



Análisis de sentimientos haciendo uso de *MapReduce*

Sentiment analysis using *MapReduce*

Kevin David SÁNCHEZ Solís [1](#); Juan Pablo CARABALÍ Rivera [2](#); Sonia JARAMILLO Valbuena [3](#)

Recibido: 15/05/2018 • Aprobado: 01/07/2018

Contenido

- [1. Introducción](#)
 - [2. Estado del arte](#)
 - [3. Descripción del sistema propuesto](#)
 - [4. Implementación del analizador de sentimientos](#)
 - [5. Conclusiones](#)
- [Referencias bibliográficas](#)

RESUMEN:

En la actualidad diversas aplicaciones generan grandes volúmenes de datos. Esta gran cantidad de información, conocida bajo el término de Big Data, se define en torno a 4 dimensiones: volumen, variedad, velocidad y veracidad y, contempla un conjunto de técnicas y tecnologías que permiten la captura, procesamiento, análisis y presentación de los datos. El análisis de Big Data beneficia los procesos de negocio y puede ser usado para aportar al mejoramiento de la vida diaria de los habitantes. En este contexto, el análisis de sentimientos surge como una importante alternativa para comprender la actitud que una persona tiene sobre un tema. El objetivo de este artículo es presentar un nuevo analizador de sentimientos que identifica la polaridad presente en una opinión. El sistema propuesto fue implementado haciendo uso del modelo de programación MapReduce. Para la evaluación del mismo se usan datasets del mundo real. Los resultados reflejan la eficacia del sistema, al proporcionar una buena correspondencia de los resultados con el dominio de búsqueda. La evaluación realizada se incluye al final del artículo.

Palabras clave: Análisis de sentimientos; clasificación; polaridad

ABSTRACT:

At present, several applications generate large volumes of data. This large amount of information, known as Big Data, is defined around 4 dimensions: volume, variety, speed and veracity, and includes a set of techniques and technologies that allow the capture, processing, analysis and presentation of data. The analysis of Big Data benefits the business processes and can be used to contribute to the improvement of the daily life of the inhabitants. In this context, the analysis of feelings emerges as an important alternative to understand the attitude that someone has on a topic. The objective of this paper is to present a new sentiment analyzer implemented using the MapReduce programming model. For the evaluation of it, real-world datasets are used. The results reflect the effectiveness of the system, by providing a good match of the results with the search domain. The evaluation carried out is included at the end of the article.

Keywords: Sentiment analysis, classification, polarity.

1. Introducción

Una aplicación del Big Data es el análisis de opiniones o sentimientos. Mediante esta

disciplina es posible detectar información subjetiva de un texto en que se expresa indirectamente cierta clase de emoción y clasificar la orientación expresada en dichos documentos. Un sentimiento es un estado de ánimo junto con una emoción conceptualizada, generada debido a un objeto, persona o circunstancia (Scherer & Ekman, 1984) (Scherer, 2005) (Castilla & Carlos, 2010). Existe una gran variedad de sentimientos, que se pueden catalogar de manera global, algunos positivos (alegría, felicidad, amor, apego, cariño, entre otros) y otros negativos (cólera, angustia, asco, desprecio, humillación). La Tabla 1 presenta un listado de las 8 emociones básicas presentadas en (Plutchik, 2002) y la Tabla 2 muestra su relación directa con cada sentimiento (Jaramillo, Cardona, & Espitia, 2017).

Tabla 1
Emociones básicas y sus opuestos (Plutchik, 1980)

Emoción Básica	Emoción opuesta
Alegría	Tristeza
Confianza	Disgusto
Miedo	Ira
Sorpresa	Anticipación

Tabla 2
Emociones básicas y relación con el contraste de sentimientos (Plutchik, 1980)

Emociones básicas	Sentimiento	Sentimiento opuesto
Anticipación + Alegría	Optimismo	Desaprobación
Alegría + Confianza	Amor	Remordimiento
Confianza + Miedo	Sumisión	Desprecio
Miedo + Sorpresa	Temor	Agresividad
Sorpresa + Tristeza	Desaprobación	Optimismo
Tristeza + Disgusto	Remordimiento	Amor
Disgusto + Ira	Desprecio	Sumisi[on]
Ira + Anticipación	Agresividad	Temor

Realizar predicciones sobre el comportamiento de las personas, procesos de fidelización de clientes, identificación de tendencias y valoración del riesgo psicosocial, son solo algunos de los usos del análisis de sentimientos.

En (Potts, 2011) (Maas, Ng, & Potts, 2011) se resalta la importancia que cumplen las teorías cognitivas de la emoción para facilitar el proceso de identificación de contraste o semejanza de los sentimientos señalados en un texto. Por su parte, (Manning & Jurafsky,

2012) indican que si el análisis se realiza de forma computacional, es necesario identificar quien genera la emoción, la razón o el motivo que la genera, la polaridad de la misma y las sentencias en las que se manifiestan dichas emociones (Jaramillo, Cardona, & Espitia, 2017).

La técnica propuesta en este trabajo permite analizar y clasificar sentimientos haciendo uso de MapReduce (Dean & Ghemawat, 2008). MapReduce es un modelo de programación para trabajar con grandes volúmenes de datos en paralelo (Jaramillo & Londoño, 2014). El modelo propuesto construye, mediante MapReduce, índices inversos que permiten relacionar las expresiones con dos posibles sentimientos, a saber: positivo y negativo.

Además, utiliza medidas de similitud entre cadenas, para poder clasificar una expresión con el clúster más cercano.

Este artículo se estructura de la siguiente forma. En la sección 2, se muestran algunas aproximaciones referentes a analizadores de sentimientos. En la sección 3, se presenta el sistema propuesto. La sección 4 muestra detalles de implementación y la evaluación experimental realizada. Finalmente, la sección 5 presenta un resumen de los principales hallazgos de este trabajo.

2. Estado del arte

El término Análisis de sentimientos hace referencia a la capacidad de detección, recolección y análisis a través de sistemas, de la información contenida en un texto, es decir, opiniones que presentan un nivel de emoción. El sistema encargado de realizar el análisis, por ende, debe ser capaz de decidir si un texto expresa sentimientos positivos, neutrales o negativos. El llamado análisis orientado a aspectos, que se relaciona mucho con el análisis de sentimientos, se define como la capacidad de seguimiento de opiniones referidas a entidades, tales como los juicios sobre eventos o características con esas mismas entidades.

En los últimos años distintas comunidades de investigación han prestado gran interés al análisis de sentimientos, convirtiéndolo en un ámbito de mucha competencia e inversión. Este interés se debe al gran crecimiento de los canales de comunicación social, tales como blogs y redes sociales, que aportan al incremento de contenidos creados por usuarios, al permitir agregar comentarios, opiniones y demás formas de opinión. En (Pang, Lee, & Vaithyanathan, 2002) se describe un método de clasificación de opiniones de películas a promocionar como positivas o negativas. La clasificación se lleva a cabo haciendo uso de técnicas de aprendizaje supervisado. (Turney, 2002) presenta una aproximación que combina el uso de diccionarios compuestos de palabras y expresiones etiquetadas (polaridad a priori) con técnicas de procesamiento de lenguaje natural.

Hoy en día existen gran cantidad de aproximaciones para realizar análisis de sentimientos, que se apoyan en el modelo de programación MapReduce. En (Ha, Back, & Ahn, 2015) se propone un método para realizar análisis de sentimientos de datos no estructurados provenientes de medios sociales, haciendo uso de Hadoop (The Apache Software Foundation, 2014). El proceso de análisis de sentimientos empieza con la carga de los diccionarios de análisis de sentimientos. A partir de allí, los datos a procesar se llevan a un sistema de archivos distribuido paralelo, denominado (HDFS) y, haciendo uso de las funciones de mapeo y reducción, se realiza el análisis de sentimientos. El sistema propuesto plantea el uso de 4 funciones especiales de mapeo, a saber: análisis de preprocesamiento de polaridad, análisis de palabras sintácticas, análisis de morfemas y análisis prohibitivo de palabras. Los resultados obtenidos se almacenan en una base de datos.

En (Khuc, Shivade, Ramnath, & Ramanathan, 2012) se presenta un analizador de sentimientos, que consta de dos componentes: un generador de léxico y un clasificador de sentimientos. El generador de léxico lleva a cabo etapas tales como normalizar los tweets, generar una matriz de co-ocurrencia (mediante MapReduce) que se almacena en HBase, computar la similaridad de coseno y crear grafos de palabras, removiendo aristas de bajo peso. El segundo componente, el clasificador de sentimientos, se apoya en dos enfoques, uno basado en léxico y otro que combina léxico con algoritmos de aprendizaje de máquina. En (Bautin, 2010) se describe un sistema de acceso de mapeo de texto, denominado Lydia,

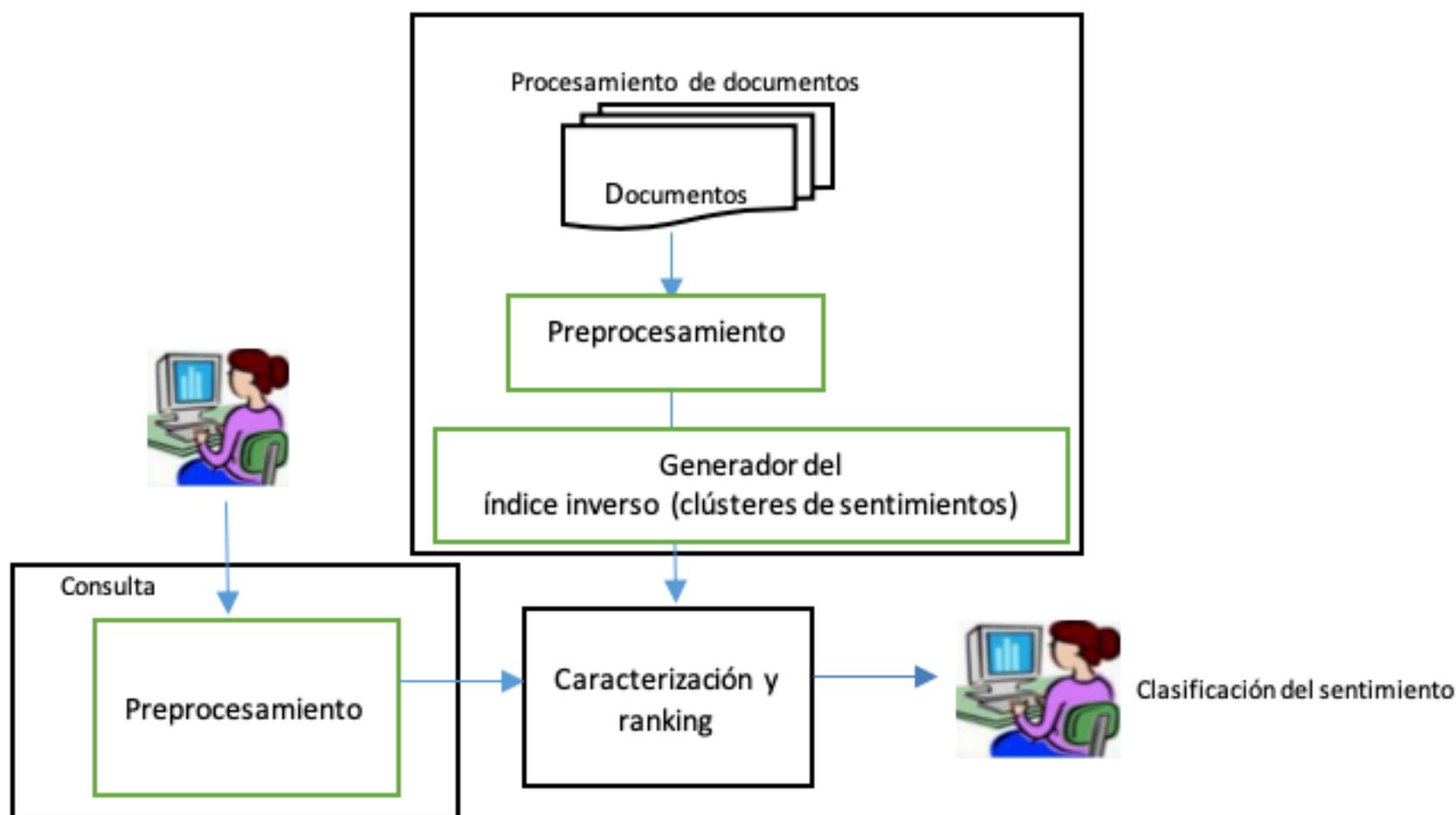
que proporciona acceso fácil a una gran cantidad de estadísticas sobre millones de personas, sitios y cosas utilizando el sistema Hadoop. Lydia consta de componentes principales tales como marcado NLP, análisis de sentimientos, análisis y agregación de entidades y visualización. El análisis de sentimientos se apoya en léxicos de palabras positivas y negativas, y consiste en la asociación de entidades con sentimientos de palabras co-ocurrentes desde esos léxicos. Los léxicos de sentimiento se construyen a partir de seis dominios específicos: negocio, crimen, salud, política, deportes y medios de comunicación. El léxico de sentimiento completo se presenta en un diccionario electrónico, tal como se describe en (Godbole, Srinivasaiah, & Skiena, 2007).

Finalmente, en (IBM, Watson AlchemyAPI., 2018) se presenta una API para dar solución a distintos problemas relacionados con la inteligencia artificial, entre ellos el análisis de sentimientos. Para realizar esta tarea se apoya en técnicas de procesamiento natural del lenguaje, lo que le permite realizar análisis de sentimientos a diferentes niveles: de documento, entidad y palabras clave.

3. Descripción del sistema propuesto

La estructura del sistema propuesto se compone por tres módulos fundamentales, a saber: el módulo de procesamiento de documentos, el módulo de procesamiento de la consulta y el módulo de caracterización-ranking, ver Figura 1.

Figura 1
Arquitectura general del modelo propuesto, basado en (Jaramillo & Londoño, 2014)



3.1. Preprocesamiento de los documentos y/o consultas

Este módulo recibe un documento con múltiples opiniones referentes a un tema de interés. Sobre cada una de las opiniones se efectúa un preprocesamiento, que contempla tres actividades. La primera consiste en eliminar de la opinión las palabras vacías o *stopwords*. Ello se logra comparando cada palabra de la misma con una lista preestablecida de *stopwords*. Se consideran *stopwords* a artículos, pronombres y preposiciones. La segunda, implica el uso del método de *stemming* o lematización, para trabajar directamente sobre la raíz semántica de cada palabra de la opinión (Jaramillo & Londoño, 2014). En la última etapa se remueven las palabras que no corresponden a sentimientos, este proceso se realiza

comparando la palabra con un lexicón de sentimientos (diccionario que asocia cada palabra a un sentimiento específico).

3.2. Módulo generador del índice inverso de los documentos

Este módulo hace uso MapReduce para generar un índice inverso, que permite agregar la opinión, luego de su preprocesamiento, a uno de los clúster de sentimientos. La opinión expresada en el documento tiene una clasificación (positiva o negativa). Se generan dos clústeres, en uno se agrupan las palabras u opiniones procesadas que contienen sentimientos negativos y en la otra los positivos.

3.3. Módulo de caracterización y ranking

Este módulo se encarga de calcular la similitud entre cada uno de los documentos y la consulta, haciendo uso del Modelo de Espacio Vectorial. Por cada opinión clasificada en los posibles clústeres de sentimientos y por cada consulta a clasificar, se obtiene un vector. (Singhal, 2001). El vector descrito se construye calculando un peso positivo por cada palabra de acuerdo a la cantidad de ocurrencias en el texto y se halla su similitud con la opinión a clasificar (Jaramillo & Londoño, 2014). La Ecuación (1) muestra el cálculo de la similitud de coseno, siendo p es el vector correspondiente a la opinión (que ya se encuentra clasificada), q la consulta ingresada, p_i es el peso de la palabra i en la opinión y q_i es el peso de la palabra i en la consulta.

$$\text{Similaridad } \cos(p, q) = \frac{\sum_{i=1}^n (p_i \cdot q_i)}{\sqrt{\sum_{i=1}^n p_i^2} \cdot \sqrt{\sum_{i=1}^n q_i^2}} \quad (1)$$

El resultado obtenido al calcular la distancia de coseno es un valor en el intervalo $[0,1]$, siendo 1 el mejor valor de cercanía que puede obtenerse (Jaramillo & Londoño, 2014).

El cálculo de la similitud de coseno se paraleliza haciendo uso modelo de programación MapReduce. Esto genera un nuevo índice inverso que asocia cada posible sentimiento (positivo o negativo) con las medidas de similaridad de coseno calculadas referentes a dicho sentimiento (distancia entre la consulta y el sentimiento). Al final, por cada sentimiento se halla la mayor similaridad obtenida. Si la similaridad obtenida por el sentimiento positivo es mayor (con un error de $\pm 5\%$), la opinión se clasifica como positiva, de lo contrario se clasifica como negativa.

4. Implementación del analizador de sentimientos

4.1. Implementación

El prototipo fue desarrollado en Java. Se utilizaron las bibliotecas Hadoop y Lucene Analyzer de Apache Software Foundation. Hadoop (The Apache Software Foundation, 2014) es un framework que implementa el modelo de programación MapReduce. Lucene (The Apache Software Foundation, 2014) es un API de Apache que permite aplicar lematización haciendo uso de la clase EnglishStemmer. Para ampliar la precisión de la clasificación se usa el Lexicón de sentimiento Subjectivity Lexicon presentado en (Wilson, Wiebe, & Hoffmann, 2005).

El proceso de creación de los dos índices inversos (clústeres de sentimientos y clústeres con similaridades de coseno) se apoya en el modelo MapReduce. Para el índice que crea los clústeres de sentimiento, se toma como entrada cada una de las opiniones procesadas y su correspondiente clasificación. El procesamiento realizado se efectúa en 2 fases: Mapeo y Reducción. Cada Mapper genera parejas de tipo clave-valor, donde la clave es el clúster (positivo o negativo) y el valor es la opinión. Esto produce un listado de parejas intermedias que se almacena en el sistema de archivos distribuido de Hadoop(HDFS). El reducer itera sobre los datos y por cada llave encontrada (en este caso, por cada clúster), pasa la llave y los valores asociados al mismo a un archivo resultante (Jaramillo & Londoño, 2014). El

procesamiento realizado da como resultado un índice en el que se mapea cada clúster con las sentencias u opiniones correspondientes al mismo. Para los clústeres de similitudes de coseno, la clave es el clúster y el valor es la similitud (distancia de coseno entre la frase extraída de dicho clúster y la opinión a clasificar). La salida es un índice en el que se mapea cada clúster con las distancias de similitud de coseno.

4.2. Evaluación

Para la evaluación del sistema se utilizó el *dataset* Sentiment Labelled Sentences descrito en (Kotzias et. al, 2015). Este *dataset* contiene opiniones etiquetadas con una clase (0 ó 1). Si el sentimiento es positivo la etiqueta es 1, si es negativo es 0. Las opiniones fueron extraídas de 3 sitios *web* diferentes *imdb.com*, *amazon.com* y *yelp.com*. Para cada sitio *web*, hay 500 opiniones marcadamente positivas y 500 negativas. No hay clasificación de sentencias con connotación neutra. Para determinar la precisión del clasificador de sentimientos, cada grupo de opiniones se divide en dos conjuntos: uno con el 70% de los datos, para entrenar el modelo y el 30% restante, para probar el modelo. Con los datos correspondientes al set de entrenamiento (70% del *dataset*, más palabras correspondientes al diccionario de sentimientos) se construye el índice inverso que permite la creación de los dos clústeres mediante MapReduce. Cada opinión del set de prueba es enviada al analizador de sentimientos, quien se encarga de clasificarla en uno de los dos clústeres. La clasificación obtenida es contrastada contra la clasificación real (*ground truth*). Para evaluar la calidad de las clasificaciones se usaron dos índices: exactitud (*accuracy*) y el índice de *Youden* (Youden, 1950). La exactitud, ver (2), es la proporción de resultados verdaderos entre la cantidad total de ejemplos observados (Metz, 1978). El índice de *Youden*, ver (3), es una estadística que corresponde a la mejor combinación de sensibilidad y especificidad en la predicción y toma valores entre -1 y 1 (Youden, 1950) (Jaramillo, Londoño, & Cardona, 2018).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$J = \frac{TP}{TP+FN} + \frac{TN}{TN+FP} - 1 \quad (3)$$

En las ecuaciones, TP es el número de verdaderos positivos, FP es el número de falsos positivos, TN el número de verdaderos negativos y FN el número de falsos negativos. Las medidas obtenidas para cada *dataset* se presentan en la Figura 2.

Figura 2

Comparación de rendimiento para los *datasets* *imdb.com*, *amazon.com* y *yelp.com*

Fuente: Imagen propia

Los resultados del análisis de sentimiento con el sistema propuesto son muy cercanos a la clasificación real. Los casos en los cuales el sistema falla, son aquellos en los cuales hay similar cantidad de palabras positivas y negativas en un mismo comentario, también cuando la opinión expresa sarcasmo.

5. Conclusiones

Este artículo presenta un analizador de sentimientos basado en el modelo de programación MapReduce. Este nuevo sistema ofrece ventajas tales como clústeres que contienen un adecuado lexicón de sentimientos y mejoras en el rendimiento al hacer uso de un proceso de indexación, que se apoya en un *framework* de computación distribuida.

El uso del lexicón de sentimientos contribuye a mejorar la calidad de la clasificación. Sin embargo, se observa la necesidad de trabajar en la construcción de diccionarios específicos por cada contexto, dado que puede darse el caso de palabras que en un contexto tengan polaridad negativa, pero aplicado a otros contextos sea positiva, como es el caso de la palabra *hot* (Rosales, Ayala, Pinto, Tovar, & Beltrán, 2016).

Como trabajo futuro se plantea la creación de un nuevo clúster que contenga opiniones neutras. Además, de la posibilidad de mejorar el proceso de identificación orientación semántica de la opinión, mediante el uso de un algoritmo de clasificación supervisado y web semántica.

Finalmente, respecto al análisis de sentimientos, es importante anotar que aún existen muchos aspectos que deben investigarse para aportar a una mejor detección de los sentimientos, entre ellos se destacan aspectos tales como la sinonimia, sarcasmo, morfología, ortografía, procesamiento de la negación y adaptación al dominio (Liu, 2012) (Pang & Lee, 2008).

Referencias bibliográficas

- THE APACHE SOFTWARE FOUNDATION. (2014). **Welcome to Apache™ Hadoop®!**. Recuperado de <http://hadoop.apache.org/>
- BAUTIN, M. W. (2010). **Access: news and blog analysis for the social sciences**. Proceedings of the 19th International World Wide Web Conference (WWW 2010).
- CASTILLA, D. P., & Carlos. (2010). **Teoría de los sentimientos**. TUSQUETS EDITORES.
- DEAN, J., & GHEMAWAT, S. (2008). **MapReduce: Simplified Data Processing on Large Clusters**. Commun. ACM, 51, 107-113.
- FOUNDATION, T. (2016). **Welcome to Apache Lucene**. Recuperado de <http://lucene.apache.org/>
- GODBOLE, N., SRINIVASIAH, M., & SKIENA, S. (2007). **Large-Scale Sentiment Analysis for News and Blogs**. Proceedings of the International Conference on Weblogs and Social Media (ICWSM).
- GOEL, A., & MUNAGALA, K. (2012). **Complexity measures for MapReduce, and comparison to Parallel computing**. Recuperado de <http://www.stanford.edu/~ashishg/papers/mapreducecomplexity.pdf>
- HA, I., BACK, B., & AHN, B. (2015). **MapReduce Functions to Analyze Sentiment Information from Social Big Data**. En International Journal of Distributed Sensor Networks, 11, 417502.
- IBM. (2018). **Watson AlchemyAPI**. Recuperado de <https://www.ibm.com/watson/alchemy-api.html>
- JARAMILLO, S., & LONDOÑO, J. M. (2014). **Búsqueda de documentos basada en el uso de índices ontológicos creados con MapReduce**. Ciencia e Ingeniería Neogranadina, 24, 57-75. Recuperado de <http://www.redalyc.org/articulo.oa?id=91132760004>
- JARAMILLO, S., CARDONA, S., & ESPITIA, E. (2017). **Minería sobre streams de datos: una nueva estrategia para afianzar las relaciones de confianza en el mundo empresarial**. Revista Espacios, Vol. 39 N° 13, pág. 6.
- JARAMILLO, S, CARDONA, S, & FERNANDEZ, A. (2015). **Minería de datos sobre streams de redes sociales, una herramienta al servicio de la Bibliotecología**. Información, cultura y sociedad: revista del Instituto de Investigaciones Bibliotecológicas, 63-74. Recuperado de <http://www.redalyc.org/articulo.oa?id=263042678005>
- JARAMILLO, S., LONDOÑO, J., & CARDONA, S. A. (2018). **ClusCTA MEWMACChart**. Revista Espacios, Vol 27.
- KHUC, V. N., SHIVADE, C., RAMNATH, R., & RAMANATHAN, J. (2012). **Towards building large-scale distributed systems for twitter sentiment analysis**. SAC.
- KOTZIAS et. al. (2015). **From Group to Individual Labels using Deep Features**. KDD 2015.
- LEVENSHTEIN. (1966). **Binary codes capable of correcting deletions, insertions, and reversals**. Soviet Physics Doklady 10: 707–10. Recuperado de Binary codes capable of correcting deletions, insertions, and reversals.
- LIU, B. (2012). **Sentiment Analysis and Opinion Mining**. Morgan & Claypool Publishers.

- MAAS, A., NG, A., & POTTS, C. (2011). **Multi-Dimensional Sentiment Analysis with Learned Representations**. unpublished.
- MANNING, & JURAFSKY. (2012). **Free online course Natural Language Processing**. Standford.
- MARTÍNEZ, A. (2013). **Big Data**. Recuperado el 22 de mayo de 2013, de <http://www.youtube.com/watch?v=pnseIuf-OfE>
- PANG, & LEE. (2008). **Foundations and Trends in Information Retrieval, Opinion mining and sentiment analysis**. Now, the essence of knowlegde.
- PANG, B., LEE, L., & VAITHYANATHAN, S. (2002). **Thumbs up? Sentiment Classification using Machine Learning Techniques**. CoRR, cs.CL/0205070. Recuperado de <http://arxiv.org/abs/cs.CL/0205070>
- PLUTCHIK. (1980). **Discovering Basic Emotion Sets via Semantic Clustering on a Twitter Corpus, Wheel of Emotions**. ResearchGate.
- PLUTCHIK, R. (2002). **Emotions and Life: Perspectives from Psychology, Biology, and Evolution**. American Psychological Association.
- PORTA, J. (2006). **Clasificación de patrones**. Recuperado de <http://arantxa.ii.uam.es/~jporta/iula/unsupervised.slides.pdf>
- POTTS, C. (2011). **On the Negativity of Negation**. En N. Li, & D. Lutz (Edits.), *Proceedings of Semantics and Linguistic Theory 20* (págs. 636-659). Ithaca, NY: CLC Publications.
- SALTON, G. (s.f.). (1971). **The Smart retrieval system–experiments**. Prentice–Hall.
- SALTON, G. W., & YANG. (s.f.). (1975). **A vector space model for automatic indexing**. *Communications of the Association for Computing (18.11)*, 613–620.
- SCHERER, K. R. (2005). **What are emotions? And how can they be measured? Social Science Information**, 44, 695-729. doi:10.1177/0539018405058216
- SCHERER, K., & EKMAN, P. (1984). **Approaches to emotion**. L. Erlbaum Associates,.
- SINGHAL, A. (2001). **Modern information retrieval: a brief overview**. *BULLETIN OF THE IEEE COMPUTER SOCIETY TECHNICAL COMMITTEE ON DATA ENGINEERING*, 24, 2001.
- TORRES, J. (21 de marzo de 2013). **Retos del Big Data**. Recuperado el 28 de mayo de 2013, de <http://www.slideshare.net/jorditorres/retos-del-big-data>
- TURNEY, P. D. (2002). **Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews**. CoRR, cs.LG/0212032. Recuperado de <http://arxiv.org/abs/cs.LG/0212032>
- WILSON, T., WIEBE, J., & HOFFMANN, P. (2005). **Recognizing Contextual Polarity in Phrase-level Sentiment Analysis**. *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing* (págs. 347-354). Stroudsburg: Association for Computational Linguistics. doi:10.3115/1220575.1220619

1. Estudiante Ingeniería de Sistemas y computación. Universidad del Quindío, UQ, Armenia, (Colombia), kdsanchezs@uqvirtual.edu.co

2. Estudiante Ingeniería de Sistemas y computación. Universidad del Quindío, UQ, Armenia, (Colombia), kdsanchezs@uqvirtual.edu.co

3. Doctora en Ingeniería. Computer Engineering Dept., Universidad del Quindío, UQ. Armenia, (Colombia), sjaramillo@uniquindio.edu.co