



niverso

de la Tecnológica®

ISSN: 2007-1450

Biodiesel

Jatropha Curcas

Bioenergéticos, alternativa
para la sustitución de
combustibles fósiles

CONOCE TAMBIÉN

Cómo se clasifican
los **textos de opinión**

La porosidad de un
biorreactor de lecho
fluidizado inverso

La **ovinocultura** en el
estado de Zacatecas

Respuesta
mecánica en
ducto agrietado

Universidad Tecnológica de Nayarit
Año V Edición N° 15 Abril/Julio 2013



ÍNDICE

Editorial	4
-----------	---

DE OPINIÓN

Desafíos y liderazgo de la educación en América Latina	5
Conoce las principales fuerzas de cambio...	

DE INVESTIGACIÓN

Artículo arbitrado	
Obtención de biodiesel a partir de <i>Jatropha Curcas</i>	7
Por Juan Antonio Alfonso Alvarez, Silvia María Balderas López, Patricia Alvarado Morán, Karina Veliz Zamorano y Juan Guillermo Hernández Pérez.- Universidad Tecnológica de Tijuana	

Artículo arbitrado	
Clasificación de textos de opinión	11
Por Adelina Escobar Acevedo y Rafael Guzmán Cabrera Universidad Tecnológica del Valle de Toluca y Universidad de Guanajuato	

Artículo arbitrado	
Correlación matemática para el cálculo de la porosidad de un biorreactor de lecho fluidizado inverso	14
Por Karla Yolanda Alvarez Pereyra y Roberto Limas Ballesteros Universidad Tecnológica de Cancún e Instituto Politécnico Nacional	

Artículo arbitrado	
La productividad de la ovinocultura en el estado de Zacatecas, México	18
Por José María Carrera Chávez y Benjamín Carrera Chávez Universidad Autónoma de Ciudad Juárez	

Artículo arbitrado	
Análisis de la respuesta mecánica de un ducto agrietado sujeto a presión interna aplicando el método de los elementos finitos	22
Por Javier Flores Méndez, Gerardo Luis Velázquez García y Marcos Espinoza Martínez Instituto Tecnológico de Puebla, Universidad Tecnológica de Puebla y Universidad Tecnológica de Huejutzingo	

Instrucciones para autores	25
----------------------------	----

UNIVERSO DE LA TECNOLÓGICA: es una revista científica de la Universidad Tecnológica de Nayarit, que publica aportaciones originales sobre resultados de investigación de diferentes áreas del conocimiento, publicada cuatrimestralmente, fundada en 2008 y disponible en el formato impreso con registro ISSN 2007-1450. También se encuentra albergada electrónicamente en <http://utnay.edu.mx/files/revista15.pdf>

UNIVERSO DE LA TECNOLÓGICA está incluida en:

LATINDEX, Sistema de Información sobre las revistas de investigación científica, técnico-profesionales y de divulgación científica y cultural que se editan en los países de América Latina, el Caribe, España y Portugal.



REVISTA **UNIVERSO DE LA TECNOLÓGICA**, Año V, No. 15, Abril/Julio 2013, es una publicación cuatrimestral editada por la Universidad Tecnológica de Nayarit, a través de la Dirección de Vinculación. Carretera Federal 200 Km. 9, Tramo Xalisco-Compostela C.P. 63780, Xalisco, Nayarit, México. Tel. 01.311.211.98.00.

www.utnay.edu.mx universodelatecnologica@utnay.edu.mx

Editor responsable: Héctor Manuel Béjar Fonseca. Reservas de Derechos al Uso Exclusivo en Trámite, ISSN: 2007-1450. Licitud de Título en Trámite. Licitud de Contenido en Trámite, ambos otorgados por la Comisión Calificadora de Publicaciones y Revistas Ilustradas de la Secretaría de Gobernación. Permiso SEPOMEX en Trámite. Impresa por los Talleres Gráficos de la Imprenta Comercial El Debate, Obregón 55 Ote. Colonia Centro en Los Mochis, Sinaloa. Este número se terminó de imprimir en agosto de 2013 con un tiraje de 1,000 ejemplares. *Universo de la Tecnológica* se distribuye en forma gratuita a nivel estatal: bibliotecas públicas y privadas, interior de la institución, empresas privadas e instituciones de gobierno con las que existen convenios. Nivel nacional: Universidades Tecnológicas del país. Internacionalmente: Universidades de España y Francia con las que se realiza movilidad estudiantil.

Las opiniones expresadas por los autores no necesariamente reflejan la postura del editor de la publicación.

Queda estrictamente prohibida la reproducción total o parcial de los contenidos e imágenes de la publicación sin previa autorización del Instituto Nacional del Derecho de Autor.

Clasificación de textos de opinión

Por Adelina Escobar Acevedo y Rafael Guzmán Cabrera
Universidad Tecnológica del Valle de Toluca y Universidad de Guanajuato

Dirección electrónica del autor de correspondencia:
adeesa@yahoo.com

Recibido: Enero 22 de 2012
Aceptado: Marzo 8 de 2013

RESUMEN

El trabajo muestra resultados de una investigación de clasificación de opiniones electrónicas en español. Las opiniones son descargadas de blogs para su posterior análisis e interpretación. El análisis automático de opiniones es un área de reciente estudio con muchas aplicaciones, como son el sondeo automático de preferencias de productos comerciales, campañas políticas, temas sociales y culturales entre otros. En nuestro caso, para validar el método propuesto, utilizamos opiniones de películas, la idea principal de nuestro trabajo consiste en que dada una opinión no vista por el sistema propuesto, permita su ubicación automática como una opinión a favor o en contra, lo cual es de mucha utilidad cuando el número de opiniones es grande. Los resultados preliminares permiten ver la viabilidad del método así como su potencial aplicación en otras áreas.

PALABRAS CLAVES: Clasificación automática, análisis de sentimientos, textos de opinión, corpus español.

ABSTRACT

This work shows results of an investigation for automatic text categorization on Spanish opinions. Opinions are downloaded from blogs for further analysis and interpretation. Automatic opinion categorization is a recent area of study with many applications, such as monitoring commercial product preferences, political campaigning, social and cultural issues among others. In our case, to validate the proposed method, we use reviews of films, the main idea of our work is that given an not seen opinion, the proposed system allows the automatic location as an opinion are favorable or not, which is very useful when the number of opinions is large. Preliminary results let see the feasibility of the method and its potential application in other areas.

Keys words: Automatic text categorization, sentiment analysis, text opinion, spanish corpus.

Introducción

En la actualidad las tecnologías de la información y comunicación tienen alta presencia en la vida diaria, permitiendo intercambio de información de cualquier parte del mundo. Los usuarios tienden a verter sus opiniones sobre varios productos y servicios a la red, haciéndolas públicas al resto del mundo, por ello, algunas organizaciones están interesadas en realizar su clasificación. La información resulta útil, por ejemplo, para guiar a las compañías en el rumbo de los consumidores. Las opiniones reflejan el gusto o desaprobación de los usuarios, indicando a la organización de forma indirecta sobre las preferencias, quejas, aceptación del producto, tendencia y expectativa de los usuarios. Las opiniones también reflejan preferencias de candidatos, percepciones ciudadanas sobre programas culturales y sociales, etc.

Debido a la imposibilidad humana de manejar enormes cantidades de textos, nace la necesidad de automatizar procesos que manipulen y organicen tales volúmenes de documentos (Galicia Haro & Gelbukh, 2007). La clasificación automática de textos es comúnmente

una tarea supervisada, se conocen las categorías y se cuenta con un conjunto de entrenamiento previamente etiquetado. Cada documento es representado por una serie de variables características de la clase. La tarea de clasificación se precede por el aprendizaje de un modelo que usa información del conjunto de entrenamiento, almacenando los valores de las variables junto con la clase a la que pertenecen (Sierra Araujo, y otros, 2006).

En el estado del arte se encuentran trabajos que intentan identificar palabras que son asociadas a priori con conceptos subjetivos o emotivos (Riloff & Wiebe, 2003); (Kim & Hovy, 2004). Por otro lado, otros trabajos intentan determinar la subjetividad de una oración. Estos trabajos (Wiebe & Riloff, 2005) intentan determinar las frases objetivas, aquellas que afirman o describen un hecho, de aquellas frases subjetivas, las cuáles expresan una opinión. Por último, existen trabajos orientados a la identificación de documentos subjetivos, los cuales clasifican un texto completo como positivo o negativo (Das & Chen, 2001); (Turney, 2002); (Pang et al., 2002). Se encuentran pocos trabajos relacionados con la clasificación de opiniones en

español (Rosá et al., 2010); (Hernández, 2009), nuestra investigación tiene como objetivo aumentar la clasificación por métodos semisupervisados.

Metodología

La creación del corpus: Implica la recopilación, lectura, segmentación y etiquetado del corpus de entrenamiento y prueba por una o varias personas que determinen a su juicio si el texto debe pertenecer a una clase u otra. El corpus fue obtenido del blog <http://www.ciao.es/>, ya que hasta el momento existen pocos en idioma español. El proceso de separación de las opiniones se llevo a cabo considerando las calificaciones asignadas en dicho blog.

El pre procesamiento: Unifica los textos provenientes de diversas fuentes a una codificación única permitiendo homogeneizar caracteres, eliminado etiquetas, símbolos, acentos, mayúsculas etc. En esta etapa se considera la eliminación de palabras vacías como artículos, conectores, etc.

Representación de documentos: Incluye esquemas de pesado y reducción de dimensionalidad, define la forma de representación utilizada basándose en los objetivos de la clasificación. La selección de estos criterios no puede ser al azar, debe probarse y compararse para determinar las ventajas de un esquema particular. La dimensionalidad es uno de los problemas más fuertes del trabajo con textos debido a que fácilmente pueden obtenerse y trabajarse matrices con miles de columnas y reglones. Los cálculos implicados repercuten fuertemente en los tiempos de solución y por ello debe considerarse alguna forma de reducción de atributos apropiada para el problema.

Construcción del clasificador: Existen clasificadores que han probado reiteradamente su eficacia en el procesamiento de textos, tales como Nàive Bayes, Vecinos más cercanos y Máquinas de vectores de soporte. Para ser comparables con otros trabajos de investigación debe considerarse la obtención de resultados con los clasificadores conocidos.

La finalidad es poder ubicar de manera automática una nueva opinión no vista por el sistema como a favor o en contra. En la siguiente sección se presentan los resultados preliminares obtenidos, así como algunos detalles de operación del método propuesto y descrito en esta sección.

Resultados

Se descargaron opiniones de tres películas, para cada una de ellas se manejaron dos categorías: a favor y en contra, consideramos una opinión a favor como positiva y una en contra como negativa. Las películas para las cuales se descargaron las opiniones, así como el número de opiniones descargadas para cada una, se presentan en la tabla 1.

Tabla 1: Corpus de trabajo

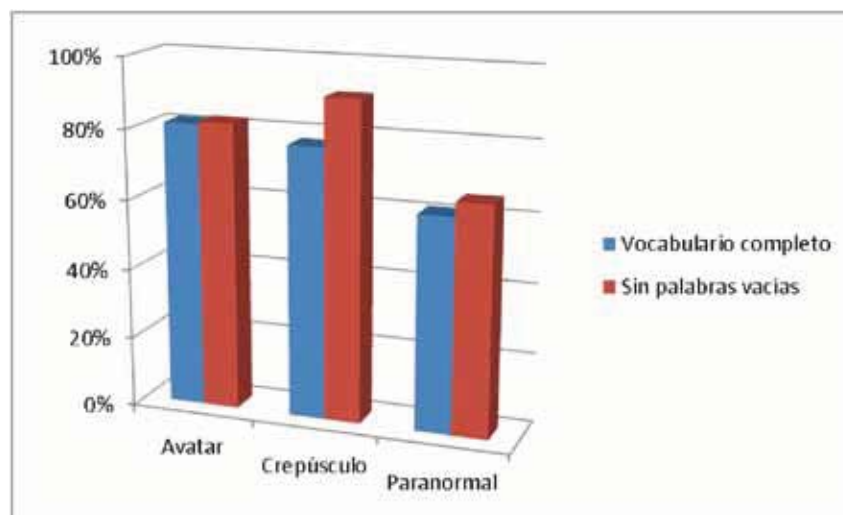
Película	Positivas	Negativas
Avatar	124	25
Crepúsculo	47	14
Paranormal activity	22	27

En los primeros experimentos utilizamos lo que comúnmente se conoce como “bolsa de palabras”, es decir, se extrae el vocabulario del conjunto de entrenamiento y se construye con ellas el clasificador automático, posteriormente se representa de forma vectorial cada documento del conjunto de prueba y se determina la exactitud de la clasificación. Para probar el método de clasificación automática se utilizó validación cruzada, específicamente la llamada “ten fold cross validation”, la cual consiste en partir el conjunto de datos en diez partes iguales, se entrena con nueve de ellas y se prueba con una, luego esta se incorpora como parte del conjunto de entrenamiento y otra parte es tomada como prueba, el resultado final se obtiene de los promedios de las mismas. El número de dobleces puede variar. La Tabla 2 presenta la exactitud de la clasificación en los escenarios de vocabulario completo y eliminación de palabras vacías (stop words) con tres dobleces en la validación cruzada.

Tabla 2: Resultados de la clasificación con validación cruzada.

Película	Vocabulario Completo	Sin palabras vacías
Avatar	81.24%	81.92%
Crepúsculo	77.27%	90.91%
Paranormal	61.15%	65.32%

Figura 1: Comparación de exactitudes con ambos escenarios



Como puede observarse, la eliminación de palabras vacías favorece la clasificación en bolsa de palabras, en el caso de la película crepúsculo, el aumento supera los 13 puntos porcentuales. Otra observación permite comparar el número de noticias de la tabla 1 y los resultados, indicando que el número de opiniones no influye en la clasificación, sino la riqueza de las mismas.

Conclusiones

La existencia de grandes cantidades de información en formato electrónico en blog exige el contar con nuevas herramientas que permitan organizar dicha información para poder sustentar una toma de decisiones correctamente. En este trabajo se presentan resultados preliminares de clasificación automática de opiniones de películas, las cuales son clasificadas como positivas o negativas, como método de clasificación se utilizó Naïve bayes, los resultados obtenidos permiten ver la viabilidad del método propuesto en el presente trabajo.

Actualmente se está trabajando en la fase semisupervisada, como trabajo futuro se espera realizar clasificación utilizando dominios cruzados (se entrena con un grupo de opiniones sobre un tema y se clasifica sobre otro tema. Y la posibilidad de realizar clasificación translingüe con idioma español como objetivo, dado que los avances más significativos se han logrado en idioma inglés.

Referencias bibliográficas

- Das, S., & Chen, M. (2001). Yahoo form Amazon Opinion Extraction from Small Talk on the Web. Proceedings of the 8th Asia Pacific Finance Association annual Conference.
- Galicia Haro, S. N., & Gelbukh, A. (2007). *Investigaciones en Análisis Sintáctico para el Español*. México: Instituto Politécnico Nacional.
- Hernández, Laritza, López-López A. & Medina, José E. (2009). *Recognizing Polarity and Attitude of Words in Text*, New Trends in Artificial Intelligence, Procs. 14th Portuguese Conference on Artificial Intelligence, EPIA 2009.
- Kim, S.-M., & Hovy, E. (2004). *Determining the sentiment of opinions*. International Conference on Computational Linguistics, (págs. 1367,1373). Geneva.
- Riloff, E., & Wiebe, J. (2003). *Learning Extraction Patterns for Subjective Expressions*. Proceedings of the Conference on Empirical Methods in Natural Language Processing.
- Rosá, Aiala, Wonsever, Dina & Minel, Jean-Luc (2010). *Opinion Identification in Spanish Texts*, Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas.
- Sierra Araujo, B., Arbelaiz, O., Armañanzas, R., Arruti, A., Bahamonde, A., Borrajo, D., y otros. (2006). *Aprendizaje Automático: Conceptos básicos y avanzados*. (M. Martín Romo, Ed.) España: Pearson Educación S.A.
- Turney, P. (2002). *Thumb Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews*. Proceedings of the 40 th Annual Meeting of the Association for Computational Linguistics.
- Wiebe, J., & Riloff, E. (2005). *Creating Subjective and Objective Sentence Classifiers form Unannotated Texts*. CiCLing.