-Score del 33 %, ya que solamente detecta una clase, lo que no representa un rendimiento significativo y eficiente del sistema.

La fórmula para calcular la exactitud se presenta en la ecuación (7).

$$Exactitud = \frac{\text{Num. de documentos correctos}}{\text{Num. de documentos en referencia}} \quad (7)$$

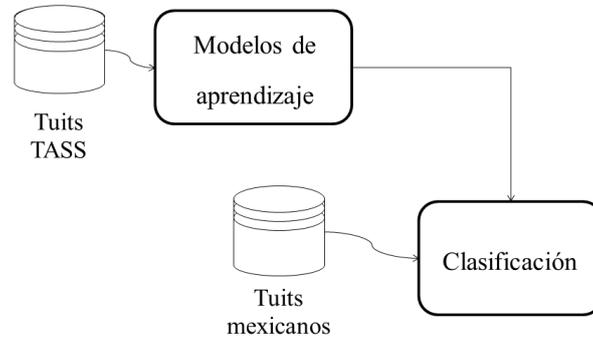


Figura 4: Proceso de clasificación usando el conjunto TASS 2013 como conjunto de entrenamiento

Y la Macro F-Score Promedio como en (8).

$$F\_Score = \frac{\sum_c \frac{2 \times (Precision_c \times Exhaustividad_c)}{Precision_c + Exhaustividad_c}}{\text{Num. de clases}} \quad (8)$$

La Precisión  $P_c$  para la clase  $c$  se determino por (9).

$$P_c = \frac{\text{Num. de documentos correctamente asignados en la clase}_c}{\text{Num. de documentos asignados en la clase}_c} \quad (9)$$

Y la Exhaustividad  $R_c$  para la clase  $c$  según (10)

$$R_c = \frac{\text{Num. de documentos correctamente asignados en la clase}_c}{\text{Num. de documentos pertenecientes a la clase}_c} \quad (10)$$

## 5.2. Aprendizaje de máquina usando datos traducidos

El conjunto Imagiweb ofrece anotación de sentimientos a nivel de persona, lo cual es más similar con respecto a nuestros tuits mexicanos proporcionando un experimento interesante. De acuerdo con la Tabla 5 el desempeño de los sistemas usando los mismos datos de contexto es realmente bajo. Las principales razones son que el vocabulario utilizado para describir a los dos principales candidatos franceses no es el mismo que el utilizado para AMLO además de que ambas distribuciones de clase parecen ser demasiado diferentes. A medida que el desempeño de los sistemas no aumentan mientras

que el tamaño del conjunto de entrenamiento sí aumenta (al considerar el conjunto Imagiweb secc. 6), se puede considerar que el rendimiento de los sistemas no está limitado a la cantidad de los datos de entrenamiento disponibles. Sin embargo, con el tamaño tan limitado de este conjunto de pruebas no somos capaces de concluir sobre la diferencia significativa entre el desempeño de los sistemas.

Tabla 5: Desempeño en la clasificación de tuits traducidos.

|          | FH sub-conjunto |               | NS sub-conjunto |               |
|----------|-----------------|---------------|-----------------|---------------|
|          | F-Score         | Exhaustividad | F-Score         | Exhaustividad |
| Baseline | 0.29            | 0.33          | <b>0.38</b>     | 0.39          |
| Coseno   | 0.17            | 0.33          | 0.34            | <b>0.42</b>   |
| MVS      | <b>0.34</b>     | <b>0.36</b>   | <b>0.38</b>     | 0.40          |

Como los resultados son ligeramente mejores cuando consideramos tuits de NS como conjunto de entrenamiento podemos considerar que los usuarios de Twitter hablan de AMLO en forma similar a la forma en que hablan de NS .

Tabla 6: Desempeño en la clasificación de tuits traducidos.

|          | Imagiweb    |               |
|----------|-------------|---------------|
|          | F-Score     | Exhaustividad |
| Baseline | 0.30        | <b>0.33</b>   |
| Coseno   | 0.26        | <b>0.38</b>   |
| MVS      | <b>0.35</b> | 0.37          |

### 5.3. Léxicos traducidos y especializados en español

Ambos enfoques léxicos, Elh-polar y Bing Liu traducido, parecen no encajar con nuestro vocabulario del conjunto de datos o este tipo de análisis no se desempeña bien como se observa en la Tabla 7.

Tabla 7: Desempeño en la clasificación usando lexicon en datos de México.

|                              | F-Score | Exhaustividad |
|------------------------------|---------|---------------|
| Lexicon ElhPolar             | 0.25    | 0.32          |
| Lexicon traducido (Bing Liu) | 0.21    | 0.33          |

El enfoque léxico muestra también limitaciones sobre el conjunto de datos de TASS'2013 como se reporta en la Tabla 8. Queríamos comparar los mejores sistemas participantes en TASS'2013, Tarea 3: Análisis de Sentimientos en Nivel Entidad pero como los

organizadores de TASS no informaron sobre el Macro Promedio F-Score no somos capaces de evaluar el "mejorrendimiento TASS en términos de F- Score y exactitud. Sin embargo, como los organizadores TASS reportan un F- Score global basado en la precisión y exhaustividad suponemos que los resultados que obtuvimos son en su mayoría similares al promedio de los participantes TASS'2013 (lo cual no se informa en este artículo).

Tabla 8: Desempeño en la clasificación usando un lexicon en el conjunto TASS'2013.

|                              | <b>F-Score</b> | <b>Exhaustividad</b> |
|------------------------------|----------------|----------------------|
| Lexicon ElhPolar             | 0.30           | 0.41                 |
| Lexicon traducido (Bing Liu) | 0.12           | 0.29                 |

#### 5.4. Aprendizaje de máquina en datos en español fuera del dominio político

En este experimento se realizó la clasificación utilizando el conjunto de RepLab 2013 como conjunto de entrenamiento. Sin embargo, RepLab también proporciona anotación a nivel persona para algunos cantantes que pueden ser similares a los de los candidatos. La Tabla 9 muestra el rendimiento de clasificación sobre contenidos en español de acuerdo con F-Score y la exactitud. MVS es inferior al método baseline cuyo rendimiento es también más alto de lo esperado. El método por distancia coseno muestra buen desempeño, utilizando los tuits traducidos. Aquí se puede concluir que es posible obtener muy buenos resultados de la clasificación, con valores de confianza cercanos a los valores de acuerdo inter-anotador observados en la literatura [3,36,26]. Se puede deducir que con este rendimiento de los clasificadores ya se puede proporcionar una buena aproximación a la anotación manual pero de manera más rápida y más económica que lo que se puede esperar con recursos humanos.

Tabla 9: Desempeño en la clasificación en tuits mexicanos.

|                              | <b>F-Score</b> | <b>Exhaustividad</b> |
|------------------------------|----------------|----------------------|
| Baseline                     | 0.50           | 0.51                 |
| Coseno                       | <b>0.74</b>    | <b>0.74</b>          |
| Lexicon ElhPolar             | 0.25           | 0.32                 |
| Lexicon traducido (Bing Liu) | 0.21           | 0.33                 |
| MVS                          | 0.17           | 0.31                 |

La anotación a nivel de entidad de RepLab ofrece una granularidad similar a la anotación de "Party/Entity" de TASS. Sin embargo, cuando consideramos TASS'2013 como conjunto de prueba se obtiene un bajo nivel de rendimiento como se muestra en la Tabla 10.

Tabla 10: Desempeño en la clasificación usando el conjunto de TASS'2013.

|          | <b>F-Score</b> | <b>Exhaustividad</b> |
|----------|----------------|----------------------|
| Baseline | 0.32           | 0.33                 |
| Coseno   | 0.32           | 0.33                 |
| MVS      | 0.33           | 0.33                 |

### 5.5. Aprendizaje de máquina usando datos de dominio

En este experimento se realizó la clasificación utilizando el conjunto de datos de TASS'2013 para entrenamiento. La Tabla 11 muestra el rendimiento de la clasificación.

Tabla 11: Desempeño en la clasificación usando los tuits mexicanos.

|          | <b>F-Score</b> | <b>Exhaustividad</b> |
|----------|----------------|----------------------|
| Baseline | 0.33           | 0.32                 |
| Coseno   | <b>0.32</b>    | <b>0.31</b>          |
| MVS      | 0.31           | 0.29                 |

El nivel más bajo de los resultados puede explicarse por el pequeño tamaño del conjunto de entrenamiento (en comparación con RepLab). Sin embargo, es bastante cercano con respecto a otros clasificadores, aunque el tamaño de la prueba es limitada. Por último, el principal hallazgo de estos experimentos es mostrar que las palabras utilizadas, este año, en este contexto son muy diferentes a las utilizadas en los dos conjuntos de México y RepLab .

### 5.6. Análisis cualitativo

Tratar con contenidos ambiguos a menudo conduce a notar errores interesantes. Algunos tuits tales como:

*“RT 1.Naces 2.Eres AMLO 3. Creces 4. No eres presidente. 5. No eres presidente. 6. No eres presidente. 7. No eres presidente. 8. Mueres. JAJA”* es etiquetado positivo por los sistemas cuando en realidad es un tuit negativo acerca del candidato AMLO.

Aquí hay otro ejemplo:

*“AMLO gran orador cada vez que abre la boca sueña #elpejeburrehastaalospejezombies”*

El cual, en este caso se trata de una ironía debido a que la gente no está soñando con un país mejor, por el contrario se está aburriendo y casi durmiendo cada vez que AMLO habla.

Los sistemas automáticos también se beneficiarían de la división del hashtag ya que no son capaces de entender las palabras unidas en una sola como *“#esunhonortuitearprobrador o #alpejenolesalencuentas”*. Sin embargo, además de las reglas lingüísticas se requerirá un procesamiento más profundo incluyendo los conocimientos de language como en el caso de *“#elpejeburrehastaalospejezombies #elpejeburrehastaasuspejezombies y #elpejeburrehastalospejezombies* donde todos son la misma declaración.

Estos son ejemplos típicos de contenidos humorísticos que los sistemas no son capaces de manejar adecuadamente. Los léxicos probablemente nunca serán capaces de considerar correctamente este tipo de mensajes. Mientras que los enfoques de AM podrían manejar estos una vez que hayan visto ejemplos similares en el conjunto de entrenamiento o en un procedimiento de aprendizaje activo.

## **6. Conclusiones**

En este artículo describimos y comparamos varios enfoques para la clasificación automática de tuits de opinión popular en política en español concernientes a la última elección presidencial en México. Este tipo de contenido es a menudo difícil de entender y la anotación es propensa a tener errores humanos. Nuestra evaluación experimental (a pesar de tener un conjunto de prueba limitado) establece que sin un entrenamiento específico del material, los enfoques de aprendizaje de máquina pueden lograr los resultados del estado del arte, mientras que en la literatura se insiste en la necesidad de datos de entrenamiento anotados manualmente.

La anotación de tuits relacionados con la política es conocida por ser una tarea difícil. Nuestros experimentos han demostrado que la necesidad de una anotación costosa por parte de un experto puede ser reconsiderada. Los enfoques de AM presentados son independientes del idioma y del dominio. Así, solo con un mínimo esfuerzo se pueden adaptar estos métodos a otro dominio, como por ejemplo, la popularidad de productos o entidades corporativas, y asimismo, se pueden manejar una gran cantidad de datos. Otra de nuestras aportaciones es el conjunto de datos anotados que pueden ser utilizados en investigaciones futuras: <https://github.com/some/url>

## **Referencias**

1. Carrillo de Albornoz, J., Amigó, E., Spina, D., Gonzalo, J.: ORMA: A semi-automatic tool for online reputation monitoring in twitter. In: *Advances in Information Retrieval - 36th European Conference on IR Research, ECIR 2014, Amsterdam, The Netherlands, April 13-16, 2014*. pp. 742–745. Springer International Publishing (2014)
2. Amigó, E., Corujo, A., Gonzalo, J., Meij, E., De Rijke, M.: Overview of replab 2012: Evaluating online reputation monitoring systems. In: *Information Access Evaluation. Multilinguality, Multimodality, and Visualization*, pp. 333–352. Springer (2012)
3. Amigó, E., De Albornoz, J.C., Chugur, I., Corujo, A., Gonzalo, J., Martín, T., Meij, E., De Rijke, M., Spina, D.: Overview of replab 2013: Evaluating online reputation monitoring systems. In: *Information Access Evaluation. Multilinguality, Multimodality, and Visualization*, pp. 333–352. Springer (2013)
4. Bing, L.: *Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies*, Morgan & Claypool Publishers (2012)
5. Bollen, J., Mao, H., P, A.: Modelling public mood and emotion: Twitter sentiment and socio-economic phenomena. In: *Fifth International AAAI Conference on Weblogs and Social Media* (2011)
6. Capstick, J., Diagne, A.K., Erbach, G., Uszkoreit, H., Leisenberg, A., Leisenberg, M.: A system for supporting cross-lingual information retrieval. *Information Processing and Management* 36(2), 275–289 (2000)

7. Cha, M., Haddadi, H., Benevenuto, F., Gummadi, K.P.: Measuring user influence in twitter: The million follower fallacy. In: ICWSM'10: Proceedings of international AAAI Conference on Weblogs and Social (2010)
8. Chung, J.E., Mustafaraj, E.: Can collective sentiment expressed on twitter predict political elections? In: Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2011, San Francisco, California, USA, August 7-11, 2011 (2011)
9. Crammer, K., Singer, Y.: On the algorithmic implementation of multiclass kernel-based vector machines. *The Journal of Machine Learning Research* 2, 265–292 (2002)
10. Di Fabrizio, G., Tur, G., Hakkani-Tur, D.: Bootstrapping spoken dialog systems with data reuse. In: Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue. pp. 72–80. Association for Computational Linguistics, Cambridge, Massachusetts, USA (2004)
11. Dodds, P., Danforth, C.: Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. *Journal of Happiness Studies* 11(4), 441–456 (2010)
12. Feczko, M., Schaye, A., Marcus, M., Nenkova, A.: Sentisummary: Sentiment summarization for user product reviews. In: proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. vol. 1, pp. 265–271 (2008)
13. González-Ibáñez, R., Muresan, S., Wacholder, N.: Identifying sarcasm in twitter: A closer look. In: The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, 19-24 June, 2011, Portland, Oregon, USA - Short Papers. pp. 581–586 (2011)
14. Gruz, A.A., Doiron, S., Mai, P.: Is happiness contagious online? a case of twitter and the 2010 winter olympics. In: HICSS. pp. 1–9 (2011)
15. Jagarlamudi, J., Kumaran, A.: Cross-lingual information retrieval system for indian languages. In: Advances in Multilingual and Multimodal Information Retrieval. pp. 80–87. Springer Berlin Heidelberg (2008)
16. Jean-Valere Cossu, R.A., Mena, A.M., Torres-Moreno, J.M., SanJuan, E.: Bilingual and cross domain politics analysis. *Avances en la Ingeniería del Lenguaje y del Conocimiento* p. 9 (2014)
17. Joachims, T.: Text categorization with support vector machines: Learning with many relevant features. Springer (1998)
18. Kramer, A.D.I.: An unobtrusive behavioral model of "gross national happiness". In: Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI 2010, Atlanta, Georgia, USA, April 10-15, 2010. pp. 287–290 (2010)
19. Lampos, V., Preoțiuc-Pietro, D., Cohn, T.: A user-centric model of voting intention from social media. In: ACL '13. p. 993–1003. Association for Computational Linguistics, Association for Computational Linguistics, Sofia, Bulgaria (08/2013 2013)
20. Malouf, R., Mullen, T.: Taking sides: User classification for informal online political discourse. *Internet Research* 18, 177–190 (2008)
21. Martínez, V., González, V.M.: Sentiment characterization of an urban environment via twitter. In: Ubiquitous Computing and Ambient Intelligence. Context-Awareness and Context-Driven Interaction - 7th International Conference, UCAmI 2013, Carrillo, Costa Rica, December 2-6, 2013, Proceedings. pp. 394–397 (2013)
22. Maynard, D., Bontcheva, K., Rout, D.: Challenges in developing opinion mining tools for social media. In: Proceedings of NLP can u tag # user generated content (2012)
23. Meng, X., Wei, F., Liu, X., Zhou, M., Xu, G., Wang, H.: Cross-lingual mixture model for sentiment classification. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1. pp. 572–581. Association for Computational Linguistics, Stroudsburg, PA, USA (2012)
24. O'Connor, B., Balasubramanyan, R., Routledge, B.R., Smith, N.A.: From tweets to polls: Linking text sentiment to public opinion time series. In: Proceedings of the Fourth Inter-

- national Conference on Weblogs and Social Media, ICWSM 2010, Washington, DC, USA, May 23-26, 2010 (2010)
25. Ohana, B., Tierney, B.: Sentiment classification of reviews using sentiwordnet. In: 9th. IT & T Conference. p. 13 (2009)
  26. Pla, F., Hurtado, L.: Political tendency identification in twitter using sentiment analysis techniques. In: COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. pp. 183–192 (2014)
  27. Sandoval, R., Matus, R.T., Rogel, R.N.: Twitter in mexican politics: Messages to people or candidates? In: 18th Americas Conference on Information Systems, AMCIS 2012, Seattle, Washington August 9-11, 2012 (2012)
  28. Sandoval-Almazán, R.: Using twitter in political campaigns: The case of the PRI candidate in mexico. IJEP 6(1), 1–15 (2015)
  29. Saralegi, X., Vicente, I.S.: Elhuyar at tass 2013. In: XXIX Congreso de la Sociedad Española de Procesamiento de lenguaje natural”. Workshop on Sentiment Analysis at SEPLN (TASS2013). pp. 143–150 (2013)
  30. Tello-Leal, E., Tello-Leal, D.A., Sosa Reyna, C.M.: Reflexiones sobre el uso de las tecnologías de información y comunicación en las campañas electorales en México: e-campañas. Revista Virtual Universidad Católica del Norte 36, 33–47 (2012)
  31. Thelwall, M., Buckley, K., Paltoglou, G.: Sentiment in twitter events. Journal of the American Society for Information Science and Technology 2(62), 406–418 (2011)
  32. Thenmozhi, D., Aravindan, C.: Cross lingual information retrieval system for agriculture society. In: International Forum for Information Technology in Tamil Conference (INFITT) (2009)
  33. Torres-Moreno, J., El-Beze, M., Bellot, P.: Bechet, opinion detection as a topic classification problem in in textual information access. chapter 9 (2013)
  34. Tumasjan, A., Sprenger, T., Sandner, P., Welpe, I.: Predicting elections with twitter: What 140 characters reveal about political sentiment. In: ICWSM. pp. 178–185 (2010)
  35. Velcin, J., Brun, C., Dormagen, J.Y., Kim, Y.M., Roux, C., Boyadjian, J., Bonnevey, S., Neihouser, M., SanJuan, E., Khouas, L., A., M., M, N.: Investigating the image of entities in social media: Dataset design and first results. In: Language Resources and Evaluation Conference (LREC) (2014)
  36. Villena-Román, J., Lana-Serrano, S., Martínez-Cámara, E., Cristóbal, J.C.G.: TASS - workshop on sentiment analysis at SEPLN. Procesamiento del Lenguaje Natural 50, 37–44 (2013)
  37. Wang, H., Can, D., Kazemzadeh, A., Bar, F., Narayanan, S.: A system for real-time twitter sentiment analysis of 2012 U.S. presidential election cycle. In: The 50th Annual Meeting of the Association for Computational Linguistics, Proceedings of the System Demonstrations, July 10, 2012, Jeju Island, Korea. pp. 115–120 (2012)
  38. Wang, X., Qian, B., Davidson, I.: Improving document clustering using automated machine translation. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management. pp. 645–653 (2012)
  39. Wijaya, V., Erwin, A., Galinium, M., Muliady, W.: Automatic mood classification of indonesian tweets using linguistic approach. In: International Conference on Information Technology and Electrical Engineering (ICITEE). pp. 41–46. IEEE (2013)