**ANALISIS DE SENTIMIENTOS**

**SentiStrength**

Un modo de superar la limitación de los métodos basados en diccionarios sería

tener en cuenta algunas reglas gramaticales y proporcionar una doble orientación del

sentimientos presentes en un texto. Estas premisas fueron incorporadas a la

herramienta

**SentiStrength** deThelwall, Buckley y Paltoglou (2012) yThelwall, Buckley, Paltoglou y Cai ( 2010).

Su sitio web es el siguiente**:** [**http://sentistrength.wlv.ac.uk/**](http://sentistrength.wlv.ac.uk/)

Inicialmente, fue diseñada para evaluar el sentimiento en los textos publicados

**en MySpace**. Sin embargo, la versatilidad de sus posibilidades y la facilidad de su manejo hacen que pueda ser aplicado a los textos cortos publicados en la web social.

SentiStrength proporciona dos valores por cada texto analizado: uno mide la

intensidad del sentimiento positivo y el otro la intensidad del sentimiento negativo.

Ambos índices varían de 1 a 5. Esta doble medida se corresponde con la presencia de

emociones diferentes combinadas en una misma expresión, lo cual viene sugerido por

estudios empíricos del campo de la psicología (Berrios, Totterdell y Kellett, 2015).

La herramienta ofrece la posibilidad de proporcionar el resultado del análisis en otros

tres formatos: binario (positivo / negativo), trinario (positivo / negativo / neutral) y en

escala simple (de -4 a +4).

SentiStrength tiene como base de trabajo una serie de archivos en formato de

diccionarios:

• **EmotionLookUpTable.txt:** Lista de palabras en la que cada una tiene asociada un

número que refleeja su intensidad de sentimiento. El valor numérico varía entre -5

y +5.

• **EmoticonLookUpTable.txt**: lo mismo que el anterior pero en lugar de palabras tiene Emoticonos.

**• IdiomLookupTable.txt**: lista de expresiones idiomáticas con su índice de

sentimiento, p.ej. a duras penas.

• **NegatingWordList.txt**: lista de palabras de negación, que servirán para invertir el

sentimiento de la palabra siguiente, p.ej. jamás, nadie.

• **BoosterWordList.txt:** lista de modificadores de la intensidad de sentimiento, p.ej.

muy, extremadamente. 3 43

• **SlangLookupTable.txt**: traducciones de expresiones coloquiales, p.ej. q equivale a

que.

Con este abanico de diccionarios SentiStrength permite una adaptación mayor al

contexto que la simple técnica basada en diccionarios. Conviene tener

presente que una de las características de un lenguaje con marcado carácter

informal se manifiesta en el uso de expresiones coloquiales y de emoticonos en los textos publicados. Por ello, resulta muy conveniente articular modos de evaluación de estas características.

Gracias a la mejora continua del programa y de los archivos asociados,

SentiStrength está siendo cada vez más utilizado en los estudios de comunicación

para analizar el sentimiento de los mensajes publicados en Twitter (Bhattacharya,

Srinivasan y Polgreen, 2014; Ferrara y Yang, 2015; Pfitzner, Garas y Schweitzer, 2012), incluido el ámbito de la comunicación política (Dang-Xuan et al., 2013; Guo y Vargo, 2015; Vargo et al., 2014).

Para la aplicación de SentiStrength a la investigación, la herramienta tiene la ventaja

de que todos los diccionarios son editables por el usuario. Esto resulta de gran interés

para la adaptación a otros idiomas. De esta forma, el usuario no requiere saber de

programación, sino que su trabajo tiene un corte más lingüístico.

Además SentiStrength ofrece la posibilidad de ajustar los pesos de los términos

para que se adapte mejor al contexto. Para ello, se requiere de un corpus codificado

manualmente que haga la función de regla de oro. El manual de la herramienta

recomienda que el volumen de este corpus sea superior a 1.000 textos y que esté

validada por tres codificadores. Este proceso resulta un tanto sofisticado pues hay que

asignar a cada texto dos índices de intensidad de sentimiento, uno para el positivo y

el otro para el negativo.

El resultado del ajuste es una tabla nueva de los términos EmotionLookUpTable en la

que el programa ha cambiado los índices de intensidad de sentimiento para que se

aproxime lo máximo posible a los valores del corpus codificado manualmente. Este

procesamiento interno está basado en técnicas de AAS.

La edición de los diccionarios y la posibilidad del ajuste de los pesos abren la posibilidad de aplicar el SentiStrength a otros idiomas. El objeto del siguiente apartado será comentar el uso de esta herramienta sobre textos escritos en castellano.

**APLICACIONES DEL SENTISTRENGTH A TEXTOS EN CASTELLANO**

El estudio de Vilares, Thelwall y Alonso (2015) tuvo un planteamiento actualizado para la puesta a punto de la herramienta. En este caso, se pretendió componer explícitamente los diccionarios necesarios para la aplicación del SentiStrength a los mensajes en castellano publicados en Twitter en el ámbito de la conversación política.

Partieron de un diccionario básico que se había nutrido fundamentalmente del

LIWC, y añadieron adjetivos, nombres, verbos y adverbios provenientes del estudio

de Brooke et al. (2009). En este punto se puso de manifiesto un inconveniente del

SentiStrength: al no realizar un procesamiento del lenguaje natural, hubo que incluir

todas las formas asociadas a los términos nuevos. O sea, incluir el masculino y el femenino junto con el singular y el plural en el caso de los nombres y los adjetivos,

y desarrollar las formas verbales de los infinitivos. El tamaño del diccionario principal

ascendió a 26.752 entradas.

Este diccionario está disponible en la siguiente dirección:

<http://sentistrength.wlv.ac.uk/SpanishSentiDataDavidVilares.zip>.

La evaluación de la herramienta con estos diccionarios fue llevada a cabo con especial

cuidado. Un corpus de 1.600 tweets fue manualmente codificado por tres lingüistas

ajenos al proyecto. La selección del equipo se basó en que hubiera unas medidas

de consistencia de intercodificación lo más altas posible. Los niveles de precisión alcanzados se midieron con un intervalo de error de +/-1. Así pues, el ajuste a los índices de sentimiento positivo fue de un 79,9% y el ajuste del sentimiento negativo fue de un 86,2% (Vilares et al., 2015).

El estudio de Vilares et al. (2015) estudio proporciona una base inicial de trabajo

A continuación se presentan algunos diccionarios de sentimientos que se pueden utilizar:

**The General Inquirer**  
  
Es un lexicón en idioma ingles que hace bastante tiempo que existe (desde 1966) y clasifica las palabras en distintas categorías. Tiene, por ejemplo, una clasificación muy simple de palabras entre “positivas” y “negativas”, pero tiene también otros grupos de categorías como por ejemplo: las dimensiones semánticas de *Osgood*, o la clasificación entre: placer, dolor, virtud y vicio. Pueden ver las distintas categorías, o mejor dicho grupos de categorías en este sitio:

<http://www.wjh.harvard.edu/~inquirer/homecat.htm>  
  
En resumen:

* Tiene 1915 palabras en la categoría: “positivas” y 2291 palabras en la categoría “negativas”
* Tiene clasificaciones complejas como por ejemplo: Fuerte vs. Débil o Activa vs. Pasiva
* Está en ingles
* Es gratis para su uso en investigación.

Enlaces:

* Página principal del sitio:   
   <http://www.wjh.harvard.edu/~inquirer/>
* Base de datos palabras:   
   <http://www.wjh.harvard.edu/~inquirer/inquirerbasic.xls>
* Categorías:   
   <http://www.wjh.harvard.edu/~inquirer/homecat.htm>

**LIWC (Linguistic Inquiry and Word Count)**

**LIWC**es otro Lexicón de Sentimientos, fue creado por James W. Pennebaker, Roger J. Booth, y Martha E. Francis. Tiene 2300 palabras y más de 70 clases. Una de las características más interesantes de este lexicón es que está disponible en español (y otros lenguajes). Entre las categorías dentro de las cuales podemos ubicar una palabra tenemos, por ejemplo, la de “Procesos Afectivos”, esta categoría nos permite clasificar las palabras en dos subcategorías: “emociones negativas” y “emociones positivas”. Otra categoría interesante es la de “Procesos Cognitivos” que nos permite clasificar palabras en las subcategorías: “tentativo” o “inhibición”, por mencionar solo algunas. Pueden ver la lista completa en este sitio:[**http://liwc.net/liwcespanol/descriptiontable1.php**](http://liwc.net/liwcespanol/descriptiontable1.php)

En resumen:

* 2300 palabras y más de 70 clases.
* Soporta idioma español.
* Tiene clasificaciones complejas.
* **No es gratuito**, tiene dos versiones con costos de US$29.95 y US$89.95.

Enlaces:

* Página principal**:** [**http://www.liwc.net/**](http://www.liwc.net/)
* Página principal en español:  [**http://liwc.net/liwcespanol/index.php**](http://liwc.net/liwcespanol/index.php)
* Categorías:  [**http://liwc.net/liwcespanol/descriptiontable1.php**](http://liwc.net/liwcespanol/descriptiontable1.php)

**MPQA Subjectivity Cues Lexicon**

**MPQA**es un lexicón creado por Theresa Wilson, Janyce Wiebe, y Paul Hoffmann en 2005. Clasifica las palabras en “positivas” y “negativas”, pero cada palabra tiene *anotada*su intensidad (fuerte vs. débil).

En resumen:

* 2718 palabras en la categoría “positivas” y 4912 en la categoría “negativas”
* Está en idioma ingles
* Se distribuye bajo licencia GNU GPL

Enlace:

* Página principal: [**http://mpqa.cs.pitt.edu/lexicons/subj\_lexicon/**](http://mpqa.cs.pitt.edu/lexicons/subj_lexicon/)