



Polibits

ISSN: 1870-9044

polibits@nlp.cic.ipn.mx

Instituto Politécnico Nacional

México

Vallés Balague, Enrique; Rosso, Paolo; Locoro, Angela; Mascardi, Viviana

Análisis de Opiniones con Ontologías

Polibits, vol. 41, 2010

Instituto Politécnico Nacional

Distrito Federal, México

Disponible en: <http://www.redalyc.org/articulo.oa?id=402640454005>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

Análisis de Opiniones con Ontologías

Enrique Vallés Balaguer, Paolo Rosso, Angela Locoro y Viviana Mascardi

Resumen—En este artículo presentamos un trabajo sobre análisis de opiniones llevado a cabo gracias a un enfoque innovador basado en fusión de ontologías. El objetivo de este trabajo es permitir que dos empresas puedan intercambiar y compartir los resultados de los análisis de las opiniones de sus productos y servicios.

Palabras clave—Minería de opiniones, mapeo y fusión de ontologías, Web 2.0, Empresa 2.0.

Opinion Mining using Ontologies

Abstract—In this paper we present a work dealing with opinion analysis carried out thanks to an innovative approach based on ontology matching. The aim of this work is to allow two enterprises to share and merge the results of opinion analyses on their own products and services.

Index Terms—Opinion mining, ontology matching and merging, Web 2.0, Enterprise 2.0.

I. INTRODUCCIÓN

PARA una pequeña y mediana empresa (empresa) tanto la cantidad como la calidad de información no tiene precio. El principal objetivo de la información es la de apoyar a la toma de decisiones, puesto que con ella se tendrán más bases sustentables para poder decidir cuáles son los pasos a seguir y qué rumbo hay que tomar para lograr los objetivos que se planificaron. Es por ello que en una empresa se le debe de poner una atención sumamente especial a la información que se genera cada día, la adecuada interpretación de ésta establecerá los cimientos necesarios para consolidarse como una Empresa 2.0 de éxito en el mercado. La información le permitirá identificar cuáles son las fortalezas con las que cuenta y cuáles las debilidades y sectores vulnerables que presenta como organización. Teniendo en cuenta estos datos podrá tener una planificación más alcanzable y factible, ya que podrá identificar donde tiene que aumentar los esfuerzos y que parte de la empresa necesita mayor atención. Hoy en día, la principal fuente de información para conocer los puntos fuertes y débiles de una organización es a través de las opiniones que generan los propios consumidores.

En la actualidad, una gran cantidad de las compras y de los servicios contratados que se efectúan, no están condicionados

por las sugerencias de las campañas de publicidad y los trucos del marketing, sino por los comentarios que otros consumidores han escrito en los múltiples foros virtuales (públicos y privados) que hoy ofrece la Web 2.0. Con la explosión de la Web 2.0, plataformas como blogs y redes sociales, los consumidores tienen a su disposición un lugar donde compartir sus experiencias con las diferentes marcas y donde poder dar sus opiniones, positivas o negativas sobre cualquier producto o servicio. Las principales empresas empiezan a darse cuenta que estos comentarios de los consumidores pueden manejar la enorme influencia en la formación de las opiniones de otros consumidores. Las empresas que deseen crecer deben de responder a las perspicacias de los consumidores, y es por todo esto que tienen la obligación de analizar los medios de comunicación sociales, para obtener la información adecuada para modificar sus mensajes de marketing, modificar el desarrollo de los productos, etc [1].

Gracias a la Web 2.0 gana peso la opinión del ciudadano frente a las marcas y sus técnicas comerciales más tradicionales. Según el Instituto Nacional de Estadística (INE¹), el 80% de los internautas reconoce que acude a la red para informarse sobre productos, marcas y servicios. En otro estudio realizado en 2009 por la Asociación para la Investigación de Medios de Comunicación (AIMC²), el 75.5% de internautas españoles admite haberse documentado en internet durante el último año, como paso previo a formalizar una compra de productos o servicios.

Dadas estas circunstancias, los responsables del marketing de las empresas tienen la obligación de supervisar en las redes sociales la información relacionada con sus productos y servicios. Sin embargo, en los últimos años se ha producido una explosión en la Web 2.0 sin precedentes, ocasionando que la supervisión manual de las opiniones de los consumidores se convierta en un trabajo irrealizable. La empresa Technorati³ especializada en blogs estima que 75.000 nuevos blogs son creados diariamente, con 1,2 millones de nuevos comentarios cada día en las que el consumidor comenta sobre productos y servicios. Con estos datos las empresas se ven en la necesidad de aunar esfuerzos por encontrar un método automático que sea capaz de analizar las opiniones de los consumidores.

Efectivamente, encontrar un método capaz de clasificar automáticamente, como positivas o negativas, las opiniones en un texto, sería de enorme utilidad para el marketing de

Manuscrito recibido el 24 de febrero del 2010. Manuscrito aceptado para su publicación el 31 de mayo del 2010.

Enrique Vallés Balaguer y Paolo Rosso trabajan en NLE Lab - ELiRF, DSIC, Universidad Politécnica de Valencia, Valencia, España (e-mail: {evalles,proso}@dsic.upv.es).

Angela Locoro y Viviana Mascardi trabajan en DISI, Università degli Studi di Genova, Genova, Italia (e-mail: {angela.locoro,viviana.mascardi}@unige.it).

¹<http://www.ine.es>

²<http://www.aimc.es/aimc.php>

³<http://www.technorati.com/>

las empresas; del mismo modo que para aquellos que buscan información a partir de grandes cantidades de noticias y de datos Web [2]; incluso también se beneficiarían los sistemas de recomendación y colaboración [3].

En este artículo proponemos un método automático para clasificar la polaridad de las opiniones. Sin embargo, al igual que las empresas están interesadas en las opiniones de los consumidores, también están interesadas en conocer de qué producto o característica del producto están opinando. Este hecho se ve reforzado puesto que en una misma opinión se puede comentar dos conceptos diferentes del mismo dominio con polaridades distintas; es más, incluso en una misma frase. Por ejemplo:

El hotel era bonito aunque las habitaciones estaban sucias

En esta frase el *hotel* tiene una polaridad positiva; pero por otro lado, *las habitaciones* tiene una polaridad negativa; por otro lado, en general esta opinión es positiva. El interés de una empresa no es solamente conocer la polaridad de la opinión en conjunto, sino conocer que características de un producto, gustan o no a los consumidores. Esto ayudará a una empresa a mejorar el producto o servicio según los gustos de los clientes. Volviendo al ejemplo, no sólo es interesante conocer que el cliente esté de acuerdo, sino poder saber que cierto aspecto, en este caso concreto *las habitaciones*, podrían mejorarse.

Es por esto que, dado que las empresas disponen de una ontología propia en la cual se incluyen todos sus productos, hemos basado nuestro algoritmo en ontologías para poder encontrar todas aquellas opiniones realizadas sobre los productos pertenecientes a ésta.

Sin embargo, dado el coste de conseguir la opinión de los consumidores, puede que varias empresas decidan compartir e intercambiar la información que poseen sobre las opiniones de los consumidores, o incluso, llegado el caso extremo en el que dos empresas se fusionen. En estos casos, se debe de encontrar algún método que sea capaz de poder analizar automáticamente las opiniones de los clientes y además que sea compatible con las diferentes ontologías. Por este motivo, proponemos un algoritmo que incluye dentro del análisis de opiniones, una fusión de ontologías.

El objetivo de este artículo es un estudio preliminar sobre el análisis de opiniones con ontologías dentro del dominio del turismo. Para ello nos ponemos en la piel de dos empresas dedicadas al turismo las cuales analizarán opiniones de consumidores sobre conceptos del dominio del turismo y después simularemos que dichas empresas deciden compartir la información.

El artículo está organizado de la siguiente manera: en la sección 2 se introduce el problema del análisis de opiniones, el análisis de opiniones basado en ontologías, y expone el algoritmo propuesto. En la sección 3, se introduce el problema de la fusión de ontologías. En la sección 4 exponemos los resultados obtenidos en los experimentos de fusión de ontologías, y posteriormente, los experimentos en el análisis

de opiniones con ontologías. Finalmente en la sección 5 comentamos las conclusiones y la línea a seguir.

II. ANÁLISIS DE OPINIONES

En la actualidad la información se ha convertido en los cimientos que sostiene al mundo empresarial. La información facilita a la empresa la identificación de sus puntos fuertes, y mucho más importante, sus puntos débiles donde deberá aumentar sus recursos para crecer en el mercado. La mayor fuente de información existente es la Web 2.0, puesto que gracias a las redes sociales, los consumidores tienen a su merced un espacio donde compartir las experiencias vividas con las diferentes marcas. Para las empresas es de vital importancia poder disponer de dicha información.

No obstante, en la actualidad se generan al día una gran cantidad de blogs con sus respectivos comentarios, llegando incluso a generar millones de comentarios al día. Como consecuencia, una supervisión manual de las diferentes opiniones de los consumidores se convierte en una tarea irrealizable. Sin embargo, la información proporcionada por los comentarios de los consumidores es la materia prima más preciada con la que una Empresa 2.0 puede construir su futuro; y es por esto que las empresas realizan continuos esfuerzos para encontrar un método automático que sea capaz de analizar las opiniones de los consumidores.

A. Subjetividad y Opiniones

La adquisición de la polaridad de las palabras y frases es en sí misma una línea activa de investigación dentro del análisis de sentimientos, promovido por el trabajo de Hatzivassiloglou y McKeown en la predicción de la polaridad o la orientación semántica de los adjetivos [4]. Desde entonces se han propuesto varias técnicas para determinar la polaridad de palabras: desde medidas probabilísticas de asociación de palabras [5], así como técnicas que explotan la información sobre relaciones léxicas [6], [7], y utilizando glosas [8] por ejemplo de WordNet [9].

Hasta la actualidad se han realizado diversos trabajos dentro de la disciplina de minería de opiniones con diferentes objetivos: desde determinar si un texto contiene opiniones, es decir, clasificar un texto como subjetivo u objetivo tales como [10], [11], hasta trabajos que se centran en determinar la polaridad (orientación semántica) a nivel de palabras y frases, tales como [12], [13], [14], [15]; existen también estudios donde no sólo determinan la polaridad sino el nivel de ésta, es decir, si es alto/medio/bajo positivo/negativo [16], [17], [18], [19].

Otros estudios relacionados con el análisis de opiniones se centran en la extracción de emociones a partir del texto [20], [21], [22], [23]; otros en cambio se centran en la extracción del opinante para tareas de búsqueda de respuesta (QA, por sus siglas en inglés) de opiniones [24]; también, existen estudios destinados a encontrar las tendencias de los consumidores en los blogs [25], [26]. Otros trabajos relacionados miden la

influencia que tienen las opiniones introducidas en los blogs sobre los consumidores [27], [28], [29].

B. Análisis de Opiniones basado en Ontologías

Sin embargo, la mayoría de los trabajos anteriormente mencionados, se centran en la obtención de la opinión sobre un tema específico. Muy pocos tienen en cuenta que un tema puede estar dividido en diferentes subtemas, los cuales describen dos aspectos distintos del tema. Es más, dos subtemas del mismo tema pueden tener polaridades opuestas. Uno de los primeros estudios que se centra en este aspecto fue el trabajo realizado por Hu y Liu [30], el cual se centra en la obtención de las opiniones sobre las propiedades de un producto. Dichas propiedades se obtenían utilizando minería de asociación entre palabras.

Recientemente se han realizado dos trabajos con el objetivo de calcular la polaridad por medio de las propiedades de un concepto, pero a diferencia del trabajo de Hu y Liu, en estos trabajos construyen primero las ontologías del dominio al que pertenecen los textos del corpus y, a partir de éstas, extraen los calificativos de las propiedades de los productos que aparecen en los textos.

El primer trabajo de ellos es el estudio de Zhou y Chaovalit en [3]. En dicho estudio el primer paso que se realiza es generar una ontología; la generación de esta ontología se realiza manualmente por analistas a partir del corpus que posteriormente se utilizará para los experimentos de minería de opiniones. Una vez construida la ontología, suponen que dado un concepto c de la ontología, el cual está descrito con n propiedades: p_1, p_2, \dots, p_n , y que en un texto t se puede expresar opiniones sobre cualquiera de estas propiedades; por tanto, de acuerdo con el número de propiedades de c , t puede dividirse en m segmentos (siendo $m < n$). Es decir que, la predicción de la polaridad de t se define como:

$$polaridad(t) = \begin{cases} positivo, & \text{si } \left(\frac{1}{m} \sum_{i=1}^n w_i v_i \geq 0 \right) \\ negativo, & \text{en caso contrario} \end{cases} \quad (1)$$

siendo $w_i \in [0, 1]$ el peso de la propiedad p_i y $v_i \in [-1, 1]$ el valor de la polaridad de la propiedad p_i calculadas por estimación de máxima verosimilitud.

El otro estudio que se ha realizado hasta ahora sobre análisis de opiniones basado en ontologías, es el trabajo de Zhao y Li en [31]. En este estudio, se genera una ontología automáticamente a partir del corpus que posteriormente se utilizará para el análisis de opiniones. Una vez generada la ontología, los autores proponen extraer los calificativos de las propiedades de los conceptos por medio de la ontología, para posteriormente identificar la polaridad de dichos calificativos utilizando la herramienta SentiWordNet⁴. Una vez extraídos los calificativos y calculado sus polaridades, obtienen la

⁴<http://sentiwordnet.isti.cnr.it/>

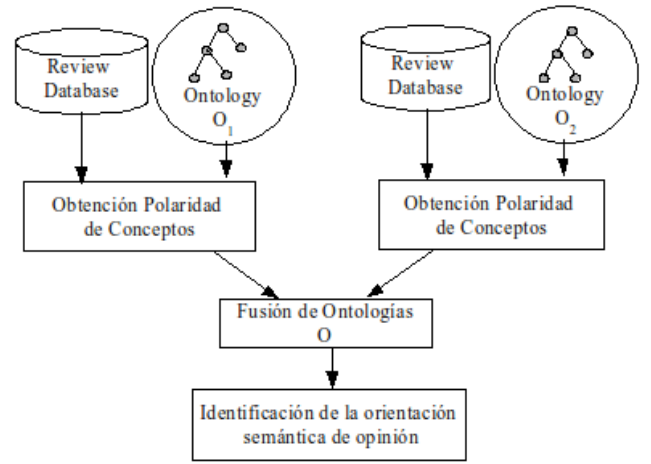


Fig. 1. Algoritmo para el análisis de opiniones via fusión de ontologías.

orientación semántica del texto a partir de la jerarquía de la ontología, para ello calculan la orientación negativa, positiva y neutra según las siguientes ecuaciones:

$$op_{hlc}(neg) = \frac{\sum_{ch_node_{ws_i \in neg}} score(ch_node_{ws_i \in neg})}{|ch_node_{ws_i \in neg}|} \quad (2)$$

$$op_{hlc}(pos) = \frac{\sum_{ch_node_{ws_i \in pos}} score(ch_node_{ws_i \in pos})}{|ch_node_{ws_i \in pos}|} \quad (3)$$

$$op_{hlc}(neu) = \frac{\sum_{ch_node_{ws_i \in ne}} score(ch_node_{ws_i \in ne})}{|ch_node_{ws_i \in ne}|} \quad (4)$$

donde $|ch_node_{ws_i}|$ representa la cardinalidad de todos los hijos con la misma opinión. Por último, escogen como orientación del texto aquella de las tres que es mayor que el resto.

C. Análisis de Opiniones via Fusión de Ontologías

Sin embargo, cuando dos empresas tengan la necesidad de compartir información sobre las opiniones de los productos (o llegado el caso en que dos empresas se fusionen), hay que encontrar una manera para poder analizar las opiniones y posteriormente enviar la información a ambas empresas intentando perder lo menos posible de ésta. Para dicho problema proponemos un algoritmo de análisis de opiniones via fusión de ontologías. Se puede ver un esquema de dicho algoritmo en la figura 1. El algoritmo propone que la empresa e_1 obtenga la polaridad de los conceptos de su ontología (O_1), del mismo modo la empresa e_2 obtendrá la polaridad de los conceptos de su ontología O_2 . Posteriormente se realizará una fusión de ontologías mediante una ontología general O (*upper ontology*, de nivel más alto) y a través de ésta, se realizará un cálculo de la orientación semántica de la opinión t mediante la ecuación:

$$os_{neg}(t) = \sum_{c \in O} v_{neg}(c) \quad (5)$$

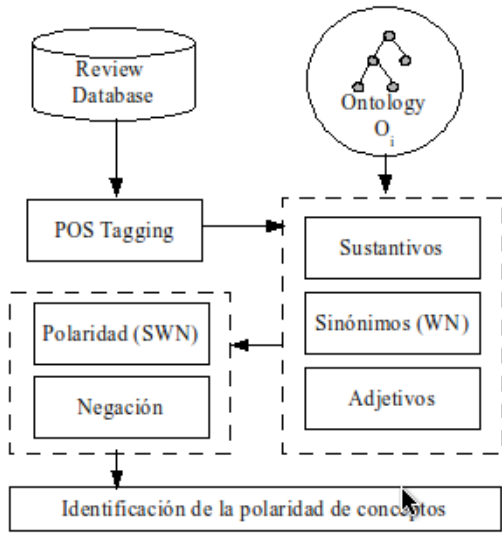


Fig. 2. Algoritmo para la identificación de la polaridad de conceptos.

$$os_{pos}(t) = \sum_{c \in O} v_{pos}(c) \quad (6)$$

donde c son los conceptos que pertenecen a la ontología O , $v_{neg}(p)$ es la polaridad negativa de la propiedad p y $v_{pos}(p)$ la polaridad positiva. Por tanto, la orientación semántica de la opinión (t) se define como:

$$polaridad(t) = \begin{cases} \text{positivo}, & \text{si } os_{pos}(t) > os_{neg}(t) \\ \text{negativo}, & \text{en caso contrario} \end{cases} \quad (7)$$

Para la obtención de la polaridad de los conceptos y propiedades de las ontologías se propone los siguientes pasos (se puede ver un esquema en la figura 2):

- En el primer paso etiquetamos cada una de las palabras de los textos (Part Of Speech tagger); para este paso hemos utilizado el toolkit GATE⁵.
- Posteriormente, buscamos las frases que contienen algún concepto (c) de la ontología (O_i) ; para ello buscamos en cada texto los nombres (o grupos de nombres) que coinciden con un concepto de la ontología. Para aquellas frases que no contengan ningún concepto de la ontología, utilizamos WordNet⁶ (WN) para encontrar sinónimos de los nombres que aparecen en la frase, que pueden ser sinónimos a su vez, de algún concepto de la ontología.
- Seguidamente, extraemos de las frases obtenidas en el paso anterior, los adjetivos adyacentes de cada concepto $(adj(c))$.
- En el siguiente paso obtenemos la polaridad de los adjetivos utilizando SentiWordNet (SWN).

- Comprobamos que la frase es afirmativa, en caso contrario, invertimos la polaridad que nos devuelve SentiWordNet.
- Guardamos las polaridades de cada concepto en un archivo XML con los datos del concepto y el valor de la polaridad.

Hay que destacar, como ya hemos comentado, tanto en el estudio de Zhou y Chaovalit como en el de Zhao y Li, la ontología utilizada se crea a partir del corpus que posteriormente se utilizará en los experimentos para el análisis de opiniones, por otro lado en nuestro trabajo utilizamos ontologías existentes con anterioridad. Nuestra opinión es que esta forma es la más cercana al problema del mundo real, puesto que las empresas tienen una ontología propia y no creemos que sea adecuado modificar su ontología para realizar la búsqueda de opiniones.

Sin embargo para la realización de nuestro algoritmo hemos de encontrar un método de fusión de ontologías que sea eficaz.

III. FUSIÓN DE ONTOLOGÍAS

Las empresas interactúan con la información almacenada mediante ontologías, las cuales proveen un vocabulario de un dominio específico y una especificación de los términos utilizados en dicho vocabulario [32]. Sin embargo, la capacidad de los humanos para tener diferentes puntos de vista de un mismo concepto no tiene límites. Por ello, no es de extrañar que diferentes analistas, desarrolladores e ingenieros del conocimiento tengan diferentes conceptos de la realidad sobre un mismo dominio, generando diferentes ontologías.

Sin embargo, por circunstancias del mercado las empresas pueden fusionarse, y en este caso es necesario no perder ninguna información almacenada anteriormente por cada una de ellas, sobre todo teniendo en cuenta lo esencial que es la información. Tampoco sería coherente utilizar dos ontologías distintas para un mismo dominio en una misma empresa. Por tanto, el primer paso técnico con relación a la información que se realiza es encontrar las relaciones semánticas entre los conceptos de las diferentes ontologías. Este proceso se conoce como *fusión de ontologías* [32].

Realizar la fusión manualmente es un trabajo laborioso, y que en la mayoría de los casos se convierte en imposible. Por consiguiente, es necesario encontrar un método (semi-)automático que facilite la fusión entre diferentes ontologías. El problema de la fusión entre ontologías puede ser abordado desde diversos puntos de vista y este hecho se refleja en la variedad de métodos de fusión que se han propuesto en la literatura. Muchos de ellos tienen sus raíces en el problema clásico de la fusión de esquemas en el área de las bases de datos, tales como Artemis [33], COMA [34], Cupid [35], Similarity Flooding [36]...; mientras que otros han sido específicamente diseñados para trabajar con ontologías, como son GLUE [37], QOM [38], OLA [39], S-Match [40], ASCO [41], PROMPT [42], HCONEE-merge [43] y SAMBO [44]. Algunos de estos métodos se basan en el razonamiento

⁵<http://gate.ac.uk/>

⁶<http://wordnet.princeton.edu/>

formal de la estructura de las descripciones de la entidad como S-Match y OLA; otros en cambio, utilizan una combinación de similitud y grafos de razonamiento como en Similarity Flooding; e incluso otros se basan en algoritmos de aprendizaje automático como GLUE.

En estas investigaciones se explota la información lingüística, la estructura de la información así como el dominio del conocimiento para encontrar un buen alineamiento entre los elementos de cada ontología [45]. Algunos resultados experimentales de algunas herramientas sobre un test estándar de parejas de ontologías aparecen en I3CON 2003⁷, EON 2004⁸ y OAEI 2005⁹. Dos buenas revisiones sobre los trabajos realizados para el estudio de la fusión de ontologías puede encontrarse en [46], [47].

A. Ontologías de Nivel más Alto

Una ontología de nivel más alto (*upper ontology*), como se define en [48], es una ontología independiente del dominio, que proporciona un marco por el cual distintos sistemas pueden utilizar una base común de conocimiento y desde el cual se pueden derivar ontologías de dominio específico. Los conceptos expresados son destinados a ser fundamentales y universales para garantizar la generalidad y la expresividad de una amplia gama de dominios [49]. Una ontología de nivel más alto se caracteriza a menudo como la representación de conceptos que son básicos para la comprensión humana del mundo [50].

Existen varias ontologías de nivel más alto implementadas, como BFO [51], Cyc [52], DOLCE [53], GFO [54], PROTON [55], Sowa's ontology [56] y SUMO [57]. Se puede encontrar una comparación de las distintas ontologías de nivel más alto mencionadas anteriormente en [58].

En nuestros experimentos hemos utilizado SUMO y OpenCyc. En los siguientes sub-apartados describimos cada una de estas ontologías generales.

1) *SUMO*: SUMO¹⁰ (*Suggested Upper Merged Ontology*) es una ontología creada por Teknowledge Corporation¹¹ con una amplia contribución de la lista de correo SUO¹² (*Standard Upper Ontology*), y fue propuesta como documento de iniciación para el Grupo de Trabajo SUO [57], un grupo de trabajo con colaboradores de los campos de la ingeniería, la filosofía y ciencias de la información. SUMO es una de las más grandes ontologías formales públicas existentes hoy día. Cuenta con 20.000 términos y 70.000 axiomas cuando todos los dominios son combinados. SUMO está compuesto por una ontología de nivel medio (*Mid-Level Ontology* (MILO)), ontologías de comunicaciones, países y regiones, computación distribuida, economía, finanzas, componentes de ingeniería, geografía, gobierno, militar, sistema de clasificación industrial

⁷<http://www.atl.external.lmco.com/projects/ontology/i3con.html>

⁸<http://km.aifb.uni-karlsruhe.de/ws/eon2004/>

⁹<http://oaei.inrialpes.fr/2005/>

¹⁰<http://dream.inf.ed.ac.uk/projects/dor/sumo/>

¹¹<http://www.teknowledge.com/>

¹²<http://suo.ieee.org/>

TABLA I
ONTOLOGÍAS USADAS EN LOS EXPERIMENTOS.

Ontologías	Conceptos
ETP-tourism	194
qallme-tourism	125
Tourism-ProtegeExportOWL	86
TravelOntology	35
e-tourism	20

de Norte América, gente, los elementos físicos, cuestiones internacionales, transporte, aeropuertos mundiales y armas de destrucción masiva.

2) *OpenCyc*: El *Cyc Knowledge Base*¹³ (KB) es una base de conocimiento multicontextual desarrollada por Cycorp. Cyc es una representación formalizada de una cantidad enorme de conocimiento fundamental humano: hechos, reglas básicas, y heurísticas para razonar sobre los objetos y los acontecimientos de la vida cotidiana. Cyc KB consiste en términos y las aserciones que relacionan los términos.

KB Cyc se divide en varias *microteorías*, cada una es esencialmente un conjunto de las aserciones que comparten un juego común de suposiciones; algunas *microteorías* son enfocadas en un dominio particular de conocimiento, en un nivel particular de detalle, etc.

Actualmente, el KB Cyc contiene casi doscientos mil términos y varias docenas de aserciones introducidas manualmente sobre cada término. Cyc es un producto comercial, sin embargo Cycorp dispone de una versión de código abierto OpenCyc¹⁴.

IV. EXPERIMENTOS EN EL DOMINIO DEL TURISMO

La intención de los experimentos es la de validar el algoritmo propuesto de análisis de opiniones via fusión de ontologías en el dominio del turismo. Para ello hemos realizado dos fases de experimentos: en la primera fase, se ha realizado un estudio para encontrar el método más idóneo para la etapa de fusión de ontologías en el dominio del turismo; y en la segunda fase, se han realizado los experimentos dedicados a la validación del algoritmo propuesto de análisis de opiniones via fusión de ontologías en el dominio del turismo.

A. Fusión de Ontologías con SUMO y OpenCyc

Para realizar la búsqueda de ontologías existentes de en el dominio del turismo, hemos utilizado Swoogle¹⁵. Tras estudiar las ontologías existentes en el dominio del turismo, hemos seleccionado para nuestros experimentos las ontologías que se muestran en la tabla I.

Hemos empezado creando un alineamiento manual entre todas las ontologías, para utilizarlo como base para medir la calidad de los alineamientos generados por las técnicas utilizadas en los experimentos. Posteriormente, hemos

¹³<http://www.cyc.com>

¹⁴<http://opencyc.org/>

¹⁵<http://swoogle.umbc.edu/>

realizado los tests sobre los métodos directos. Para estas pruebas hemos utilizado la API de Euzenat¹⁶, la cual es una API para la ejecución de alineamientos entre ontologías. El Alignment API permite la ejecución de diferentes medidas de distancia entre textos. En los tests hemos utilizado las medidas *equal* que compara si el texto es exacto, *SMOA* [59] y por último, la distancia de edición de *Levenshtein* [60].

Seguidamente, hemos realizado las pruebas con las técnicas vía ontologías de nivel más alto. Para ello hemos utilizado la API desarrollada en [61]. La API está desarrollada en Java y permite la ejecución de dos métodos diferentes¹⁷:

- El método no estructurado: los conceptos $c \in o$ y $c' \in o'$ están relacionados si corresponden al mismo concepto $c_u \in u$, donde u es la ontología de nivel más alto; o y o' son las ontologías a fusionar.
- El método estructurado: los conceptos $c \in o$ y $c' \in o'$ están relacionados si:
 - Los conceptos c y c' corresponden al mismo concepto $c_u \in u$, por tanto están relacionados con una medida de confianza de $conf_1 * conf_2$.
 - El concepto c corresponde al concepto c_u , c' corresponde con el concepto c'_u , y c'_u es un super-concepto de c_u en u (o viceversa), por tanto están relacionados con una medida de confianza de $conf_1 * conf_2 * df$.
 - El concepto c corresponde al concepto c_u , c' corresponde con el concepto c'_u , y c'_u tiene algún super-concepto en común con c_u , por tanto están relacionados con una medida de confianza de $conf_1 * conf_2 * df^2$.

En la ejecución primero hemos utilizado la API de Euzenat entre la ontología de nivel más alto y cada una de las ontologías a alinear. Al utilizar la API de Euzenat nos permitía ejecutar cada una de las funciones de distancia anteriormente citadas. Una vez realizados los alineamientos de las ontologías seleccionadas con las ontologías de nivel más alto utilizamos la API de [62] para realizar los alineamientos estructurados y no estructurados. Este proceso lo realizábamos para cada una de las dos ontologías de nivel más alto.

En la tabla II¹⁸ aparecen los resultados obtenidos en los experimentos para el alineamiento entre las ontologías *ETP-tourism*¹⁹ y *qallme-tourism*²⁰. Los campos que aparecen en la tabla corresponden a: función de distancia (Dis), ontología de nivel más alto utilizada (UO), método estructurado o no estructurado (Met), número de alineamientos encontrados (Enc), número de alineamientos correctos (Cor), y las tres medidas: precisión (Pre), recall (Rec) y F-measure

TABLA II
RESULTADOS DE MAPEO DE ONTOLOGÍAS.

Dis	UO	Met	Enc	Cor	Pre	Rec	F-M
Man.	None	None	98	98	1.00	1.00	1.00
equal	None	None	195	80	0.41	0.82	0.55
smoa	None	None	221	84	0.38	0.86	0.52
Lev.	None	None	205	82	0.40	0.84	0.54
equal	Sumo	NoSt	18	14	0.78	0.14	0.24
smoa	Sumo	NoSt	415	83	0.20	0.85	0.32
Lev.	Sumo	NoSt	264	74	0.28	0.76	0.41
equal	Sumo	Str	20	14	0.70	0.14	0.24
smoa	Sumo	Str	461	83	0.18	0.85	0.30
Lev.	Sumo	Str	264	74	0.28	0.76	0.41
equal	Cyc	NoSt	38	16	0.42	0.16	0.24
smoa	Cyc	NoSt	143	80	0.56	0.82	0.67
Lev.	Cyc	NoSt	122	78	0.64	0.80	0.71
equal	Cyc	Str	53	16	0.30	0.16	0.21
smoa	Cyc	Str	200	80	0.40	0.82	0.54
Lev.	Cyc	Str	144	78	0.54	0.80	0.64

(F-M). Por último, hemos señalado en la tabla los mejores resultados para cada test en las diferentes medidas.

Observando los datos podemos destacar que al aplicar la técnica de *matching* vía ontologías de nivel más alto con el método no estructurado [61] obtenemos una ganancia media de precisión frente a los métodos directos, de alrededor del 4,3 % para SUMO y de un 36,6 % para OpenCyc. En recall se produce una pérdida media de 31,2 % para SUMO y de 30 % para OpenCyc. Y finalmente, en F-measure se produce una pérdida media de 39,6 % para SUMO y una ganancia media de 1,3 % para OpenCyc. En cambio, al aplicar la técnica de *matching* vía ontología de nivel más alto con el método estructurado obtenemos una pérdida media de precisión frente a los métodos directos del 3,97 % para SUMO y una ganancia media de un 4,5 % para OpenCyc. En recall se produce una pérdida media de 31,2 % para SUMO y de 30 % para OpenCyc. Y finalmente, en F-measure se produce una pérdida media del 41 % para SUMO y de 13,2 % para OpenCyc.

Estos datos nos podrían indicar que se obtienen mejores resultados con las técnicas directas que realizando el mapeo vía ontologías de nivel más alto. Sin embargo, en los diferentes experimentos se ha comprobado que conforme aumenta el número de términos de cada ontología, es notable una mejora en los métodos utilizando ontologías de nivel más alto respecto a los métodos directos en cuanto a las medidas de precisión y F-measure, y una mínima pérdida en recall. Esta mejora se nota sobre todo al utilizar OpenCyc como ontología de nivel más alto. Este hecho se puede comprobar en la tabla II. Esto nos puede dar un indicio de que cuanto mayor son las ontologías es preferible utilizar las técnicas vía ontología de nivel más alto.

Por tanto, hemos decidido utilizar para el paso de fusión de ontologías, los métodos no estructurados vía ontología de nivel más alto, seleccionando Cyc como ontología de nivel más alto, y utilizando como ontologías de origen *ETP-tourism* y *qallme-tourism*.

¹⁶<http://alignapi.gforge.inria.fr/>

¹⁷Información detallada puede verse en [62]

¹⁸Las limitaciones de espacio nos impiden ilustrar el resto de alineamientos; para una descripción más detallada véase <http://users.dsic.upv.es/grupos/nle/?file=kop4.php>

¹⁹http://www.info.uqam.ca/Members/valchev_p/mbox/ETP-tourism.owl

²⁰<http://qallme.fbk.eu/>

TABLA III
RESULTADOS OBTENIDOS DIVIDIENDO EL CORPUS.

Ontología	Num.	Acc.
ETP Tourism	1.500	72,41 %
qallme-tourism	1.500	70,92 %
Mapeo de ontologías	3.000	71,13 %

B. Análisis de Opiniones con OpenCyc

Una vez validado el método para la fusión de ontologías, hemos realizado los experimentos para el análisis de opiniones vía fusión de ontologías. Para la realización de dichos experimentos hemos creado un corpus formado por 3.000 textos de opiniones (1.500 positivos y 1.500 negativos) extraídas desde TripAdvisor²¹. Los textos corresponden a conceptos como hoteles, restaurantes y ciudades. TripAdvisor nos ofrecía una ventaja, ya que en este blog los usuarios no sólo escriben sus opiniones sino que además puntúan el producto entre Excelente, Muy bueno, Regular, Malo y Pobre. De esta forma no teníamos que analizar las opiniones manualmente para clasificar el valor de la orientación semántica de cada una de ellas.

En los experimentos hemos intentado simular como actuarían dos empresas dedicadas al dominio del turismo para obtener información sobre algunos productos ofertados. Para ello primero realizarán cada una de ellas un proceso de análisis de opiniones sobre el corpus creado con dos diferentes ontologías del dominio del turismo. Posteriormente, suponemos que las empresas desean intercambiar información, o compartirla, o en un caso extremo que dos empresas se fusionen, por tanto realizarán un proceso de fusión de ontologías. Con estos experimentos podremos tener una base para decidir la validez de nuestro algoritmo basado en la fusión de ontologías para el análisis de opiniones.

Para poder medir la eficacia del algoritmo propuesto, hemos realizado dos diferentes experimentos: en el primer experimento hemos separado el corpus para cada una de las dos empresas, con la intención de simular que ocurriría si dos empresas analizan diferentes textos antes de compartir la información sobre el análisis de opiniones; y en el segundo, hemos utilizado el corpus completo para las dos ontologías, simulando que dos empresas analizan anteriormente los mismos textos.

En la tabla III se muestran los resultados obtenidos en el primer experimento. Un dato destacable es que tras realizar el proceso de fusión de ontologías se obtiene una tasa de aciertos (Acc) de 71,33 % que es cercano al obtenido por separado en cada ontología, incluso es superior al obtenido con la ontología *qallme-tourism*. Estos resultados nos dan a entender que al realizar el proceso de fusión de ontologías no se pierden datos referentes al proceso de análisis de opiniones realizado con antelación.

TABLA IV
RESULTADOS OBTENIDOS CON EL CORPUS COMPLETO.

Ontología	Num.	Acc.
ETP Tourism	3.000	72,2 %
qallme-tourism	3.000	71,2 %
Mapeo de ontologías	3.000	71,33 %

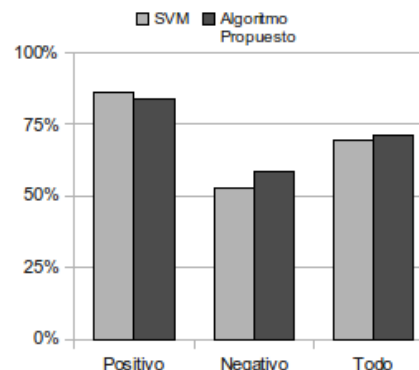


Fig. 3. Resultados de minería de polaridad.

Los resultados del segundo experimento realizado el proceso de calcular la polaridad para cada ontología con el corpus completo se muestra en la tabla IV. Como se observa en los resultados tras realizar el proceso de fusión de ontologías es muy similar al obtenido en el experimento anterior. Estos resultados nos dan a entender que al realizar el proceso de fusión de ontologías no se pierden datos referentes al proceso de análisis de opiniones realizado con antelación.

En la figura 3 se muestra una comparación entre los resultados obtenidos mediante nuestro algoritmo propuesto y utilizando un método de aprendizaje automático como *support vector machine* (SVM). La figura muestra que nuestro algoritmo incrementa el porcentaje de aciertos casi en dos puntos, de un 69,5 % a un 71,33 %. Es más también se observa que nuestro algoritmo tiene menos diferencia entre los resultados obtenidos con las opiniones positivas y las negativas, que utilizando SVM donde los resultados con las opiniones positivas son mucho mayores que en las negativas.

Una interesante observación que se desprende de los resultados de los experimentos y que se ve reflejado en la figura 3 es la diferencia de porcentajes de aciertos que existe cuando examinamos únicamente las opiniones positivas frente a cuando examinamos las opiniones negativas. Esto no ocurre únicamente con nuestro algoritmo propuesto, sino que también cuando utilizamos SVM. Analizando los textos que componen nuestro corpus hemos observado que las personas cuando estamos en desacuerdo con algún producto o servicio tenemos cierta tendencia a utilizar la ironía y el sarcasmo [63], [64]. Éste provoca que al extraer los adjetivos para obtener la orientación semántica, éstos tendrán la polaridad cambiada, llevando a clasificar los textos con la polaridad incorrecta. Pero este hecho, aunque es mucho más frecuente en

²¹<http://www.tripadvisor.com>

opiniones negativas, también se utiliza pero con menor medida en opiniones positivas.

V. CONCLUSIONES Y TRABAJO FUTURO

En este trabajo se ha presentado un método para el análisis de opiniones vía fusión de ontologías. Esta fusión permitirá a empresas poder compartir y/o intercambiar información sobre las opiniones de los consumidores con respecto a productos. Sin embargo, se ha demostrado la dificultad que entraña la tarea de analizar las opiniones en la Web 2.0 de los usuarios de blogs a través de una ontología, principalmente cuando esta ontología esta preestablecida. En estudios anteriores como [31] y [3], las ontologías en cambio se generaban a partir del corpus con lo que facilitaba la búsqueda de conceptos. Sin embargo, creemos que nuestros experimentos están más próximos al problema del mundo real, puesto que las empresas ya poseen de antemano una ontología del dominio. No obstante, hemos comprobado cómo al realizar el proceso de fusión de ontologías no se pierde prácticamente ningún dato de los anteriormente calculados por el análisis de opiniones.

Un hecho resaltante en el proceso de fusión de ontologías es que conforme aumenta el número de términos de las ontologías orígenes, se obtiene mejores resultados utilizando los métodos vía ontologías de nivel más alto respecto a los métodos directos en relación con las medidas de precisión y F-measure, y una leve pérdida en recall. En cuanto a las ontología de nivel más alto, se observa que con OpenCyc hay una leve mejora frente a SUMO, en precisión y F-measure, y una mayor mejoría de más del 10 % en recall. Por tanto, es lógico pensar que es preferible utilizar en el dominio del turismo OpenCyc.

Para un futuro estudio sería interesante utilizar no sólo los adjetivos para hallar la polaridad de los textos sino también verbos o adverbios. Otro aspecto interesante a tener en cuenta en futuros trabajos sería introducir algún método automático para detectar la ironía y el sarcasmo. Esto nos ayudaría a poder clasificar correctamente la polaridad, ya que si utilizamos el sarcasmo podemos invertir la polaridad de sus palabras. Detectar automáticamente el sarcasmo sería efectivo sobre todo para mejorar el porcentaje de aciertos en las opiniones negativas.

AGRADECIMIENTOS

El trabajo de los dos primeros autores se engloba dentro del proyecto del MICINN: *TEXT-ENTERPRISE 2.0: Técnicas de Comprensión de textos aplicadas a las necesidades de la Empresa 2.0* (TIN2009-13391-C04-03). El trabajo ha sido el resultado de una estancia de Erasmus-Máster de 4 meses en la Universidad de Génova.

REFERENCIAS

[1] J. Zabin and A. Jefferies, "Social media monitoring and analysis: Generating consumer insights from online conversation," Aberdeen Group Benchmark Report, January 2008.

[2] T. Wilson, J. Wiebe, and R. Hwa, "Just how mad are you? finding strong and weak opinion clauses," *AAAI'04: Proceedings of the 19th national conference on Artificial intelligence*, pp. 761–769, 2004.

[3] L. Zhou and P. Chaovalit, "Ontology-supported polarity mining," *Journal of the American Society for Information Science and Technology*, vol. 59, no. 1, pp. 98–110, 2008.

[4] V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives," *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*, pp. 174–181, 1997.

[5] P. D. Turney and M. L. Littman, "Measuring praise and criticism: Inference of semantic orientation from association," *ACM Trans. Inf. Syst.*, vol. 21, no. 4, pp. 315–346, 2003.

[6] J. Kamps and M. Marx, "Words with attitude," *1st International WordNet Conference*, pp. 332–341, 2002.

[7] S. Kim and E. Hovy, "Determining the sentiment of opinions," *COLING '04: Proceedings of the 20th international conference on Computational Linguistics*, pp. 1267–1373, 2004.

[8] A. Esuli and F. Sebastiani, "Determining the semantic orientation of terms through gloss classification," *CIKM '05: Proceedings of the 14th ACM international conference on Information and knowledge management*, pp. 617–624, 2005.

[9] A. Andreevskaia and S. Bergler, "Mining wordnet for a fuzzy sentiment: Sentiment tag extraction from wordnet glosses," *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2006)*, pp. 209–216, 2006.

[10] E. Riloff, J. Wiebe, and W. Phillips, "Exploiting subjectivity classification to improve information extraction," in *Proc. of the NCAI*, vol. 20, 2005, p. 1106.

[11] J. Wiebe, T. Wilson, and C. Cardie, "Annotating expressions of opinions and emotions in language," in *Language Resources and Evaluation*, 2005, pp. 165–210.

[12] A. Esuli and F. Sebastiani, "Determining term subjectivity and term orientation for opinion mining," *EACL 2006, 11th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 193–200, 2006.

[13] H. Kanayama and T. Nasukawa, "Fully automatic lexicon expansion for domain-oriented sentiment analysis," *EMNLP '06: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pp. 355–363, 2006.

[14] P. D. Turney, "Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews," *ACL '02: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pp. 417–424, 2004.

[15] B. Pang and L. J. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," *ACL '04: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, pp. 271–278, 2004.

[16] K. Dave, S. Lawrence, and D. M. Pennock, "Mining the peanut gallery: opinion extraction and semantic classification of product reviews," *WWW '03: Proceedings of the 12th international conference on World Wide Web*, pp. 519–528, 2003.

[17] A. B. Goldberg and X. Zhu, "Seeing stars when there aren't many stars: graph-based semi-supervised learning for sentiment categorization," *TextGraphs '06: Proceedings of TextGraphs: the First Workshop on Graph Based Methods for Natural Language Processing on the First Workshop on Graph Based Methods for Natural Language Processing*, pp. 45–52, 2006.

[18] K. Shimada and T. Endo, "Seeing several stars: A rating inference task for a document containing several evaluation criteria," *PAKDD 2008: Proceedings of Advances in Knowledge Discovery and Data Mining, 12th Pacific-Asia Conference*, pp. 1006–1014, 2008.

[19] S. Baccianella, A. Esuli, and F. Sebastiani, "Multi-facet rating of product reviews," *ECIR '09: Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval*, pp. 461–472, 2009.

[20] C. Strapparava and A. Valitutti, "WordNet-Affect: an affective extension of WordNet," *LREC '04: Proceedings of the 4th International Conference on Language Resources and Evaluation*, vol. 4, pp. 1083–1086, 2004.

[21] C. Strapparava and R. F. Mihalcea, "Semeval-2007 task 14: affective text," *SemEval '07: Proceedings of the 4th International Workshop on Semantic Evaluations*, pp. 70–74, 2007.

- [22] A. Balahur and A. Montoyo, "Applying a culture dependent emotion triggers database for text valence and emotion classification," *Procesamiento del lenguaje natural*, vol. 40, pp. 107–114, 2008.
- [23] C. Strapparava and R. Mihalcea, "Learning to identify emotions in text," *SAC '08: Proceedings of the 2008 Association for Computational Linguistics Symposium on Applied Computing*, pp. 1556–1560, 2008.
- [24] S. Kim and E. Hovy, "Identifying opinion holders for question answering in opinion texts," *Proc. of AAAI Workshop on Question Answering in Restricted Domains*, 2005.
- [25] N. S. Glance, M. Hurst, and T. Tomokiyo, "Blogpulse: Automated trend discovery for weblogs," *WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics*, 2004.
- [26] M. Platakis, D. Kotsakos, and D. Gunopulos, "Discovering hot topics in the blogosphere," in *Proc. of the Panhellenic Scientific Student Conference on Informatics, Related Technologies and Applications EUREKA*, 2008, pp. 122–132.
- [27] K. E. Gill, "How can we measure the influence of the blogosphere?" *WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics*, 2004.
- [28] A. Java, P. Kolar, T. Finin, and T. Oates, "Modeling the spread of influence on the blogosphere," *Proceedings of the 15th International World Wide Web Conference*, 2006.
- [29] A. Kale, "Modeling trust and influence in the blogosphere using link polarity," *ICWSM '07: Proceedings of the International Conference on Weblogs and Social Media*, 2007.
- [30] M. Hu and B. Liu, "Mining and summarizing customer reviews," *KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 168–177, 2004.
- [31] L. Zhao and C. Li, "Ontology based opinion mining for movie reviews," in *Proc. of KSEM*, 2009, pp. 204–214.
- [32] P. Shvaiko, "Iterative schema based semantic matching," in *PhD-Thesis, International Doctorate School in Information and Communication Technology*, 2006, univ. Trento, Italia.
- [33] S. Castano, V. De Antonellis, and S. De Capitani di Vimercati, "Global viewing of heterogeneous data sources," *IEEE Trans. on Knowl. and Data Eng.*, vol. 13, no. 2, pp. 277–297, 2001.
- [34] H.-H. Do and E. Rahm, "COMA: A system for flexible combination of schema matching approaches," *VLDB '02: Proceedings of the 28th international conference on Very Large Data Bases*, pp. 610–621, 2002.
- [35] J. Madhavan, P. A. Bernstein, and E. Rahm, "Generic Schema Matching with Cupid," *VLDB '01: Proceedings of the 27th International Conference on Very Large Data Bases*, pp. 49–58, 2001.
- [36] S. Melnik, H. Garcia-Molina, and E. Rahm, "Similarity Flooding: A Versatile Graph Matching Algorithm and Its Application to Schema Matching," *ICDE '02: Proceedings of the 18th International Conference on Data Engineering*, p. 117, 2002.
- [37] A. Doan, J. Madhavan, R. Dhamankar, P. Domingos, and A. Halevy, "Learning to match ontologies on the semantic web," *The VLDB Journal*, vol. 12, no. 4, pp. 303–319, 2003.
- [38] M. Ehrig and S. Staab, "Qom - quick ontology matching," pp. 683–697, 2004.
- [39] J. Euzenat, P. Gagan, and P. Valtchev, "OLA in the OAEI 2005 alignment contest," *Proceedings of the K-CAP Workshop on Integrating Ontologies*, pp. 61–71, 2005.
- [40] F. Giunchiglia, P. Shvaiko, and M. Yatskevich, "S-match: an algorithm and an implementation of semantic matching," in *Proc. of ESWS 2004*, Y. Kalfoglou and et al., Eds. Springer, 2004, pp. 61–75.
- [41] B. Le, R. Dieng-Kuntz, and F. Gandom, "On ontology matching problems - for building a corporate semantic web in a multi-communities organization," in *Proc. of the Sixth International Conference on Enterprise Information Systems*, no. 4, Abril 2004, pp. 236–243.
- [42] N. Noy and M. Musen, "The PROMPT suite: interactive tools for ontology merging and mapping," *International Journal of Human-Computer Studies*, vol. 59, no. 6, pp. 983–1024, 2003.
- [43] K. Kotis, G. Vouros, and K. Stergiou, "Capturing semantics towards automatic coordination of domain ontologies," in *the 11th International conference of Artificial Intelligence: Methodology, Systems, Architectures - Semantic Web Challenges - AIMS 2004*. Springer-Verlag, 2004, pp. 22–32.
- [44] P. Lambrix and H. Tan, "SAMBO-A system for aligning and merging biomedical ontologies," *Web Semant.*, vol. 4, no. 3, pp. 196–206, 2006.
- [45] Y. Qu, W. Hu, and G. Cheng, "Constructing virtual documents for ontology matching," in *WWW '06: Proceedings of the 15th international conference on World Wide Web*, 2006, pp. 23–31.
- [46] N. F. Noy, "Semantic integration: a survey of ontology-based approaches," *SIGMOD Rec.*, vol. 33, no. 4, pp. 65–70, 2004.
- [47] E. Rahm and P. A. Bernstein, "A survey of approaches to automatic schema matching," *The VLDB Journal*, vol. 10, no. 4, pp. 334–350, 2001.
- [48] C. Phytla, "An Analysis of the SUMO and Description in Unified Modeling Language," 2002, no publicado.
- [49] S. Semy, M. Pulvermacher, and L. Obrst, "Toward the use of an upper ontology for U.S. government and U.S. military domains: An evaluation," in *Submission to Workshop on IIWeb*, 2004.
- [50] A. Kiryakov, K. Simov, and M. M. Dimitrov, "Ontomap: portal for upper-level ontologies," in *Proc. of the FOIS*. ACM, 2001, pp. 47–58.
- [51] P. Grenon, B. Smith, and L. Goldberg, "Biodynamic ontology: applying BFO in the biomedical domain," in *Ontologies in Medicine*, D. M. Pisanelli, Ed. IOS Press, 2004, pp. 20–38.
- [52] D. Lenat and R. Guha, *Building Large Knowledge-Based Systems; Representation and Inference in the Cyc Project*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1989.
- [53] A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider, "Sweetening ontologies with DOLCE," in *Proc. of EKAW*. Springer, 2002, pp. 166–181.
- [54] H. Herre, B. Heller, P. Burek, R. Hoehndorf, F. Loebe, and H. Michalek, "General formal ontology (GFO): A foundational ontology integrating objects and processes. Part I: Basic principles," Research Group Ontologies in Medicine (Onto-Med), Univ. Leipzig, Tech. Rep. Nr. 8, 2006.
- [55] N. Casellas, M. Blzquez, A. Kiryakov, P. Casanovas, M. Poblet, and V. Benjamins, "OPIK into PROTON: Legal domain ontology integration into an upper-level ontology," in *Proc. of WORM 2005*. Springer, 2005, pp. 846–855.
- [56] J. Sowa, *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Brooks Cole Publishing, 2000.
- [57] I. Niles and A. Pease, "Towards a standard upper ontology," in *FOIS '01: Proceedings of the international conference on Formal Ontology in Information Systems*. New York, NY, USA: ACM, 2001, pp. 2–9.
- [58] V. Mascardi, V. Cord, and P. Rosso, "A comparison of upper ontologies," in *Atti del Workshop Dagli Oggetti agli Agenti*, WOA, M. Baldoni and et al., Eds. Seneca Editore, 2007, pp. 55–64.
- [59] G. Stoilos, G. Stamou, and S. Kollias, "A string metric for ontology alignment," in *Proc. of the ISWC*, 2005, pp. 624–637.
- [60] V. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," *Soviet Physics Doklady*, vol. 10, no. 8, pp. 707–710, 1966.
- [61] V. Mascardi, A. Locoro, and P. Rosso, "Automatic ontology matching via upper ontologies: A systematic evaluation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 99, no. 1, 2009, doi: 10.1109/TKDE.2009.154.
- [62] A. Locoro, "Ontology Matching using Upper Ontologies and Natural Language Processing," in *PhD-Thesis Course in Electronic and Computer Engineering, Robotics and Telecommunications*, 2010, univ. Genova, Italia.
- [63] A. Utsumi, "A unified theory of irony and its computational formalization," *Proceedings of the 16th conference on Computational linguistics*, pp. 962–967, 1996.
- [64] A. Reyes, P. Rosso, and D. Buscaldi, "Humor in the blogosphere: First clues for a verbal humor taxonomy," *Journal of Intelligent Systems*, vol. 18, no. 4, 2009.