

MORENO-ORTIZ, A. *Lingmotif 1.0* [Software]. Málaga: Universidad de Málaga, 2016.

<http://tecnolengua.uma.es/lingmotif>

Dirección

Clara Martínez
Cantón
Gimena del Río
Riande
Ernesto Priani

Secretaría

Romina De León

Reseña realizada por:

Aroa ORREQUIA-BAREA

Universidad de Jaén

orrequia@ujaen.es

La enorme presencia de internet en nuestras vidas diarias y el auge de la web 2.0 – páginas que permiten a los usuarios compartir contenidos, por ejemplo, redes sociales– han hecho que el Análisis de Sentimiento, campo perteneciente al Procesamiento del Lenguaje Natural, tenga un gran desarrollo en los últimos quince años. Este apogeo se debe, en parte, a la gran cantidad de información que existe en internet. El hecho de que los consumidores busquen opiniones antes de comprar o consumir productos, y que las proporcionen una vez consumidos los mismos, ha permitido a las empresas, en general, y a los investigadores, en particular, acceder a una cantidad ingente de datos susceptibles de ser analizados. En este contexto Antonio Moreno Ortiz, profesor titular de Filología Inglesa en la Universidad de Málaga, junto con su grupo de investigación Tecnolengua (Tecnologías lingüísticas y comunicación intercultural)¹, creó Lingmotif en 2016. Esta aplicación está financiada por el Ministerio de Ciencia e Innovación español a través de los proyectos de investigación FFI2011-25893 (2012-2015) y FFI2016-78141-P (2017-2019). Este software resulta muy útil, no solo para el campo de la Lingüística, sino también para otros en los que se utiliza el lenguaje evaluativo como, por ejemplo, la publicidad, el turismo o las Humanidades Digitales.

Lingmotif es una aplicación multiplataforma que realiza automáticamente una clasificación y un análisis lingüístico de los textos en términos de Análisis de Sentimiento. La misma está basada en el léxico, o sea, que analiza los textos en función de la orientación semántica que contienen las palabras y frases del texto introducido.

¹ Accesible desde: <http://tecnolengua.uma.es/>.

Actualmente puede analizar textos en español, inglés y catalán, pero el equipo de investigación continúa trabajando para ampliar este análisis a textos en francés, alemán e italiano. La novedad de esta aplicación con respecto a otros software existentes radica en que no se trata de un simple clasificador que etiqueta los textos como negativos o positivos, sino que, además, ofrece un análisis cuantitativo y cualitativo, así como unas representaciones visuales de los resultados. Otra ventaja con respecto a dichos clasificadores es que los resultados se proporcionan en formato HTML, permitiendo que se puedan publicar de forma inmediata en páginas web.

Para analizar los textos, Lingmotif los divide en oraciones, por lo cual el análisis se lleva a cabo a dicho nivel. Uno de los problemas que se derivan de esta división es que la aplicación deja fuera del análisis muchos marcadores que indican la polaridad a nivel extra-oracional. El segundo paso que realiza es una búsqueda de las palabras y multipalabras que tienen polaridad y les asigna un valor: si es positivo de 5 a 2 y si es negativo de -5 a -2. De esta forma, los textos se clasifican en positivos, negativos y neutros (cuando el valor es 0). No importa la temática o la longitud, pues este software pretende ser un recurso léxico general que sea capaz de analizar cualquier tipo de texto, desde novelas hasta textos tan cortos como los tuits.

El funcionamiento de Lingmotif depende de dos partes fundamentales: el lexicón nuclear y las reglas de contextos. Existen, además, léxicos extendidos o diccionarios específicos que también se pueden añadir a este software, permitiendo que sea una herramienta adaptable a todo tipo de dominios.

El lexicón nuclear es la parte más importante de la aplicación y contiene unos 207.000 ítems incluyendo palabras, expresiones multipalabras y emoticonos. Las expresiones multipalabras se hacen necesarias en este análisis, pues la polaridad no siempre viene dada por una única palabra. Por ejemplo, *mejor* es una palabra con polaridad positiva, pero una expresión que contenga dicha palabra no tiene que ser necesariamente positiva, de hecho, la expresión *antes era mejor* tiene una polaridad negativa. En este sentido, Lingmotif clasifica las palabras en positivas, negativas, neutras y *none*, correspondiéndose esta última clase con las palabras funcionales o gramaticales.

Sin embargo, una de las limitaciones que tiene es que, en algunas ocasiones, las faltas de ortografía en textos, tales como opiniones de usuarios, pueden cambiar la polaridad global del texto. De esta forma, la aplicación proporciona un análisis a nivel global, dando la orientación semántica del texto como un todo, y otro atendiendo a la polaridad de cada uno de los ítems que dicho texto contiene.

Otra parte fundamental dentro de esta herramienta son las reglas de contexto. Lingmotif contiene unas cuatrocientas reglas de contexto para cada idioma, mecanismos que analizan las palabras en función de otras que están en su entorno. En el caso del término *culpa*, por ejemplo, tiene una polaridad negativa, pero si aparece con *no*, como en la expresión *no es*

culpa de, las reglas de contexto permiten invertir la polaridad de dicha palabra y, por tanto, a esta expresión multipalabra se le asigna la polaridad positiva. Asimismo, las reglas de contexto también suponen un problema para la aplicación en el ámbito de las expresiones multipalabras, pues no detecta las comparaciones, traduciéndose en numerosos fallos de polaridad de ítems y por ende de polaridad global. En el caso de contenidos generados por usuarios, si se está comparando el producto o el servicio con algo peor, el software detectará las palabras con polaridad negativa de manera aislada, clasificando así el texto como negativo. Lo mismo ocurre con las comparaciones positivas.

Por último, existen unos complementos opcionales, plugin, extensiones o diccionarios específicos. Debido a que, de manera general, el tema del texto suele determinar de manera notable su orientación semántica, estos diccionarios de léxico específico permiten refinar el análisis en textos de una temática concreta. Dentro del Análisis de Sentimiento, este analiza los textos en función de los aspectos más importantes del dominio. En el caso de textos relacionados con la temática de electrodomésticos, el plugin añadiría palabras que, en un lexicón general, como el que posee Lingmotif, no tendrían ninguna polaridad. Con respecto al análisis, con la extensión implementada, la aplicación busca primero las palabras en este diccionario y a las que consigue identificar les asigna la polaridad. Luego buscará las palabras en el lexicón nuclear, ignorando las que ya posean una polaridad. Por ejemplo, el adjetivo *silencioso* es analizado por el software como neutro, es decir, le asigna valencia 0. Sin embargo, si estamos hablando de una plancha o de una cafetera, *silencioso* es algo positivo por lo que en el *plugin* tendrá una valencia entre 2 y 5. Uno de los inconvenientes que esta aplicación plantea en relación con las palabras de dominio específico, en particular, y a todos los términos, en general, son los nombres propios. En numerosas ocasiones estos coinciden con vocablos que portan una polaridad. Esto da lugar a errores en la asignación de la orientación semántica por parte de la herramienta. Este problema ocurre con bastante frecuencia en el caso de nombres de hoteles, como, por ejemplo, Hotel Conquistador, donde la palabra *conquistador* se analiza como positiva, o con marcas de electrodomésticos, como, por ejemplo, Dolce Gusto, donde *gusto* también es analizada como positiva.

Con respecto al funcionamiento de Lingmotif, una de las funciones que resulta muy interesante es la posibilidad de introducir varios documentos para su análisis. Esta función permite realizar estudios en relación con el lenguaje evaluativo, pues se podrían comparar expresiones de diferentes hablantes o autores, así como estudiar la evolución de un emisor o autor particular. Independientemente de si se introducen uno o más documentos, esta aplicación siempre proporciona una serie de métricas para cada texto: *Text Sentiment Score* (TSS), *Text Sentiment Intensity* (TSI), perfil de sentimiento, análisis de texto, análisis cuantitativo de sentimiento y un Análisis de Sentimiento detallado.

Las medidas más importantes son el TSS y el TSI, pues van a proporcionar los datos relativos a la polaridad. Por un lado, el TSS da la polaridad de texto a nivel general; por otro, el TSI ofrece la cantidad de elementos con polaridad que hay en cada texto. Estas medidas son muy útiles, ya que permiten obtener una medida objetiva de la polaridad de un texto. Para estas dos métricas, se presenta una representación gráfica, similar a los cuentakilómetros de los coches, que resulta muy visual y clara a la hora de interpretar los resultados. Los ítems del texto que proporcionaron estas métricas se pueden observar en el análisis cuantitativo de sentimiento que ofrece un pormenor detallado donde aparecen los elementos coloreados en función de su polaridad. Mediante esta función se pueden identificar fragmentos o secciones de un texto que sean inusualmente positivas o negativas y averiguar a qué se debe.

Otra de las métricas interesantes que proporciona esta aplicación es el perfil de sentimiento. Se trata de un gráfico de líneas de diferentes colores que permite visualizar de manera precisa cómo va variando el flujo de polaridad de un pasaje. El análisis de texto también puede resultar útil para cualquier tipo de estudio lingüístico, ya que proporciona las palabras *token* y tipo, el número de oraciones, las palabras léxicas y las funcionales, etc.

De manera general, y aunque esta herramienta pueda ser usada para cualquier tipo de texto, es bastante útil en el estudio del lenguaje evaluativo y, por tanto, en el análisis de contenidos generados por usuarios en internet. Resulta de gran ayuda para analizar pasajes extraídos de redes sociales, tales como Twitter o de páginas de opiniones de usuarios, como, por ejemplo, TripAdvisor.

El principal problema que esta aplicación encuentra con este tipo de textos es la ironía. Como máquina que es, Lingmotif no es capaz de detectar si un usuario está escribiendo de forma irónica o sarcástica o, por el contrario, diciendo la verdad. En este sentido, asignará a las palabras la polaridad que les corresponda por su significado denotativo, esto es, el que está en el lexicón nuclear o, de manera particular, en el *plugin*. Por lo tanto, este tipo de opiniones serán clasificadas de manera errónea por parte de la aplicación requiriendo en estos casos una revisión manual por parte del investigador.

En conclusión, aunque Lingmotif tenga algunas limitaciones e inconvenientes que difícilmente serán subsanables, debido a que se tratan de problemas inherentes al trabajo con una máquina, resulta una aplicación bastante útil y recomendable para todos aquellos investigadores que utilicen de una forma u otra lenguaje evaluativo y Análisis de Sentimiento. Las mayores ventajas son la cantidad de métricas que ofrece, así como la posibilidad de obtener de manera rápida resultados que se pueden publicar fácilmente, debido a su formato, y que constituyen una sólida base para cualquier estudio empírico. Desde que fue distribuida en 2016, tanto la aplicación como el léxico han sido enormemente ampliados y afinados

habiéndose convertido, actualmente, en un producto comercial que puede ser adquirido a través de Corpus Level².

² Esta empresa de base tecnológica nace tras haber ganado el XXII Concurso Spin-off organizado por el Vicerrectorado de Innovación Social y Emprendimiento de la Universidad de Málaga. Accesible desde: <http://www.corpuslevel.com>.