

# Caracterización de una entidad basada en opiniones: un estudio de caso\*

## *Opinion-based entity profiling: a case study*

Damiano Spina, Enrique Amigó, Julio Gonzalo

Grupo de PLN y RI de la UNED

C/Juan de Rosal, 16.

28040 Madrid, España

{damiano,enrique,julio}@lsi.uned.es

Bernardino Beotas

Grupo ALMA

C/Valentín Beato, 23.

28037 Madrid, España

b.beotas@almatech.es

**Resumen:** En este artículo introducimos la tarea de *caracterización de una entidad* a través de las *valoraciones* que sobre ella se vierten en la red, entendida como la agrupación de valoraciones (positivas y negativas) equivalentes o muy similares. Se trata de un problema práctico que se encuentra en el ámbito de la gestión de la reputación online, para el que todavía no existen tareas ni benchmarks de referencia dentro del Procesamiento del Lenguaje Natural. A través de un estudio de caso se muestran las dificultades intrínsecas de la tarea y se proporciona un baseline basado en el algoritmo de clustering Lingo.

**Palabras clave:** minería de opiniones, gestión de la reputación online, resumen de opiniones, agrupamiento

**Abstract:** We introduce the *opinion-based entity profiling* task, understood as the automatic grouping of equivalent or very similar online opinions (both positive and negative) about an entity. This is a highly practical problem that arises in the context of online reputation management, but it has not been directly addressed by any Natural Language Processing evaluation campaign or specific benchmarking corpus. We use a case study to illustrate the challenges of the task, and we provide an initial baseline based on the Lingo clustering algorithm.

**Keywords:** opinion mining, online reputation management, opinion summarization, clustering

## 1. Introducción

Las personas comparten sus opiniones sobre productos, personas y organizaciones en distintos sitios de la web como foros, blogs, redes sociales y sitios de valoraciones (Pollach, 2006; Dellarocas, Awad, y Zhang, 2004). Monitorizar este tipo de información es crucial para gestionar la reputación de una entidad. El proceso de Gestión de Reputación Online (Online Reputation Management, ORM) consiste en monitorizar los medios, detectar contenido relevante, analizar qué se dice de la entidad y participar en las conversaciones identificadas, interactuando con los consumidores.

Conocer cuáles son los comentarios (positivos y negativos) más recurrentes de los usuarios es esencial para una buena gestión

de la reputación, ya que reflejan la imagen que los usuarios tienen de la entidad. A esta tarea la hemos denominado *caracterización (profiling)* de una entidad basada en opiniones. Curiosamente, es una tarea para la que todavía no existen benchmarks específicos de evaluación en el área de las Tecnologías de la Lengua. La tarea más próxima es el "Opinion Question Answering and Summarization Task" (Dang, 2008) del TAC, en la que hay que resumir las opiniones de los usuarios sobre una entidad determinada; sin embargo, una diferencia sustancial es que en éste último caso no es necesario tener en cuenta las frecuencias con las que se emite cada juicio de valor. En la tarea que nos ocupa, sin embargo, éste es un dato fundamental, y por tanto se requiere algún proceso de agrupación de opiniones de contenido similar.

En este artículo queremos perfilar el problema y estudiar el tipo de retos que plan-

\* Este trabajo se enmarca dentro del proyecto WebOpinion, financiado por el Ministerio de Industria y Alma Technologies.

tea desde el punto de vista de su resolución automática. Nótese que realizar un profiling de forma manual es bastante costoso: según nuestras estimaciones, una persona puede tardar alrededor de cinco días en analizar mil opiniones.

Con el fin de analizar la dificultad de la tarea de forma cualitativa y ayudar a la comunidad a entender el problema, se ha realizado un estudio preliminar sobre una entidad, la Universidad Nacional de Educación a Distancia (UNED). Por un lado, se han agrupado manualmente las valoraciones más comentadas en las ventajas (valoraciones positivas) y las desventajas (valoraciones negativas) que aparecen en las opiniones de un sitio de valoraciones. En segundo lugar, se han analizado los resultados para estimar la complejidad de abordar la tarea de forma automática.

El artículo está organizado de la siguiente manera: la sección 2 resume el estado del arte. A continuación se muestra el análisis manual llevado a cabo para realizar la tarea de caracterización basado en opiniones de la UNED. En la sección 4 se analizan las tecnologías de la lengua que se podrían aplicar para automatizar la tarea y cómo se podría evaluar. La sección 5 muestra en qué casos se podría aplicar WordNet Castellano para mejorar los resultados. Finalmente, en la sección 6 se discuten los principales problemas encontrados y se presenta el trabajo futuro.

## 2. Estado del arte

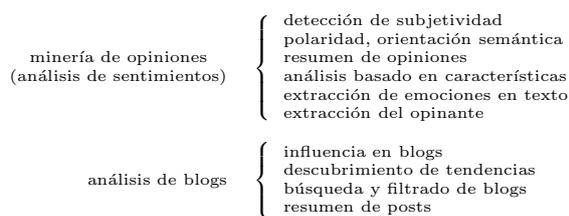


Figura 1: Principales ramas de investigación en minería de opiniones y análisis de blogs.

Dado que se trata de un problema emergente, aunque los experimentos se centran en un aspecto concreto de la tarea, abordamos el estado del arte desde una perspectiva global para una mejor contextualización del mismo.

Nuestro trabajo está relacionado con dos campos de investigación que se han ido consolidando en los últimos años, esquematizados en la figura 1. El primero investiga pro-

blemas relacionados con el análisis de opiniones e información subjetiva (Pang y Lee, 2008). El segundo campo se centra en investigar aspectos relacionados con la blogosfera (Agarwal y Liu, 2008). Respecto al primero, uno de los problemas que se tratan en la minería de opiniones (*opinion mining*) consiste en la detección de opiniones o de subjetividad en textos (Wiebe, Wilson, y Cardie, 2005; Kim y Hovy, 2005a; Riloff, Wiebe, y Phillips, 2005). Otra cuestión consiste en el análisis de la polaridad u orientación semántica de las opiniones (Turney, 2002; Esuli y Sebastiani, 2005; Pang, Lee, y Vaithyanathan, 2002). En el TREC Blog Track (Macdonald y Soboroff, 2008) se evalúan sistemas que atacan estos dos problemas. También se estudia el análisis de opiniones desde el punto de vista multilingüe (Seki et al., 2008; Mihalcea, Banea, y Wiebe, 2007; Balahur y Montoyo, 2008). Denecke (2008) traduce automáticamente al inglés valoraciones en alemán y utiliza el recurso SentiWordNet (Esuli y Sebastiani, 2006) para extraer la polaridad de las mismas. Otros trabajos relacionados con el análisis de sentimientos se centran en la extracción de emociones a partir del texto (Strapparava y Mihalcea, 2008; Strapparava y Mihalcea, 2007; Balahur, de Correos, y Montoyo, 2008) usando WordNet-Affect (Strapparava y Valitutti, 2004)...

Por otro lado, Kim y Hovy (2005b) estudian la extracción del opinante (*opinion-holder*) usando un algoritmo de máxima entropía entrenado con datos anotados manualmente para aplicarlo a tareas de búsqueda de respuesta (QA) de opiniones.

Con respecto al análisis de la blogosfera, hay trabajos que proponen técnicas para descubrir las tendencias y los temas más candentes discutidos en los blogs (Glance, Hurst, y Tomokiyo, 2004; Platakis, Kotsakos, y Gunopulos, 2008). Otro problema consiste en medir la influencia de los blogs (Gill, 2004; Nallapati y Cohen, 2008; Java et al., 2006; Kalle, 2007; Marlow, 2004). Además, se estudia la búsqueda y el filtrado de blogs (Mishne y de Rijke, 2006; Mishne, Carmel, y Lempel, 2005).

Por otra parte, Hu y Liu (2004) proponen la tarea de agregar las opiniones de productos que se encuentran en sitios de valoraciones a nivel de atributos e introducen un corpus compuesto por valoraciones de cinco productos electrónicos descargados desde Amazon

y Cnet. Por ejemplo, en el caso de cámaras digitales el sistema debe detectar valoraciones positivas o negativas para cada uno de sus atributos principales: resolución, calidad de la óptica, duración de la batería, etc. Usan información morfo-sintáctica (POS) para extraer los nombres y los sintagmas nominales más frecuentes como características del producto. Este tipo de extracción puede ser suficiente para una amplia gama de productos; sin embargo, cuando trabajamos con otros tipos de entidades, como organizaciones y personas, las opiniones de los usuarios no son fácilmente reducibles a su polaridad en términos de atributos. En el caso de la UNED, por ejemplo, “estudiar a tu ritmo” o “es más difícil que una universidad presencial” son opiniones que nos interesa conocer explícitamente: es decir, en un informe sobre la UNED queremos saber que una de las opiniones es que es más difícil que las universidades convencionales, y cuan generalizada está esa opinión.

Liu, Hu, y Cheng (2005) describen un prototipo que compara valoraciones de diferentes productos. Éste trata de detectar automáticamente los atributos relevantes para los productos que se valoran, usando técnicas de reglas de asociación y WordNet para extraer sinónimos de las características. Aunque se obtienen buenos resultados para valoraciones de productos, no es aplicable a las relaciones semánticas más complejas que se establecen al comentar aspectos de organizaciones o personas (p.e. “No presencial”, “Es a distancia” al valorar un programa de estudios).

Gamon et al. (2005) aplican algoritmos de clustering para agrupar opiniones, obteniendo resultados muy prometedores desde un punto de vista cualitativo. Éste es el tipo de trabajos más cercanos a nuestro interés, aunque en el artículo no se hace una evaluación sistemática del enfoque propuesto.

Otros trabajos (Ding, Liu, y Yu, 2008; Popescu y Etzioni, 2005; Carenini, Ng, y Zwart, 2005) y herramientas como *Swotti*<sup>1</sup> clasifican las opiniones de productos según una taxonomía predefinida de características.

El problema que tratamos de definir aquí está relacionado sobre todo con la extracción de características y la generación de resúmenes de opiniones. La experimentación descrita en este artículo aprovecha la infor-

mación estructurada (comentarios separados en ventajas y desventajas) que ofrecen los sitios de valoraciones. Esto permite obviar la detección de subjetividad y el análisis de la polaridad de las opiniones. Sin embargo, esta tarea se puede llevar a cabo tratando las opiniones que aparecen en foros, blogs y medios online, teniendo en cuenta que en este caso es necesario llevar a cabo un análisis previo para extraer las opiniones y la polaridad de las mismas.

### 3. Estudio de caso

Con el fin de analizar de forma cualitativa la tarea de caracterización de una entidad basada en opiniones, hemos abordado un estudio de caso sobre la Universidad Nacional de Educación a Distancia (UNED).

En este artículo nos centramos en las opiniones recuperadas del sitio de valoraciones *Ciao!*<sup>2</sup>, por dos motivos: (i) más del 50 % de las opiniones recolectadas en el estudio pertenecen a sitios de valoraciones (256 frente a un total de 384); y (ii) los comentarios separados en ventajas y desventajas, permiten obviar la detección de subjetividad en la información y el análisis de la polaridad de las opiniones.

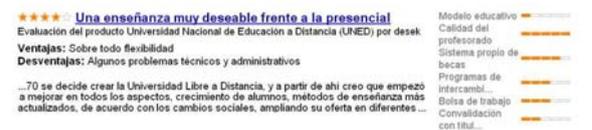


Figura 2: Una opinión sobre la UNED en Ciao!

Cada opinión tiene un título, un autor, un conjunto de características predefinidas de la categoría de la entidad (modelo educativo, calidad del profesorado, sistema propio de becas, programas de intercambio con univ. extranjeras...), un texto libre y valoraciones separadas en ventajas y desventajas. Agrupar manualmente las ventajas y desventajas de las 206 opiniones ha costado 8 personas-hora. Podríamos concluir que una persona tardaría unos cinco días en realizar esta tarea sobre 1000 opiniones.

Las figuras 3 y 4 muestran, respectivamente, las ventajas y desventajas más frecuentes.

En 227 ventajas se han identificado un total de 31 grupos distintos, de los cuales 17 están compuestos por menos de tres valoraciones. Entre las valoraciones menos fre-

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

<sup>2</sup><http://www.ciao.es>



Figura 3: Ventajas más comentadas sobre la UNED en ciao.es



Figura 4: Desventajas más comentadas sobre la UNED en ciao.es

cuentas (no recogidas individualmente en la gráfica), se encuentran “biblioteca”, “aumento de curriculum”, “todo está en los libros no hay profesores inútiles que aguantar”, “página web”, “exámenes normales”, “sobre todo el sistema en sí”, “foros”, “la posibilidad de estudiar a cualquier edad”... El 26 % de los grupos identificados recogen el 80 % de las valoraciones. Finalmente, el 13 % de las valoraciones están compuestas sólo por palabras vacías (“todas”, “muchas”, “no hay”...).

En un total de 213 desventajas se han identificado 39 grupos, de los cuales 19 tienen menos de 3 valoraciones. Algunas de las valoraciones menos comentadas son: “plazos”, “material”, “pocas ayudas”, “biblioteca penosa”, “tardan mucho en darte las calificaciones”. El 36 % de los grupos recogen el 80 % de las valoraciones. Las valoraciones compuestas únicamente por palabras vacías representan el 9 % de las desventajas.

Los cuadros 1 y 2 muestran algunas de las

Aspecto	Expresiones
flexibilidad de horarios	libertad de horarios, estudias a tu ritmo, adaptabilidad a tiempos, gestionas tu propio tiempo, tu marcas el ritmo, vas a tu aire
no tener que asistir a clase	estudias en casa, no hay que ir a clase, no requiere desplazarse, no presencial
se puede estudiar y trabajar a la vez	poder estudiar si trabajas, compatibilizar los estudios con un trabajo, compatible con otras ocupaciones, única opción para aquellos que tienen que currar, la posibilidad de compaginar trabajo y estudio, puedes trabajar y estudiar a la vez

Cuadro 1: Ejemplos de expresiones que componen las ventajas más comentadas por los usuarios

expresiones más frecuentes en las valoraciones. Estas expresiones ilustran la dificultad del problema: se trata de expresiones cortas en las que se puede estar comentando un mismo aspecto incluso sin usar el mismo vocabulario y con una relación semántica compleja entre opiniones equivalentes (p.e. “más difícil que la presencial” y “el sacrificio que supone”).

Aspecto	Expresiones
estudiar por cuenta propia	has de ser muy constante, algunas veces te sientes solo, necesidad de mucha voluntad, soledad, saber organizarse, tienes que echarle muchas ganas, te puedes desanimar estudiando solo, sentimiento de dejadez, debe de gustarte hablarte a ti mismo
más difícil que la presencial	es mucho más difícil, las dificultades para el alumno son altas, nivel más alto que la presencial, muy dura, requiere más esfuerzo, el sacrificio que supone
precio	gasto económico, podían bajar los precios, qué cara, demasiado cara

Cuadro 2: Ejemplos de expresiones equivalentes entre las desventajas más comentadas por los usuarios

Las valoraciones más comentadas por los usuarios reflejan la imagen que los usuarios

tienen de una entidad, por lo tanto son esenciales para gestionar la reputación de la misma.

#### 4. *Automatización de la caracterización de una entidad basada en opiniones*

Con el fin de estimar la complejidad de automatizar la tarea de profiling de una entidad basada en opiniones, hemos adaptado al castellano un algoritmo de clustering de uso extendido que además está especialmente diseñado para documentos muy cortos. El algoritmo de clustering Lingo (Osinski, Stefanowski, y Weiss, 2004), implementado en el proyecto Carrot<sup>3</sup>, está pensado para agrupar resultados de un motor de búsqueda (a través de los snippets correspondientes a cada URL), pero también se puede aplicar para automatizar el profiling de una entidad basada en opiniones.

Una vez realizado el preprocesamiento (identificación de idioma, filtro de palabras vacías y stemming), Lingo induce los descriptores de los clusters usando Singular Value Decomposition y Latent Semantic Indexing (Deerwester et al., 1990) para luego asignar los documentos a los clusters usando Vector Space Model.

La adaptación al castellano ha consistido en extender la lista de palabras vacías, que por defecto es de 146 palabras, a 446 palabras, y en incluir una fase de eliminación de tildes en el preprocesamiento. Además, se ha añadido una penalización para los descriptores compuestos por un sólo término.

##### 4.1. Configuración del experimento

Las ventajas y desventajas de las 206 opiniones se han considerado de forma independiente. Los umbrales del algoritmo se han entrenado usando los dos conjuntos de forma cruzada (entrenando ventajas para evaluar desventajas y viceversa). Los valores de los umbrales obtenidos por los dos entrenamientos son los mismos, excepto para el umbral que controla el número de clusters creados por el algoritmo, que varía en dos centésimas.

La evaluación de los dos clustering se ha realizado usando dos pares de métricas: pureza y pureza inversa (por ser las más utilizadas) y precisión, cobertura BCubed por

ser formalmente las más adecuadas (Amigó et al., 2009). Se ha calculando la medida  $F$  sobre ambos pares con  $\alpha = 0,5$  (es decir, haciendo la media armónica de las dos medidas individuales). Se han optimizado los umbrales del algoritmo para maximizar los valores de medida  $F$  sobre las métricas BCubed.

##### 4.2. Resultados globales

El cuadro 3 muestra los resultados para el clustering de ventajas y el clustering de desventajas realizado por Lingo comparado con los baseline “uno en uno” (cada documento forma un cluster) y “todos en uno” (todos los documentos se agrupan en un único cluster). Los resultados del baseline “uno en uno” indican que las ventajas están más agrupadas que las desventajas. Los resultados de Lingo mejoran la calidad de los clusters con respecto a ambos baseline. Sin embargo, sigue habiendo un gran margen de mejora.

La evaluación de agrupaciones es de por sí una tarea compleja (Rosenberg y Hirschberg, 2007), y sugiere un estudio en más detalle de los resultados.

##### 4.3. Análisis detallado de resultados

Para obtener conclusiones más concretas, se han analizado los clusters asociados a las valoraciones más comentadas (aquellas con frecuencia mayor que 6).

La mayoría de los errores de precisión son debidos a palabras que no se filtran en el preprocesamiento como “permite” y “puedes” o por palabras compartidas por valoraciones distintas como “estudiar a distancia” y “hay que estudiar más”. Curiosamente, añadir estos términos a las stopwords empeora los resultados; entre otras cosas, porque resulta en muchas expresiones que se componen sólo por palabras vacías (“no hay”, “muchas”, “ninguna”...) y que, por tanto, el algoritmo no consigue agrupar correctamente.

Con respecto a la cobertura, el resultado global tiene un sesgo positivo debido a los clusters que asocian aspectos menos comentados, la mayoría de éstos compuestos por un solo comentario). Así, el 83% de los clusters más significativos, es decir, los asociados a los aspectos más comentados, tienen cobertura menor que 0.3, mientras que la cobertura media es del 0.5. Esto significa que los aspectos más comentados tienden a dispersarse en va-

<sup>3</sup><http://project.carrot2.org/>

		Pureza	Pureza Inversa	$\mathcal{F}(P, PI)$	Precisión BCubed	Cobertura BCubed	$F$ BCubed
Ventajas	Uno en uno	1.00	0.14	0.25	1.00	0.14	0.25
	Todos en uno	0.25	1.00	0.40	0.12	1.00	0.22
	Lingo	0.59	0.59	<b>0.59</b>	0.46	0.50	<b>0.48</b>
Desventajas	Uno en uno	1.00	0.18	0.31	1.00	0.18	0.31
	Todos en uno	0.16	1.00	0.27	0.07	1.00	0.13
	Lingo	0.48	0.61	<b>0.54</b>	0.34	0.50	<b>0.40</b>

Cuadro 3: Calidad de los clustering de ventajas y de desventajas

rios clusters. Por ejemplo, la ventaja “No tener que asistir a clase”, que se repite en 42 valoraciones, se agrupa en 5 clusters distintos caracterizados por las expresiones: “No hay que ir a clase”, “No tienes que desplazarte”, “No requiere presencia”, “Comodidad” y “No acudir a clases”.

### 5. ¿Puede Wordnet contribuir a solucionar el problema?

Queremos evaluar si el uso de recursos como WordNet Castellano (Atserias, Villarejo, y Rigau, 2004) pueden mejorar la calidad del clustering, al menos de manera inmediata con la información explícita que proporcionan. Por ello, se han elegido 10 ejemplos representativos del problema de dispersión y se han buscado manualmente posibles nexos entre términos en la ontología por medio de la interfaz web del Repositorio Central Multilingüe (MCR, Multilingual Central Repository)<sup>4</sup>.

El cuadro 4 recoge los ejemplos de dispersión que se podrían resolver con el uso de WordNet Castellano 1.6.

El synset  $\{ir_2, desplazarse_1\}$  permitiría asociar las dos expresiones del ejemplo 1. Las expresiones del ejemplo 2 se podrían agrupar en un mismo cluster teniendo en cuenta el siguiente camino: el synset  $\{presencia_4\}$  tiene como hiperónimo  $\{proximidad_3\}$ . El synset  $\{próximo_6, cercano_1\}$  es sinónimo cercano (*near synonym*) de  $\{cercano_2\}$ , que es un estado (relación *be in state*) de  $\{distancia_1\}$ . Las expresiones que aparecen en el ejemplo 3 se podrían asociar mediante el synset  $\{ocupación_5, empleo_2, trabajo_{11}\}$ . Las palabras “difícil” y “esfuerzo” del ejemplo 4 también están relacionadas: el synset  $\{esfuerzo_1, afán_1, dedicación_1, desvelo_1\}$  tiene 11 hipónimos entre los cuales se encuentra el

synset  $\{dificultad_2\}$ . El synset  $\{dificultad_6\}$  tiene como estado  $\{difícil_1, duro_3\}$ . El synset  $\{caro_1, costoso_2\}$  está relacionado con el synset  $\{coste_1, costo_1\}$ , que tiene como hipónimo al synset  $\{importe_1, precio_1\}$ . El camino mínimo entre las palabras “precio” y “gasto” podría ser el siguiente: el synset  $\{importe_1, precio_1\}$  es hipónimo de  $\{coste_1, costo_1\}$  que a su vez es hipónimo de  $\{gastos_1, gasto_1, desembolso_2\}$ .

Existen herramientas computacionales que calculan la similitud y la relación entre dos conceptos en WordNet (Pedersen, Patwardhan, y Michelizzi, 2004; Agirre y Rigau, 1996). Por otro lado, Hotho, Staab, y Stumme (2003) han demostrado que el uso de WordNet puede mejorar los resultados de clustering de documentos. La idea consiste en extender los vectores de términos de la matriz término-documento con conceptos de WordNet que aparecen en el documento. Para ello proponen dos estrategias de desambiguación: La primera consiste en considerar el sentido de WordNet más común, y la segunda consiste en desambiguar el sentido del término considerando el documento como contexto. Además, usan las relaciones de hiperonimia de los conceptos que aparecen en el documento para modificar la frecuencia en la que aparecen términos de más alto nivel.

Nótese que hay casos en los que es muy difícil encontrar un camino entre dos términos. El cuadro 5 muestra algunos ejemplos en los que no se han encontrado caminos cortos de forma manual.

### 6. Discusión y trabajo futuro

La caracterización (profiling) de una entidad basada en las opiniones online que se vierten sobre ella es una tarea relevante desde un punto de vista práctico y costosa de realizar manualmente. Los aspectos más comentados en las ventajas y desventajas de valoraciones sobre la entidad son aquellos que un

<sup>4</sup><http://garraf.epsevg.upc.es/cgi-bin/wei4/public/wei.consult.perl>

Ejemplo	Expresión 1	Expresión 2
1	No hay que ir a clase	No requiere desplazarse
2	No es presencial	Es a distancia
3	Compatible con otras ocupaciones	La posibilidad de compaginar trabajo y estudio
4	Muy difícil	Requiere mucho esfuerzo
5	Precio	Es cara
6	Precio	Gasto económico

Cuadro 4: Problemas de dispersión (*joining gaps*) que se podrían resolver con WordNet Castellano

Ejemplo	Expresión 1	Expresión 2
7	Estudiar a tu ritmo	Gestionas tu propio tiempo
8	Gestionas tu propio tiempo	Libertad de horarios
9	Adquisición efectiva de conocimientos	Se aprende de verdad
10	Soledad	Debe de gustarte hablarte a ti mismo

Cuadro 5: Ejemplos de joining gaps que no pueden conectarse mediante EuroWordNet

gestor de reputación de la entidad debería de tener en cuenta de forma prioritaria, ya que reflejan la imagen que los usuarios tienen de la misma. En este artículo hemos comprobado que la resolución automática de este problema no es trivial, puesto que las valoraciones suelen ser expresiones cortas difíciles de relacionar semánticamente mediante técnicas convencionales de indexación.

Hasta ahora sólo hemos abordado un estudio de caso sobre el que no se pueden extraer conclusiones cuantitativas generalizables. Nuestro trabajo a corto plazo incluye la creación de un corpus de opiniones sobre entidades de distintos dominios, incluyendo otras fuentes de información como foros, blogs y redes sociales. Este tipo de corpus nos permitirá realizar una evaluación comparativa de distintos algoritmos de clustering. También queremos estudiar el uso de otras bases de conocimiento como eXtended WordNet<sup>5</sup>, WordNet Domains (Bentivogli et al., 2004) o SUMO (Pease, Niles, y Li, 2002) y el uso de algoritmos de desambiguación como el presentado en Agirre y Soroa (2009). Por último, otra línea que pretendemos abordar consiste en estudiar la agrupación de opiniones considerando la variable temporal, para analizar cómo varía lo que se dice de una entidad en función del tiempo.

### Bibliografía

Agarwal, N. y H. Liu. 2008. Blogosphere: research issues, tools, and applications.

<sup>5</sup><http://xwn.hlt.utdallas.edu/>

Agirre, E. y G. Rigau. 1996. Word sense disambiguation using conceptual density. En *Proceedings of COLING*, volumen 96, páginas 16–22.

Agirre, E. y A. Soroa. 2009. Personalizing pagerank for word sense disambiguation. En *Proceedings of EACL-09*, Athens, Greece.

Amigó, E., J. Gonzalo, J. Artiles, y F. Verdejo. 2009. A comparison of extrinsic clustering evaluation metrics based on formal constraints. *Information Retrieval*, páginas 1–26.

Atserias, J., L. Villarejo, y G. Rigau. 2004. Spanish wordnet 1.6: Porting the spanish wordnet across princeton versions. En *Proceedings of the 4th International Conference on Language Resources and Evaluation*.

Balahur, A., A. de Correos, y A. Montoyo. 2008. Applying a culture dependent emotion triggers database for text valence and emotion classification. *Procesamiento del Lenguaje Natural*, 40:107–114.

Balahur, A. y A. Montoyo. 2008. Multilingual Feature-Driven Opinion Extraction and Summarization from Customer Reviews. En *Proceedings of the 13th international conference on Natural Language and Information Systems: Applications of Natural Language to Information Systems*, páginas 345–346.

- Bentivogli, L., P. Forner, B. Magnini, y E. Pianta. 2004. Revising wordnet domains hierarchy: Semantics, coverage, and balancing. En *Proceedings of COLING Workshop on Multilingual Linguistic Resources*.
- Carenini, G., R.T. Ng, y E. Zwart. 2005. Extracting knowledge from evaluative text. En *Proceedings of the 3rd International Conference on Knowledge Capture*.
- Dang, H.T. 2008. Overview of the TAC 2008 Opinion Question Answering and Summarization Tasks. En *TAC 2008 Workshop Notes*.
- Deerwester, S., S.T. Dumais, G.W. Furnas, T.K. Landauer, y R. Harshman. 1990. Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6):391–407.
- Dellarocas, C., N. Awad, y X. Zhang. 2004. Exploring the value of online reviews to organizations: Implications for revenue forecasting and planning. En *Proceedings of the International Conference on Information Systems*.
- Denecke, K. 2008. Using SentiWordNet for multilingual sentiment analysis. En *IEEE 24th International Conference on Data Engineering Workshop*, páginas 507–512.
- Ding, X., B. Liu, y P.S. Yu. 2008. A holistic lexicon-based approach to opinion mining. En *Proceedings of the international conference on Web search and web data mining*, páginas 231–240.
- Esuli, A. y F. Sebastiani. 2005. Determining the semantic orientation of terms through gloss analysis. En *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, páginas 617–624.
- Esuli, A. y F. Sebastiani. 2006. SentiWordNet: A publicly available lexical resource for opinion mining. En *In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC'06)*, páginas 417–422.
- Gamon, M., A. Aue, S. Corston-Oliver, y E. Ringger. 2005. Pulse: Mining customer opinions from free text. *Lecture Notes in Computer Science*, 3646:121–132.
- Gill, K.E. 2004. How can we measure the influence of the blogosphere. En *WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis, and Dynamics*.
- Glance, N., M. Hurst, y T. Tomokiyo. 2004. BlogPulse: Automated trend discovery for weblogs. En *WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics*.
- Hotho, A., S. Staab, y G. Stumme. 2003. Wordnet improves text document clustering. En *Proceedings of the SIGIR 2003 Semantic Web Workshop*, páginas 541–544.
- Hu, M. y B. Liu. 2004. Mining and summarizing customer reviews. En *Proceedings of the 10th ACM SIGKDD international conference on Knowledge discovery and data mining*, páginas 168–177.
- Java, A., P. Kolari, T. Finin, y T. Oates. 2006. Modeling the spread of influence on the blogosphere. En *Proceedings of the 15th International World Wide Web Conference*.
- Kale, A. 2007. *Modeling trust and influence on blogosphere using link polarity*. Ph.D. tesis, University of Maryland.
- Kim, S.M. y E. Hovy. 2005a. Automatic detection of opinion bearing words and sentences. En *Proceedings of International Joint Conference on Natural Language Processing*, volumen 5.
- Kim, S.M. y E. Hovy. 2005b. Identifying opinion holders for question answering in opinion texts. En *Proceedings of AAAI Workshop on Question Answering in Restricted Domains*.
- Liu, B., M. Hu, y J. Cheng. 2005. Opinion observer: Analyzing and comparing opinions on the web. En *Proceedings of the 14th international conference on World Wide Web*, páginas 342–351.
- Macdonald, C. y I. Soboroff. 2008. Overview of the TREC-2008 Blog Track. En *Proceedings of the 17th Text REtrieval Conference (TREC 2008)*.
- Marlow, C. 2004. Audience, structure and authority in the weblog community. En *International Communication Association Conference*, páginas 1–9.

- Mihalcea, R., C. Banea, y J. Wiebe. 2007. Learning multilingual subjective language via cross-lingual projections. En *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, página 976.
- Mishne, G., D. Carmel, y R. Lempel. 2005. Blocking blog spam with language model disagreement. En *Proceedings of the First International Workshop on Adversarial Information Retrieval on the Web (AIRWeb)*.
- Mishne, G. y M. de Rijke. 2006. A study of blog search. *Lecture Notes in Computer Science*, 3936:289.
- Nallapati, R. y W. Cohen. 2008. Link-PLSA-LDA: A new unsupervised model for topics and influence of blogs. En *International Conference for Weblogs and Social Media*.
- Osinski, S., J. Stefanowski, y D. Weiss. 2004. Lingo: Search results clustering algorithm based on singular value decomposition. En *Proceedings of the International IIS: Intelligent Information Processing And Web Mining*, página 359.
- Pang, B. y L. Lee. 2008. *Opinion mining and sentiment analysis*. Now Publishers.
- Pang, B., L. Lee, y S. Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. En *Proceedings of the ACL'02 conference on Empirical methods in natural language processing*, volumen 10, páginas 79–86.
- Pease, A., I. Niles, y J. Li. 2002. The Suggested Upper Merged Ontology: a large ontology for the semantic web and its applications. En *Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web*, volumen 28.
- Pedersen, T., S. Patwardhan, y J. Michelizzi. 2004. Wordnet::similarity-measuring the relatedness of concepts. En *Proceedings of the National Conference on Artificial Intelligence*, páginas 1024–1025.
- Platakis, M., D. Kotsakos, y D. Gunopulos. 2008. Discovering Hot Topics in the Blogosphere. páginas 122–132.
- Pollach, I. 2006. Electronic word of mouth: A genre analysis of product reviews on consumer opinion web sites. En *Proceedings of the 39th Annual Hawaii International Conference on System Sciences*, volumen 3.
- Popescu, A.M. y O. Etzioni. 2005. Extracting product features and opinions from reviews. En *Proceedings of HLT/EMNLP*, volumen 5, páginas 339–346.
- Riloff, E., J. Wiebe, y W. Phillips. 2005. Exploiting subjectivity classification to improve information extraction. En *Proceedings of the National Conference on Artificial Intelligence*, volumen 20, página 1106.
- Rosenberg, A. y J. Hirschberg. 2007. V-measure: A conditional entropy-based external cluster evaluation measure. En *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*.
- Seki, Y., D.K. Evans, L.W. Ku, L. Sun, H.H. Chen, y N. Kando. 2008. Overview of multilingual opinion analysis task at NTCIR-7. En *Proceedings of the 7th NTCIR Workshop Meeting on Evaluation of Information Access Technologies*.
- Strapparava, C. y R. Mihalcea. 2007. Semeval-2007 task 14: Affective text. En *Proceedings of SemEval-2007*.
- Strapparava, C. y R. Mihalcea. 2008. Learning to identify emotions in text. En *Proceedings of the 2008 ACM symposium on Applied computing*, páginas 1556–1560.
- Strapparava, C. y A. Valitutti. 2004. WordNet-Affect: an affective extension of WordNet. En *Proceedings of LREC*, volumen 4, páginas 1083–1086.
- Turney, P.D. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. En *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, páginas 417–424.
- Wiebe, J., T. Wilson, y C. Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2):165–210.