

# Técnicas de clasificación de opiniones aplicadas a un corpus en español

## *Opinion classification techniques applied to a Spanish corpus*

Eugenio Martínez Cámara

M. Teresa Martín Valdivia

José M. Perea Ortega

L. Alfonso Ureña López

Departamento de Informática, Escuela Politécnica Superior de Jaén

Universidad de Jaén, E-23071 – Jaén

{emcamara, maite, jmperea, laurena}@ujaen.es

**Resumen:** El análisis de sentimientos es una nueva tarea que combina técnicas de minería de texto y Procesamiento del Lenguaje Natural (PLN). Aunque existen ya varios trabajos relacionados con la temática, la mayoría de ellos únicamente usan textos en inglés. Sin embargo, el número de páginas webs, blogs u opiniones vertidas en Internet que usan cualquier idioma, no sólo el inglés, crece exponencialmente. De hecho, otros idiomas como el español incrementan su presencia en la red día a día. Es por ello que en este trabajo se presenta un estudio experimental sobre un corpus de opiniones sobre películas escrito en español. Nuestro objetivo principal consiste en comprobar cómo se comportan varios clasificadores entrenados para determinar la polaridad de las opiniones en dicho corpus. Para ello hemos usado dos clasificadores (SVM y Naïve Bayes) variando además distintos parámetros como el esquema de pesado o la utilización o no de *stopper* y *stemmer*. Los experimentos realizados muestran que SVM se comporta mejor que Naïve Bayes y que el uso de *stopper* y *stemmer* también mejora los resultados.

**Palabras clave:** minería de opiniones, clasificación de la polaridad, corpus de opiniones en español, análisis de sentimientos

**Abstract:** Sentiment analysis is a new challenging task related to Text Mining and Natural Language Processing (NLP). Although there are some current works, most of them only focus on English texts. However, web pages, blogs and opinions on the Internet are increasing every day in any language and not only in English. Other language like Spanish is increasingly present so we have carried out an experimental study with a Spanish films reviews corpus. Our main goal is to check the results obtained using several classifiers trained in order to determinate the opinion polarity. We have tested two classification algorithms (SVM, Naïve Bayes) and several weighting schemes and different linguistic preprocessing (stopper and stemmer). The accomplished experiments show that SVM works better than Naïve Bayes. In addition, the stopper and stemmer also obtain a slight improvement.

**Keywords:** Opinion mining, sentiment polarity classification, opinion corpus in Spanish, sentiment analysis

## 1 *Introducción*

La minería de opiniones (*Opinion Mining* – OM) o análisis de sentimientos está cobrando cada vez mayor importancia debido fundamentalmente a la gran cantidad de comentarios que se escriben en Internet por parte de millones de usuarios de todo el

mundo a través de blogs, foros o redes sociales. En Internet ya proliferan las *webs* de opiniones, generalistas o específicas, sobre cualquier ítem como hoteles, productos electrónicos, coches, películas, temas políticos... Algunos ejemplos son *Epinions*<sup>1</sup> o *Booking*<sup>2</sup>. La razón fundamental

---

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

para la, cada vez, mayor presencia de este tipo de *webs*, es que la forma de consumir y de relacionarse de las personas ha cambiado. Los usuarios que van a realizar una compra prefieren informarse lo mejor posible antes de gastar su dinero. Tanto es así, que los grandes buscadores de Internet ya proporcionan opciones que permiten a sus usuarios buscar opiniones sobre los productos que están consultando. Un ejemplo muy claro es la versión estadounidense de *Bing*<sup>3</sup>. Pero no solo son importantes las *webs* específicas de opiniones, sino que también hay que tener muy en cuenta la multitud de redes sociales y la cantidad de blogs existentes, en las que los usuarios de todo el mundo comentan constantemente información de todo tipo y, en la mayoría de los casos, están opinando sobre algún tema. Además, aunque se podría pensar que los usuarios solo opinan sobre productos o servicios que consumen, la realidad no es así. Por ejemplo, la política, ya sea local, nacional o internacional, es otro tema sobre el que se escribe, opina o comenta en abundancia a cada instante en Internet.

Así pues, la minería de opiniones trata de clasificar los documentos en función de lo que expresa su autor. Esta nueva disciplina que combina PLN y minería de textos, incluye una gran cantidad de tareas que han sido tratadas en mayor o menor medida (Pang y Lee, 2008). Una de las principales aplicaciones consiste en determinar la polaridad de las opiniones a nivel de documento, frase o característica. De esta manera, podemos clasificar binariamente las opiniones en positivas o negativas.

Por otra parte, y a pesar de que las opiniones y comentarios compartidos en Internet no tienen restricción en cuanto al idioma utilizado, la gran mayoría de la investigación llevada a cabo relacionada con la minería de opiniones se centra casi exclusivamente en textos escritos en inglés. Sin embargo, cada vez son más los textos subjetivos que utilizan otros idiomas. Si bien el inglés es la lengua predominante en Internet, hay otros idiomas como el chino o el español que cada vez tienen más presencia

en la red<sup>4</sup> (Figura 1). Así pues, la investigación en OM no se debería centrar exclusivamente en un idioma sino que tendría también que estudiar otras lenguas, e incluso, el análisis de sentimientos desde una perspectiva multilingüe.

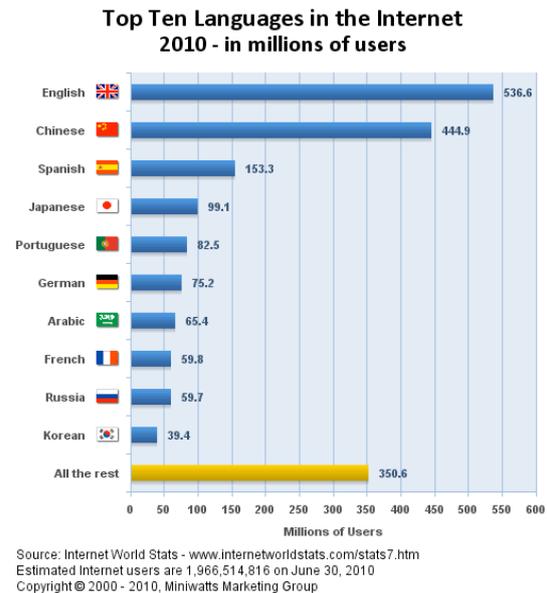


Figura 1: Presencia de los 10 primeros idiomas en Internet

Precisamente, el uso de otras lenguas en OM es uno de los motivos del presente trabajo. En este artículo se describen los experimentos realizados sobre un corpus de opiniones en español. Como una primera aproximación, se ha realizado un estudio con varios clasificadores para determinar la polaridad de los comentarios.

El presente artículo se estructura como sigue: en la segunda sección se presenta una breve descripción de otros trabajos relacionados prestando atención a aquellos que usan un idioma diferente al inglés. Posteriormente, se describe el corpus utilizado así como el proceso de preparación de datos para ser usados en los experimentos. A continuación, se describen los algoritmos utilizados en la experimentación. En la quinta sección se explica la ejecución de los algoritmos seleccionados, los resultados obtenidos, y su interpretación. Por último, planteamos las

<sup>2</sup><http://www.booking.com>

<sup>3</sup><http://www.bing.com/>

<sup>4</sup><http://www.internetworldstats.com/stats7.htm>

líneas de trabajo que se seguirán en el futuro.

## 2 Minería de opiniones en otros idiomas

A pesar de tratarse de una disciplina relativamente nueva, muchos han sido ya los trabajos realizados relacionados con la minería de opiniones y, más concretamente, con la clasificación de la polaridad. Se pueden distinguir dos metodologías principales a la hora de abordar el problema. Por una parte, la aproximación basada en aprendizaje automático que utiliza una colección de datos para entrenar los clasificadores. Por otra parte, la aproximación basada en orientación semántica, que no necesita un entrenamiento previo, sino que se tiene en cuenta la orientación de las palabras (positiva o negativa). Aunque la mayoría de los trabajos llevados a cabo en este ámbito utilizan textos escritos exclusivamente en inglés como conjunto de datos, también existen algunas investigaciones sobre otros idiomas.

Por ejemplo, el trabajo de Cruz y otros (Cruz et al., 2008) describe la creación del corpus en español utilizado en este artículo. Además se presenta un conjunto de experimentos realizados sobre dicha colección utilizando, en primer lugar, una orientación semántica con un clasificador no supervisado y, posteriormente, realizan experimentos con un clasificador supervisado obteniendo resultados ligeramente superiores. Denecke (Denecke, 2008) utiliza un corpus en alemán con revisiones de productos de Amazon para entrenar un clasificador que permita determinar la polaridad de las opiniones. Se utiliza un traductor automático para pasar los comentarios en alemán a inglés y posteriormente aplica SentiWordNet<sup>5</sup> (Esuli y Sebastiani, 2006), un recurso léxico para el análisis de sentimientos en inglés. Los resultados obtenidos son comparables a los conseguidos con un corpus en inglés. Zhang y otros (Zhang et al., 2009) aplican algoritmos basados en reglas y aprendizaje automático (SVM, Naïve Bayes y árboles de decisión) sobre dos corpus en chino. Uno de ellos trata comentarios sobre la eutanasia

mientras que el segundo se construyó a partir de los comentarios en chino de seis tipos de productos de Amazon. Otro trabajo es el realizado por Agić, Ljubešić y Tadić (Agić, Ljubešić, y Tadić, 2010) en el que se presenta un corpus con textos en croata recopilado de mercados financieros y anotado manualmente. Ghorbel y Jacot (Ghorbel y Jacot, 2010) usan una aproximación basada en aprendizaje automático sobre comentarios de películas en francés. Al igual que el trabajo de Denecke (2008) se traduce el corpus al inglés y se integra con SentiWordNet.

Por otra parte, también existen algunos trabajos que utilizan un entorno *multilingüe*. Por ejemplo, Ahmad, Cheng y Almas (Ahmad, Cheng, y Almas, 2006) trabajan sobre un corpus con noticias financieras en inglés, chino y árabe, mientras que Boldrini y otros (Boldrini et al., 2009) presentan el corpus EmotiBlog que incluye comentarios sobre varios temas en tres idiomas: español, inglés e italiano.

## 3 Corpus Muchocine

Para la realización de los distintos experimentos hemos utilizado un corpus en español de críticas de cine (Cruz et al., 2008). La colección está constituida por 3.878 críticas recogidas de la web *muchocine*<sup>6</sup>.

Las críticas que componen el corpus no están escritas por profesionales, sino por usuarios de la web. Esto que puede aparecer anecdótico aumenta la dificultad de la tarea, ya que los textos pueden no ser gramaticalmente correctos, pudiendo aparecer faltas de ortografía o expresiones informales.

Las críticas están puntuadas en un rango de 1 a 5, significando el 1 una película muy mala, y el 5 una película muy buena. Las películas puntuadas con 3 se pueden catalogar como neutras, o que el crítico no las considera ni malas ni buenas. Para la realización del estudio no se han considerado estas críticas neutras puntuadas con 3. Así, las cuatro clases restantes se van a reducir a dos, positiva o negativa. Las críticas con una valoración inferior a 3 se considerarán como negativas, mientras que

<sup>5</sup> <http://sentiwordnet.isti.cnr.it/>

<sup>6</sup> <http://www.muchocine.net>

las valoradas con un 4 o un 5, se etiquetan como críticas positivas. Finalmente, el número total de documentos sobre los que se han realizado los experimentos es de 2.625, de los cuales 1.274 se corresponden con críticas negativas, y 1.351 positivas.

Para cada documento se tiene el nombre del autor que realiza la crítica, el título de la película, la puntuación asignada, la máxima nota que se puede asignar a una película, la fuente de los datos, un breve resumen a modo de titular y la crítica en sí. De todos esos campos, para la experimentación, sólo nos interesa la puntuación de la película, el resumen de la crítica (Re), y la crítica desarrollada (Cr).

Por otra parte, no todas las críticas tienen la misma extensión, pero por lo que sí se caracterizan es porque expresan, de una forma más o menos profesional, la opinión del autor sobre la película. Cuando queremos expresar nuestro parecer sobre algún tema suele ocurrir que cuanto más nos extendemos más diluida queda nuestra opinión, a no ser que la hayamos hecho de forma muy cuidadosa. Las críticas que forman el corpus no tienen todas la misma extensión, y además no están redactadas por profesionales, por lo que lo más natural es que en muchas críticas la orientación de la misma no sea muy clara. En cambio, los resúmenes al estar formados por una frase o en algunos casos dos, expresan con mayor claridad la impresión que a los autores ha causado la película. Debido a esto, se han generado tres corpora: uno formado por los resúmenes (Re), otro solamente por la crítica (Cr), y por último, el constituido por los resúmenes y las críticas (Re\_Cr).

#### 4 Clasificadores utilizados

La mayoría de los trabajos sobre clasificación de documentos subjetivos se han realizado sobre textos en inglés. La falta de experimentación con opiniones en español es otro de los motivos de este

artículo, y por ello se han aplicado dos algoritmos de clasificación: SVM y Naïve Bayes. A continuación, se describen brevemente estos algoritmos de aprendizaje automático.

##### 4.1 SVM

El fundamento del algoritmo SVM es la capacidad de separar en dos clases un conjunto de datos (Vapnik, 1998). Se pueden aplicar a problemas de clasificación y de regresión en una infinidad de dominios. Debido a su versatilidad cuentan con diversas opciones de configuración. Para este trabajo se ha utilizado la implementación que ofrece libSVM, tipo C-SVC (Chang y Lin, 2001) y kernel lineal, tomando el valor del parámetro C igual a 0.

##### 4.2 Naïve Bayes

El clasificador Naïve Bayes (Mitchell, 1997) se fundamenta en el teorema de Bayes. Existen diversos métodos de estimación, siendo el elegido el estimador de Densidad de Kernel. Este varía su comportamiento en función del valor de su parámetro *bandwidth* (ancho de banda), que ha sido fijado a 0.05.

## 5 Experimentación

La experimentación se ha realizado sobre los tres corpora descritos en la Sección 3. A cada uno de estos corpora se les ha aplicado distintos tipos de *preprocesado*. Por un lado, se ha experimentado con la utilización o no tanto de *stopper* como de *stemmer*. Por otro lado, se han aplicado diferentes esquemas de pesado:

1. TF-IDF: Suele usarse en tareas de recuperación de información.
2. Frecuencia de ocurrencias (TF): Para cada documento es la frecuencia relativa de los términos de dicho documento.

	Stop	Stem	Re_Cr			Cr			Re		
			Prec.	Recall	F1	Prec.	Recall	F1	Prec.	Recall	F1
TF-IDF	√	√	86,84%	86,67%	86,75%	85,67%	85,49%	85,58%	76,23%	76,07%	76,15%
	√		87,66%	87,60%	87,63%	86,40%	86,34%	86,37%	75,75%	75,51%	75,63%
		√	86,80%	86,64%	86,72%	85,77%	85,56%	85,66%	75,95%	75,80%	75,87%
			87,73%	87,69%	87,71%	86,18%	86,10%	86,14%	75,57%	75,37%	75,47%
TF	√	√	79,81%	79,44%	79,62%	78,47%	78,20%	78,33%	74,74%	74,51%	74,62%
	√		77,48%	77,08%	77,28%	76,12%	75,65%	75,88%	72,12%	71,92%	72,02%
		√	79,74%	79,42%	79,58%	78,16%	77,91%	78,03%	74,74%	74,56%	74,65%
			77,06%	76,65%	76,85%	75,83%	75,34%	75,58%	72,29%	72,07%	72,18%
TO	√	√	77,66%	76,89%	77,27%	74,66%	73,13%	73,89%	73,82%	73,51%	73,66%
	√		74,05%	73,00%	73,52%	72,59%	70,96%	71,77%	71,84%	71,49%	71,66%
		√	77,86%	77,09%	77,47%	74,64%	73,03%	73,83%	74,08%	73,81%	73,94%
			74,25%	73,12%	73,68%	72,45%	70,77%	71,60%	71,87%	71,49%	71,68%
BTO	√	√	84,23%	84,20%	84,21%	83,61%	83,63%	83,62%	75,24%	75,07%	75,15%
	√		83,94%	83,91%	83,92%	83,69%	83,62%	83,65%	73,53%	73,24%	73,38%
		√	84,13%	84,12%	84,12%	84,10%	84,09%	84,09%	74,63%	74,45%	74,54%
			84,11%	84,16%	84,13%	83,84%	83,78%	83,81%	73,96%	73,70%	73,83%

Tabla 1: Resultados obtenidos con SVM

- Número de ocurrencia (TO): Para cada documento es la frecuencia absoluta de los términos de dicho documento.
- Ocurrencia binaria (BTO): A un término se le asigna el valor 1 si aparece en el documento, y 0 si no aparece.

Las medidas de evaluación de los clasificadores son las que se suelen usar en recuperación de información: *precision*, *recall*, y F1 (Sebastiani, 2002). Además, la evaluación se ha realizado utilizando validación cruzada con 10 particiones (10-*cross validation*).

Los resultados obtenidos para los experimentos realizados con SVM y Naïve Bayes se muestran en la Tabla 1 y Tabla 2, respectivamente.

Con respecto a la experimentación

realizada con SVM, el mejor resultado se ha obtenido con el corpus Re\_Cr, con el esquema de pesado TF-IDF y sin aplicar ningún mecanismo de reducción de características. Aunque esta configuración sea la mejor, no es la más eficiente debido a la cantidad de recursos que consume. Con el mismo esquema de pesado pero utilizando *stopper* y *stemmer* se consigue un resultado muy bueno (0,8675), llegando al 1.11% de diferencia con respecto al mejor (0,8771), por lo que la mínima pérdida que se produce en exactitud se gana en eficiencia.

			Re_Cr			Cr			Re		
	Stop	Stem	Prec.	Recall	F1	Prec.	Recall	F1	Prec.	Recall	F1
TF-IDF	√	√	83,16%	83,14%	83,15%	81,35%	81,32%	81,33%	75,37%	74,63%	75,00%
	√		84,08%	84,01%	84,04%	82,10%	82,01%	82,05%	75,23%	74,86%	75,04%
		√	83,47%	83,44%	83,45%	81,40%	81,35%	81,37%	75,87%	75,22%	75,54%
			83,62%	83,55%	83,58%	81,91%	81,82%	81,86%	74,83%	74,43%	74,63%
TF	√	√	68,49%	68,19%	68,34%	67,18%	66,89%	67,03%	65,45%	65,25%	65,35%
	√		66,54%	66,24%	66,39%	64,97%	64,63%	64,80%	62,48%	62,40%	62,44%
		√	67,53%	67,24%	67,38%	66,72%	66,47%	66,59%	65,77%	65,62%	65,69%
			65,57%	65,29%	65,43%	64,37%	64,10%	64,23%	63,32%	63,24%	63,28%
TO	√	√	69,21%	68,65%	68,93%	68,61%	66,71%	67,65%	74,80%	74,76%	74,78%
	√		72,12%	71,72%	71,92%	67,62%	65,49%	66,54%	73,59%	73,56%	73,57%
		√	69,25%	68,69%	68,97%	68,56%	66,66%	67,60%	74,62%	74,60%	74,61%
			72,32%	71,94%	72,13%	67,72%	65,61%	66,65%	73,46%	73,41%	73,43%
BTO	√	√	75,32%	78,58%	76,92%	75,18%	74,91%	75,04%	74,61%	74,58%	74,59%
	√		76,40%	75,88%	76,14%	76,29%	75,74%	76,01%	73,59%	73,54%	73,56%
		√	75,53%	75,28%	75,40%	75,37%	75,10%	75,23%	74,23%	74,20%	74,21%
			76,58%	76,07%	76,32%	76,21%	75,63%	75,92%	73,44%	73,38%	73,41%

Tabla 2: Resultados obtenidos con Naïve Bayes

De los resultados hay que destacar el buen comportamiento que ofrece el esquema de pesado BTO, que obtiene unos resultados que se acercan a los obtenidos por TF-IDF en los tres corpora (-3% para Re\_Cr, -2,34% para Cr y -1,33% para Re, siempre aplicando *stopper* y *stemmer* para la medida F1). En relación a TF y a TO se debe resaltar su mal comportamiento para Re\_Cr y para Cr, no siendo tan malo para Re si se compara con los resultados obtenidos para ese corpus.

La experimentación con Naïve Bayes no mejora los resultados obtenidos con SVM, pero tiene un comportamiento similar. Los mejores resultados se obtienen en el corpus Re\_Cr habiendo aplicado *stopper* y TF-IDF. Al igual que ocurría con SVM, la diferencia al aplicar conjuntamente *stopper* y *stemmer* (0,8315) llega prácticamente a un punto (-1,07%) con respecto a aplicar únicamente *stopper* (0,8404), por lo que la mejora en

rendimiento es mucho mayor que lo que se pierde en exactitud. A diferencia de SVM, en NB el esquema BTO no obtiene unos resultados cercanos a los de TF-IDF, por lo que para NB se puede concluir claramente que TF-IDF es el mejor esquema de pesado.

Si comparamos los resultados obtenidos con los experimentos realizados sobre el mismo corpus por Cruz y otros (Cruz et al., 2008), los cuales siguen un enfoque semántico, podemos observar que la configuración óptima de su sistema ofrece una precisión de 77,5%, lo que supone 9,34 puntos menos que la precisión obtenida por nuestra configuración más eficiente (86,84%). De esta forma se puede concluir, con los trabajos publicados hasta la fecha, que el enfoque de aprendizaje automático se comporta mejor en la tarea de clasificación de la polaridad en español.

## 6 Conclusiones y trabajo futuro

En este trabajo se ha realizado una primera aproximación a la minería de opiniones en español. Se ha estudiado el comportamiento de dos algoritmos de aprendizaje automático sobre tres corpora de tamaño diferente, a los que se les han aplicado distintas estrategias de reducción de características y de pesado. Con los resultados obtenidos se puede concluir que la combinación de *stopper*, *stemmer*, TF-IDF y SVM es la que mejor rendimiento ofrece.

Aunque en principio la experimentación realizada se ha centrado en el español, la idea es abordar el problema desde una perspectiva multilingüe logrando que, independientemente del idioma utilizado, se permita la clasificación de la polaridad en un texto subjetivo.

Por otra parte, los textos de opinión tienen un alto componente semántico y de conocimiento del dominio del problema. Por ello, nuestro trabajo futuro irá dirigido a la incorporación de información semántica a los análisis de los textos, mediante la utilización de recursos externos como SentiWordNet, DAL (“*Dictionary of Affect in Language*”) (Whissell, 1989), o General Inquirer<sup>7</sup> (Stone, 1966).

### Agradecimientos

Esta investigación ha sido parcialmente financiada por el Fondo Europeo de Desarrollo Regional (FEDER), proyecto TEXT-COOL 2.0 (TIN2009-13391-C04-02) del Gobierno Español, por la Junta de Andalucía, proyecto GeOasis (P08-TIC-41999), y por el Instituto de Estudios Giennenses, proyecto RFC/IEG2010.

### Bibliografía

- Agić, Z., N. Ljubešić, M. Tadić. 2010. Towards Sentiment Analysis of Financial Texts in Croatian. *In Proceedings of Language Resources and Evaluation (LREC)*.
- Ahmad, Cheng y Almas. 2006. Multi-lingual sentiment analysis of financial news streams. *Proceedings of Science, GRID2006*.

- Boldrini, E., A. Balahur, A., P. Martínez-Barco, and A. Montoyo. 2009. Emotiblog: an annotation scheme for emotion detection and analysis in non-traditional textual genres. *In DMIN*, pp 491–497. CSREA Press.
- Chang, C.C. y C.J. Lin, 2001. LIBSVM: a Library for Support Vector Machines. Software disponible en <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Cruz, F.L., J.A. Troyano, F. Enríquez, y J. Ortega. 2008. Clasificación de documentos basada en la opinión: experimentos con un corpus de críticas de cine en español. *Sociedad Española para el Procesamiento de Lenguaje Natural*, nº 41
- Del-Hoyo, R., I. Hupont, F.J. Lacueva, D. Abadía. 2009. Hybrid Text Affect Sensing System for Emotional Language Analysis. *Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots*
- Denecke, K. 2008. Using SentiWordNet for multilingual sentiment analysis. *In ICDE Workshops, pp 507–512. IEEE Computer Society*.
- Esuli, A. and F. Sebastiani. 2006. SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. *In Proceedings of Language Resources and Evaluation (LREC)*.
- Genkin, A., D.D. Lewis, D. Madigan. 2004. Large-Scale Bayesian Logistic Regression for Text Categorization. *Technometrics*, Vol. 49, No. 3
- Ghorbel, H. y D. Jacot. 2010. Sentiment analysis of French movie reviews. *Proceedings of the 4th international Workshop on Distributed Agent-based Retrieval Tools (DART 2010)*, June 2010, Geneva.
- Mitchell, T. 1997. Machine Learning. Ed. McGraw-Hill.
- Ortiz, Antonio Jesús, M.T: Martín, L.A. Ureaña, M.A. García. 2005. Detección automática de Spam utilizando Regresión Logística Bayesiana. *Procesamiento del Lenguaje Natural*. Nº 35, pp. 127-133. 2005.

<sup>7</sup> <http://www.wjh.harvard.edu/~inquirer/>

- Pang, B. and L. Lee, 2008. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135.
- Sebastiani, F. 2002. Machine learning in automated text categorization. *ACM Comput. Surv.*, 34(1):1-47.
- Stone, P. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. The MIT Press.
- Vapnik, V. 2008. *Statistical Learning Theory*. Wiley, Chichester, GB.
- Whissell, C.M. 1989. *The Dictionary of Affect in Language. Emotion: Theory, Research and Experience. Vol 4, The Measurement of Emotions*. R. Plutchik and H. Kellerman Eds. New York: Academic.
- Zhang, C., D. Zeng, J. Li, F.Y. Wang, and W. Zuo, 2009. Sentiment analysis of chinese documents: From sentence to document level. *JASIST*, 60(12):2474–2487.