

UNED LSI @ TASS 2013: Considerations about Textual Representation for IR based Tweet Classification

UNED LSI en TASS 2013: Consideraciones acerca de la Representación Textual para la clasificación de Tweets basada en Recuperación de Información

Ángel Castellanos González

Universidad Nacional de
Educación a Distancia
C/ Juan del Rosal 16, Madrid
acastellanos@lsi.uned.es

Juan Cigarrán Recuero

Universidad Nacional de
Educación a Distancia
C/ Juan del Rosal 16, Madrid
juanci@lsi.uned.es

Ana García Serrano

Universidad Nacional de
Educación a Distancia
C/ Juan del Rosal 16, Madrid
agarcia@lsi.uned.es

Resumen: Este artículo resume el trabajo planteado para nuestra participación en el TASS 2013, planteado como una extensión del trabajo realizado para el TASS 2012. El trabajo planteado el año anterior estaba enfocado en la clasificación de tweets basada en un enfoque de Recuperación de Información: las clases son modeladas de acuerdo a la información textual de los tweets pertenecientes a ellas y los tweets a ser clasificados son utilizados como query contra estos modelos. Este año hemos aplicado esta aproximación sobre las tareas de Sentiment Analysis y Topic Classification, pero este año nuestro trabajo está enfocado a analizar el tipo de información de los tweets a utilizar para llevar a cabo la clasificación y qué proceso debe seguirse para tener en cuenta esta información. En este sentido, hemos planteado diferentes tipos de modelado, así como diferentes maneras de llevar a cabo el proceso de Recuperación de Información de acuerdo a los diferentes tipos de información. Los resultados obtenidos sugieren que si bien la utilización de este tipo de información es valiosa (especialmente las entidades nombradas), debe hacerse siempre en conjunto con el contenido global de los tweets.

Palabras clave: Modelado de Temáticas, Análisis de Sentimientos, Modelado KLD, Etiquetado POS, Recuperación de Información

Abstract: This article summarizes the work proposed for our participation at TASS 2013, which is proposed as an extension of work done for TASS 2012. The work carried out the previous year was focused on the tweet classification based on an Information Retrieval (IR) approach: the classes are modeled according to the textual information of the tweets belonging to each class, and the tweets to be classified are used as query. This year we have applied this approach on Sentiment Analysis and Topic Classification tasks, but this year our work is focused on analyzing the type of tweet information to use to carry out the classification and what process should be followed to take this information into account. In this sense, we have proposed different types of modeling as well as different ways of performing the information retrieval process according to the different types of information. The results suggest that although the use of this type of information is valuable (especially named entities), it should always be done in conjunction with the overall content of the tweets.

Keywords: Topic Modelling, Sentiment Analysis, KLD Modelling, POS Tagging, Information Retrieval

1 Introducción

Este artículo describe nuestra participación en el Taller de Análisis de Sentimientos (TASS) 2013. El TASS está enfocado al análisis de sentimiento y reputación online en Twitter. El *análisis de sentimientos* (también conocido como minería de opinión) en Twitter consiste en la detección automática de la opinión (positiva, negativa, neutra) acerca de un tema o producto particular, por medio del análisis de un conjunto de tweets relativos a dicho tema o producto.

Debido a que el análisis de sentimientos necesita una precisión mucho mayor que la de la clasificación de textos tradicional, puesto que un pequeño cambio en el texto puede reflejar un gran cambio en la opinión (la fotografía me gusta vs. la fotografía no me gusta); no es factible abordar este problema con un enfoque basado en clasificación.

En su lugar diferentes metodologías han sido planteadas, principalmente basadas en técnicas de aprendizaje automático (Xia, Zong y Li, 2011) (Zhang et al., 2011), (Xu et al., 2011). Para una revisión más detallada sobre el trabajo en este campo se puede consultar en (Liu y Zhang, 2012) o nuestro resumen en (Castellanos, Cigarrán y García-Serrano, 2012b)

Además del análisis de sentimientos, otra de las tareas del TASS, está enfocada en la identificación automática de temáticas (Topic Detection). La identificación de temáticas se trata de un paso crucial en cualquier sistema de análisis de reputación online (ORM por sus siglas en inglés), de cara a la posterior evaluación de dichas temáticas por separado. Tanto es así que, campañas de evaluación centradas en ORM, como por ejemplo RepLab en su edición de 2012 (Amigó et al., 2012) y 2013 (Amigó et al., 2013), hacen especial énfasis en la detección automática de temáticas.

Para la detección automática de temáticas se han realizado trabajos aplicando técnicas clásicas de minería de textos; sin embargo, debido a que los mensajes en redes sociales (tweets) suelen tener una extensión reducida y a que existen otros problemas como el uso de abreviaturas, slang o jerga o vocabulario propio de estas redes, estas aproximaciones no han alcanzado resultados satisfactorios. En este sentido, se otras propuestas más exitosas son, por ejemplo, la aplicación de *Topic Models*

(Blei y Lafferty, 2009) o Author-Topic Model (Rosen-Zvi et al, 2010). Para ampliar información sobre este campo, se puede revisar los trabajos realizados en este campo en las dos ediciones de RepLab, previamente citadas.

En este trabajo se plantea una aproximación de anotación, tanto para el análisis de sentimiento como para la detección automática de temáticas, basada en un enfoque de recuperación de información. En este campo los autores ya han realizado trabajos previos, aplicando esta metodología a la anotación visual de conceptos (Benavent et al., 2013), a la detección de reputación online (Castellanos, Cigarrán y García-Serrano, 2013) y al propio campo del análisis de sentimiento y detección de temáticas (Castellanos, Cigarrán y García-Serrano, 2012b).

Este último trabajo sirve como base para el trabajo desarrollado este año en el TASS. Aplicando el mismo enfoque de trabajo allí presentado, la participación de este año busca cubrir las cuestiones que quedaron abiertas en la participación de la edición del TASS 2012 relativas al tipo de información a utilizar y la manera de utilizarla de cara a representar el contenido de los tweets. Como se vio en el trabajo del año pasado, la escasez de información textual de los tweets hace de su manejo un aspecto crucial de cara al rendimiento final de las aproximaciones de anotación.

Para la representación toda esta información, al igual que el año pasado, se aplicará una técnica de modelado basada en divergencias del lenguaje: la diferencia de Kullback-Leibler (Kullback y Liebler, 1951), probada con éxito para la representación de contenidos textuales frente a otras técnicas comunes en el estado del arte (Castellanos, 2013).

El resto del artículo se organiza como sigue: en la sección 2 se presenta el sistema desarrollado para llevar a cabo el trabajo; en las secciones 3 y 4 se exponen los diferentes experimentos realizados y se discute sus resultados; finalmente, en la sección 5 presentan las conclusiones y las líneas de trabajo futuras.

2 Descripción del Sistema

El sistema utilizado es igual al presentado en la pasada edición del TASS (Castellanos, Cigarrán y García-Serrano, 2012b). Este sistema aplica una aproximación basada en RI, utilizando el

los tweets de entrenamiento como base para la generación de los modelos y los índices utilizados.

Pueden distinguirse tres fases principales: 1) preprocesamiento, se filtra y etiqueta el conjunto de tweets original, 2) modelado, en la cual se genera la representación de cada una de las temáticas, así como de cada una de las distintas polaridades y, por último, 3) categorización, donde se lleva a cabo el proceso de asignación de cada tweet del conjunto de test a una polaridad y temática específicas utilizando el sistema de RI, utilizando el contenido del tweet como consulta contra el sistema.

Cada una de estas fases es explicada en detalle a continuación.

2.1 Preprocesado

En esta etapa se realiza un tratamiento previo del contenido de los tweets, el cual consta de los siguientes pasos:

- **Limpieza del contenido:** eliminación de caracteres especiales (puntos, comas, etc...) eliminación de palabras vacías y eliminación de términos propios de Twitter (menciones, hashtags y retweets).
- **Etiquetado POS:** Identificando las entidades nombradas y los adjetivos presentes. Para ello se han utilizado las herramientas de anotación presentadas en (Hernandez-Aranda, Granados y García-Serrano, 2012), basadas en la herramienta Stilus desarrollada por Daedalus¹.
- **Almacenamiento:** se han almacenado los resultados de las anotaciones realizadas en el paso anterior junto con la información relativa a su polaridad y temática.

2.2 Modelado

La fase de modelado toma como entrada los resultados obtenidos en la sección anterior para la generación de los modelos de los distintos niveles de polaridad y temática.

La metodología consiste en dividir el conjunto de tweets en sub-conjuntos según la polaridad que tengan (y de manera análoga para las temáticas), agregando el contenido de cada tweet a su correspondiente subconjunto. Tras ello, el proceso de modelado será aplicado de manera independiente para la polaridad y para

las temáticas utilizando la divergencia de Kullback-Liebler (KLD). Para más detalles acerca de la metodología de modelado se puede consultar el trabajo de (Castellanos, Cigarrán y García-Serrano, 2012a).

Mediante KLD es posible ordenar los términos de cada subconjunto en función de su representatividad de acuerdo a la fórmula (1), donde $pD(t)$ es la probabilidad de que el término t aparezca en el subconjunto S y $pC(t)$ la probabilidad de que el mismo término t aparezca en el resto de los subconjuntos.

$$KLD_{pS,pC} = pS(t) \cdot \log \left(\frac{pS(t)}{pC(t)} \right) \quad (1)$$

2.2.1 Modelado

Con el enfoque de modelado anterior se han desarrollado varios modelos, cada uno de ellos enfocados a experimentar con algún tipo de información a utilizar para llevar a cabo las tareas. Para la tarea de detección de temáticas se han generado los siguientes índices:

- **Modelado KLD del contenido del tweet (MTContent):** se ha aplicado KLD sobre la totalidad de los términos contenidos en los tweets asociados a las diferentes temáticas.
- **Modelado KLD de las entidades nombradas del tweet (MTNamedEntities):** se ha aplicado KLD sobre las entidades nombradas identificadas en cada uno los tweets asociados a las diferentes temáticas.
- **Modelado KLD de los hashtags del tweet (MTHashtag):** se ha aplicado KLD sobre la totalidad de los términos contenidos en los tweets asociados a las diferentes temáticas.

Por su parte, para la tarea análisis de sentimiento se han generado los siguientes índices:

- **Modelado KLD del contenido del tweet (MSContent):** se ha aplicado KLD sobre la totalidad de los términos contenidos en los tweets asociados a los diferentes niveles asociados al análisis de sentimientos.
- **Modelado KLD de los adjetivos del tweet (MSAdjectives):** se ha aplicado KLD sobre la totalidad de los adjetivos identificados en los tweets asociados a los diferentes niveles asociados al análisis de sentimientos.
- **Modelado KLD del contenido del tweet con acuerdo (MSContentWithAgreement):** se ha aplicado KLD sobre la totalidad de los

¹ <http://www.daedalus.es/productos/stilus/>

términos contenidos únicamente en los tweets asociados a los diferentes niveles asociados al análisis de sentimientos, para los cuales exista acuerdo en su polaridad (información provista por la organización).

- **Modelado KLD de los adjetivos del tweet con acuerdo (MSAdjectivesWithAgreement):** se ha aplicado KLD sobre los adjetivos identificados únicamente en los tweets asociados a los diferentes niveles asociados al análisis de sentimientos, para los cuales exista acuerdo en su polaridad (información provista por la organización).

2.3 Categorización de los Tweets

Para este proceso se ha empleado un enfoque clásico de Recuperación de Información. Para ello, los modelos previamente generados se han indexado² y cada uno de los tweets del conjunto de test se ha utilizado como consulta.

El resultado devuelto por la consulta será un ranking, donde cada uno de sus elementos se corresponderá con el modelo de una de las polaridades (o temáticas). El proceso de categorización se lleva a cabo seleccionando la polaridad (o temática) asociada al primer resultado devuelto por el sistema.

Además, para la tarea del análisis de sentimiento también se ha desarrollado otro índice, el cual no utiliza los modelos generados con anterioridad:

- **Índice del Más Similar (MostSimilarIndex):** almacena cada uno de los tweets del conjunto de entrenamiento, junto con su valor de sentimiento. La idea es que dado un tweet a anotar, se buscará en este índice el tweet más similar y se anotará con el valor de sentimiento que tenga asociado.

Para un mayor detalle de los diferentes enfoques para categorizar los tweets ver la sección de Experimentos.

3 Experimentos

Para la participación de este año en el TASS, nuestro grupo ha enviado un total de 24 Experimentos, 15 para la tarea de análisis de sentimiento y 9 más para la tarea de detección

de temática. En las siguientes subsecciones cada uno de estos experimentos se explica en detalle.

3.1 Sentiment Analysis

Esta tarea está enfocada en la identificación automática de la polaridad de cada tweet. Los diferentes valores que ésta puede tomar son: Muy Positiva (P+), Positiva (P), Neutral (NEU), Ninguna (NONE), Negativa (N) y Muy Negativa (N+).

Para esta tarea se han enviado un total de 15 experimentos, detallados a continuación:

TASK1_RUN_01: Se toma el contenido del tweet y se consulta contra el índice *MostSimilarIndex*.

TASK1_RUN_02: Se toma el contenido del tweet y se consulta contra el índice que contiene el modelo *MSContentWithAgreement*.

TASK1_RUN_03: Se toma el contenido del tweet y se consulta contra el índice que contiene el modelo *MSAdjectivesWithAgreement*.

TASK1_RUN_04: Se toma el contenido del tweet y se consulta contra el índice que contiene el modelo *MSContent*.

TASK1_RUN_05: Se toma el contenido del tweet y se consulta contra el índice que contiene el modelo *MSAdjectives*.

TASK1_RUN_06: Se toma el contenido del tweet más los adjetivos identificados y se consulta contra el índice *MostSimilarIndex*.

TASK1_RUN_07: Se toma el contenido del tweet más los adjetivos identificados y se consulta contra el índice que contiene el modelo *MSContentWithAgreement*.

TASK1_RUN_08: Se toma el contenido del tweet más los adjetivos identificados y se consulta contra el índice que contiene el modelo *MSAdjectivesWithAgreement*.

TASK1_RUN_09: Se toma el contenido del tweet más los adjetivos identificados y se consulta contra el índice que contiene el modelo *MSContent*.

TASK1_RUN_10: Se toman los adjetivos identificados y se consulta contra el índice que contiene el modelo *MSAdjectives*.

TASK1_RUN_11: Se toman los adjetivos identificados y se consulta contra el índice *MostSimilarIndex*.

TASK1_RUN_12: Se toman los adjetivos identificados y se consulta contra el índice que contiene el modelo *MSContentWithAgreement*.

² La indexación de todos los modelos se ha llevado a cabo utilizando Apache Solr.

TASK1_RUN_13: Se toman los adjetivos identificados y se consulta contra el índice que contiene el modelo *MSAdjectivesWithAgreement*.

TASK1_RUN_14: Se toman los adjetivos identificados y se consulta contra el índice que contiene el modelo *MSContent*.

TASK1_RUN_15: Se toman los adjetivos identificados y se consulta contra el índice que contiene el modelo *MSAdjectives*.

3.2 Topic Classification

Esta tarea está enfocada a detectar de manera automática la temática asociada a cada uno de los tweets. Para esta tarea se han enviado un total de 9 experimentos, detallados a continuación:

TASK2_RUN_01: Se toma el contenido del tweet más las entidades nombradas detectadas y se consulta contra el índice que contiene el modelo *MTCContent*.

TASK2_RUN_02: Se toma el contenido del tweet más las entidades nombradas detectadas y se consulta contra el índice que contiene el modelo *MTHashtags*.

TASK2_RUN_03: Se toma el contenido del tweet más las entidades nombradas detectadas y se consulta contra el índice que contiene el modelo *MTNamedEntities*

TASK2_RUN_04: Se toma el contenido del tweet y se consulta contra el índice que contiene el modelo *MTCContent*.

TASK2_RUN_05: Se toma el contenido del tweet y se consulta contra el índice que contiene el modelo *MTHashtags*.

TASK2_RUN_06: Se toma el contenido del tweet y se consulta contra el índice que contiene el modelo *MTNamedEntities*

TASK2_RUN_07: Se toma el contenido del tweet más las entidades nombradas detectadas y se consulta contra el índice que contiene el modelo *MTCContent*.

TASK2_RUN_08: Se toman las entidades nombradas detectadas y se consulta contra el índice que contiene el modelo *MTHashtags*.

TASK2_RUN_09: Se toman las entidades nombradas detectadas y se consulta contra el índice que contiene el modelo *MTNamedEntities*.

4 Resultados

En esta sección se detallan los resultados obtenidos por los experimentos enviados a cada una de las tareas (ver Tabla 1 y Tabla 2).

Los resultados han sido evaluados utilizando como métrica la precisión. Nótese que, a la hora de evaluar los resultados correspondientes a la tarea de Sentiment Analysis se tienen en cuenta 5 niveles de polaridad (P+, P, NEU, N y N+) y 3 niveles (P+ y P, NEU, N+ y N) respectivamente.

4.1 Sentiment Analysis

Sentiment Analysis (5 niveles)			
Run	Precisión	Recall	F1
BEST_RUN	0.616	0.616	0.616
AVERAGE	0.433	0.433	0.433
TASK1_RUN_01	0.354	0.354	0.354
TASK1_RUN_02	0.393	0.393	0.393
TASK1_RUN_03	0.391	0.391	0.391
TASK1_RUN_04	0.398	0.398	0.398
TASK1_RUN_05	0.395	0.395	0.395
TASK1_RUN_06	0.359	0.359	0.359
TASK1_RUN_07	0.395	0.395	0.395
TASK1_RUN_08	0.391	0.391	0.391
TASK1_RUN_09	0.402	0.402	0.402
TASK1_RUN_10	0.395	0.395	0.395
TASK1_RUN_11	0.398	0.398	0.398
TASK1_RUN_12	0.391	0.391	0.391
TASK1_RUN_13	0.386	0.386	0.386
TASK1_RUN_14	0.402	0.402	0.402
TASK1_RUN_15	0.396	0.396	0.396
Sentiment Analysis (3 niveles)			
Run	Precisión	Recall	F1
BEST_RUN	0.686	0.686	0.686
AVERAGE	0.530	0.530	0.530
TASK1_RUN_01	0.471	0.471	0.471
TASK1_RUN_02	0.474	0.474	0.474
TASK1_RUN_03	0.467	0.467	0.467
TASK1_RUN_04	0.461	0.461	0.461
TASK1_RUN_05	0.454	0.454	0.454
TASK1_RUN_06	0.479	0.479	0.479
TASK1_RUN_07	0.476	0.476	0.476
TASK1_RUN_08	0.470	0.470	0.470
TASK1_RUN_09	0.464	0.464	0.464
TASK1_RUN_10	0.457	0.457	0.457
TASK1_RUN_11	0.459	0.459	0.459
TASK1_RUN_12	0.405	0.405	0.405
TASK1_RUN_13	0.407	0.407	0.407
TASK1_RUN_14	0.408	0.408	0.408
TASK1_RUN_15	0.408	0.408	0.408

Tabla 1: Resultados obtenidos para Sentiment Analysis

Los resultados para esta tarea son analizados en dos niveles: teniendo en cuenta los 5 niveles de polaridad y agrupando los niveles de polaridad en 3 (P, NEU y N).

Aunque de manera general se puede ver que los resultados son muy similares para todos los RUNS, sin embargo se pueden destacar algunos aspectos. Fijándose en los resultados teniendo en cuenta 5 niveles, atendiendo al tipo de Modelado (ver Sección 2.2.1) los mejores resultados son obtenidos por el Modelado basado en el contenido del tweet (MSContent) (runs 4, 9 y 14). Esto parece lógico, puesto que es el que mayor información utiliza y por tanto el que resulta en un modelado más representativo. Otro aspecto a destacar en estos resultados es que parece preferible hacer uso de los adjetivos (runs 5 al 15)

Sin embargo si miramos los resultados de polaridad a 3 niveles estos aspectos no quedan tan claros. Es más, fijándose en estos resultados parece que es preferible la utilización únicamente del contenido de los tweets como query, sin tener en cuenta los adjetivos.

4.2 Topic Classification

Topic Classification			
Run	Precisión	Recall	F1
BEST_RUN	0.804	0.804	0.804
AVERAGE	0.624	0.444	0.496
TASK2_RUN_01	0.660	0.404	0.501
TASK2_RUN_02	0.649	0.381	0.480
TASK2_RUN_03	0.653	0.406	0.501
TASK2_RUN_04	0.659	0.400	0.498
TASK2_RUN_05	0.639	0.366	0.465
TASK2_RUN_06	0.646	0.377	0.476
TASK2_RUN_07	0.777	0.184	0.298
TASK2_RUN_08	0.773	0.158	0.262
TASK2_RUN_09	0.727	0.211	0.327

Tabla 2: Resultados obtenidos para Topic Classification

Centrándose en los resultados de la tarea 2, se pueden extraer las siguientes consideraciones de los datos.

Centrándose en el tipo de información a utilizar como query (Contenido del tweet, Entidades Nombradas o Ambas) parece que el uso únicamente de entidades nombradas (runs 7 al 9) ofrece unos resultados inferiores a los del uso del contenido total del tweet como query (runs 4 al 6), teniendo en cuenta la medida F. A pesar de que los runs basados únicamente en entidades nombradas no ofrecen unos buenos resultados globales, sí que se obtiene una alta precisión (i.e., se clasifican pocos tweets, pero casi sin error).

Sin embargo, si se añaden las entidades nombradas detectadas al contenido de la consulta (runs 1 al 3) parece que si se consigue mejorar también en términos de medida F los resultados de utilizar únicamente el contenido sin las entidades nombradas.

Si nos centramos en el tipo de información a utilizar para generar los modelos de cada temática, los resultados no indican diferencias relevantes entre ninguna de las aproximaciones.

5 Conclusiones y Trabajo Futuro

El trabajo desarrollado para nuestra segunda participación en el TASS, ampliando el trabajo desarrollado en nuestra participación anterior, se centraba en analizar qué tipo de información relativa a los tweets es más adecuada de cara a representar los contenidos. Al igual que en nuestra pasada participación no se ha utilizado ninguna fuente de información externa y la información utilizada es únicamente aquella que aparece en los tweets.

Los resultados obtenidos por nuestra aproximación están situados en la media de los resultados globales, tanto para la tarea de Sentiment Analysis como para la tarea de Topic Classification (en esta última están incluso por encima). Estos resultados son similares a los obtenidos en nuestra última participación.

Centrándonos ya en un análisis más detallado, para la primera de las tareas resulta complicado extraer alguna conclusión debido a la escasa diferencia entre los resultados obtenidos por cada una de nuestras aproximaciones. Parece que la utilización de adjetivos es interesante de cara a representar cada una de las polaridades: sin embargo se debe hacer un trabajo más detallado a este respecto de cara a poder obtener alguna conclusión satisfactoria.

Los resultados obtenidos para la tarea de Topic Classification, sí que ofrecen alguna conclusión más clara. El principal aspecto a destacar es que la utilización de entidades nombradas presentes en los tweets es una aproximación válida siempre y cuando se utilice en conjunto con el resto de contenidos de los tweets. Como puede verse en los resultados, su utilización por separado no consigue cubrir un número de tweets suficientes (i.e. son únicamente capaces de clasificar un número reducido de tweets); aunque si bien es cierto los que son clasificados lo son con una alta precisión. Sobre la base de estos resultados una

aproximación basada únicamente en entidades nombradas podría ser utilizada como un paso previo antes de aplicar otra aproximación, la cual clasifique únicamente los tweets no cubiertos por las entidades nombradas.

Como conclusión general del trabajo de este año, podemos decir que la información adicional de los tweets representa una valiosa fuente de conocimiento; sin embargo, su utilización ha de hacerse siempre en conjunto con el contenido global de los tweets, ya que es este contenido el que realmente aporta información representativa.

Como trabajo futuro queda pendiente aglutinar las conclusiones extraídas de nuestras dos participaciones en la tarea de cara a poder concluir y demostrar experimentalmente nuestras hipótesis sobre la utilización de diferentes contenidos para modelar los tweets, tanto de cara a definir qué tipo de información utilizar (adjetivos, entidades nombradas, hashtags...), como a definir cómo debe ser utilizada esta información

Agradecimientos

Este trabajo se ha financiado con los proyectos competitivos: MA2VICMR (S2009/TIC-1542, financiado por la Comunidad de Madrid), y HOLOPEDIA (TIN 2010-21128-C02), financiado por el Ministerio de Ciencia e Innovación).

Bibliografía

Amigó, E., Carrillo de Albornoz, J., Chugur, I., Corujo, A., Gonzalo, J., Martín, T., Meij, E., de Rijke, M., Spina, D. 2013. Overview of RepLab 2013: Evaluating Online Reputation Monitoring Systems. En *Proceedings of the International Conference of the CLEF initiative (CLEF 2013)*.

Amigó, E. Corujo, A. Gonzalo, J. Meij, E. de Rijke, M. 2012. Overview of RepLab 2012: evaluating online reputation management systems. En *Proceedings of the International Conference of the CLEF initiative (CLEF 2012)*.

Benavent, X. Castellanos, A. de Ves, E. Hernández-Aranda, D. Granados, R. García-Serrano, A. 2013. A multimedia IR-based system for the Photo Annotation Task at ImageCLEF2013. En *Working Notes of the CLEF 2013*.

Blei, D. y Lafferty, J. 2009. Topic models. *Text Mining: Theory and Application*.

Castellanos, A. 2013. Recomendación de Contenidos Digitales basada en divergencias del lenguaje: Diseño, Experimentación y Evaluación. *Master Thesis, UNED*.

Castellanos, A. Cigarrán, J. García-Serrano, A. 2012. Generación de un corpus de usuarios basado en divergencias del lenguaje. En *Proceedings of the Segundo Congreso Español de Recuperación de In-formación (CERI 2012)*.

Castellanos, A. Cigarrán, J. García-Serrano, A. 2012. UNED @ TASS: Using IR techniques for topic-based sentiment analysis through divergence models. En *Proceedings of the Workshop on Sentiment Analysis at SEPLN (TASS 2012)*.

Castellanos, A. Cigarrán, J. García-Serrano, A. 2013. Modelling Techniques for Twitter Contents: A step beyond classification based approaches. En *Working Notes of the CLEF 2013*.

Hernández-Aranda, D. Granados, R. García-Serrano, A. 2012. Servicios de anotación y búsqueda para corpus multimedia. *Revista de la Sociedad Española para el Procesamiento del Lenguaje Natural, SEPLN (49)*.

Kullback, S. Leibler. R.A. 1951. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1): 79-86.

Liu, B. Zhang, L. 2012. A survey of opinion mining and sentiment analysis. En *Mining Text Data*, páginas 415-463. Springer US.

Xia, R. Zong, C. Li, S. 2011. Ensemble of feature sets and classification algorithms for sentiment classification. *Information Sciences (181)*: 1138-1152.

Xu, K. Shaoyi-Liao, S. Li, J. Song, Y. 2011. Mining comparative opinions from customer reviews for Competitive Intelligence. *Decision Support Systems (50)*: 743-754.

Zhang, Z. Ye, Q. Zhang, Z. Li, Y. 2011. Sentiment classification of Internet restaurant reviews written in Cantonese. *Expert Systems with Applications*.