

Seguimiento y análisis automático de contenidos en redes sociales

Alonso, Miguel A.*, Gómez-Rodríguez, Carlos, Vilares, David, Doval, Yeraí, y Vilares, Jesús

Grupo LYS, Departamento de Computación, Facultad de Informática, Universidade da Coruña;
Emails: miguel.alonso@udc.es (MAA), carlos.gomez@udc.es (CGR), david.vilares@udc.es (DV),
yeraí.doval@udc.es (YD), jesus.vilares@udc.es (JV)

* Autor responsable del trabajo; miguel.alonso@udc.es

Abstract: La Minería de Opiniones es la disciplina que aborda el tratamiento automático de las opiniones contenidas en un texto. Permite, por ejemplo, determinar si en un texto se está opinando o no, o si la polaridad o sentimiento que se expresa en el mismo es positiva, negativa o mixta. También permite la extracción automática de características, lo que posibilita conocer la percepción que los autores tienen sobre aspectos concretos de un tema determinado. Este trabajo, tras realizar una introducción a dicho ámbito, presenta una aproximación propia al mismo, la cual destaca por emplear información sintáctica así como por estar especialmente adaptada a uno de los contextos de trabajo más complicados, Twitter. Dicha tecnología es fácilmente aplicable a tareas de inteligencia.

Keywords: Minería de opiniones; análisis del sentimiento; redes sociales; Twitter; procesamiento del lenguaje natural; contraterrorismo

1. Introducción

El anonimato y falta de supervisión en Internet atraen a individuos y organizaciones criminales o extremistas que emplean la Red para actividades que suponen una amenaza y que van desde el acoso al terrorismo [9,12,13]. Asimismo, dada la naturaleza de Internet así como la vasta cantidad de datos a manejar, las técnicas clásicas para la lucha contra este tipo de amenazas son insuficientes. Se hace preciso, pues, emplear técnicas y herramientas computacionales que nos permitan monitorizar y analizar este ingente flujo de datos, así como agregar la información obtenida de diversas fuentes, automatizando en lo posible el proceso mediante la utilización de técnicas de identificación, extracción y procesado inteligente de la información para, de este modo, ayudar al analista de inteligencia en su trabajo [5].

El presente trabajo aborda la experiencia de los autores en la Minería de Textos y de Opiniones, desde el punto de vista de su posible aplicación en tareas de inteligencia.

2. Contexto de la discusión: Minería de Textos, Minería de Opiniones y sus aplicaciones

La *Minería de Textos* (MT) [14] aborda la detección y extracción de información a partir de textos, combinando para ello técnicas estadísticas, de Aprendizaje Automático y de Procesamiento del Lenguaje Natural (PLN). Por abuso del lenguaje, suele hablarse simplemente de PLN en lugar de MT, cosa que también haremos aquí con permiso del lector. A su vez, dentro de dicho campo ha cobrado fuerza la denominada *Minería de Opiniones* (MO) o *Análisis del Sentimiento*, que se centra en el análisis del contenido subjetivo y emocional de los textos [17]. Entre otras cosas, permite determinar si en un texto se está expresando o no una opinión y, de ser así, si ésta es positiva, negativa o neutra, así como también identificar sobre qué entidad, e incluso sobre qué aspecto de la misma, se opina. Asimismo, actualmente este tipo de técnicas se está empleando para monitorizar la actividad radical extremista en la Web.

2.1. Medios de comunicación sociales y extremistas

Los llamados *social media* o *medios de comunicación sociales* tales como foros, blogs, redes sociales, etc., han permitido globalizar la expresión de opiniones inspirándose en las interacciones humanas así como favorecer la creación de comunidades virtuales. Desafortunadamente, en ocasiones éstas se convierten en puntos de encuentro de individuos u organizaciones con todo tipo de motivaciones e ideologías extremistas, pasando a cumplir funciones tales como [3,9,13]: (a) proselitismo y adoctrinamiento ideológico; (b) propaganda; (c) fomento de sentimientos identitarios y de compañerismo entre sus simpatizantes obviando barreras geográficas, socio-económicas e incluso culturales; (d) formación y entrenamiento; y (e) obtención de financiación.

Asimismo, en los *social media* la gente se pronuncia sobre cuestiones de las que no hablaría en su día a día, normalizando comportamientos que no tendría en su vida normal y facilitando así el proceso de radicalización. En paralelo, su interactividad promueve que las líneas divisorias entre espectador pasivo y militante activo se desdibujen, favoreciendo la adopción de roles más activos, tanto “no-violentos” y que hacen más cómoda la adhesión a la causa —p.ej. prestar apoyo financiero o hacer propaganda—, como eminentemente violentos —p.ej. la planificación de ataques terroristas [3,13,23].

Esto concordaría con declaraciones recientes del Ministro del Interior según las cuales el 80% de las personas captadas por los jihadistas en nuestro país lo serían a través de Internet y las redes sociales, cuando hace tan sólo tres años dicha cifra se refería a centros de culto y prisiones. Asimismo, el periodo de captación y adoctrinamiento es cada vez menor, reduciéndose incluso a dos meses. Por tanto, se hace perentorio abordar de forma ágil y eficaz el problema que suponen este tipo de comunidades virtuales.

2.2. Objetivo: la identificación de sospechosos

Se plantean en este contexto dos tareas [5,16]: (a) identificar aquellos foros de encuentro de extremistas que puedan constituir una amenaza y (b) identificar allí a los sujetos más peligrosos. Esto permitiría al analista centrar sus esfuerzos en éstos últimos, optimizando así los recursos disponibles.

En lo referente a la primera fase de localización de webs, foros, etc. de contenido radical, diversos trabajos demuestran la efectividad del uso de técnicas de NLP y MO. El sistema de Brynielsson et al. [4], por ejemplo, emplea técnicas de *crawling* combinadas con un proceso de doble filtrado de

candidatos en base a la temática y al extremismo de sus contenidos. Asimismo, L’Huillier et al. [16] y Ríos y Muñoz [21] señalan las bondades de conjugar técnicas de MO y de *Análisis de Redes Sociales* (ARS) —técnicas que permiten analizar las relaciones existentes entre los miembros de una comunidad (p.ej. los miembros de un foro de debate).

Respecto a la segunda fase, la identificación de los sujetos que suponen una mayor amenaza, los trabajos realizados en este campo se centran, principalmente, en dos aspectos que explicamos a continuación. El primero de dichos aspectos es la identificación de los miembros más influyentes de esas comunidades virtuales, como en los trabajos de Katsimpras et al. [15] y de L’Huillier et al. [16], que combinan técnicas de ARS y NLP; o el de Biran et al. [2], mucho más centrado en NLP y MO —obsérvese que el carisma y los mecanismos de influencia tienen un fuerte componente emocional.

Un segundo aspecto de interés es el de la identificación de los denominados *lobos solitarios*, especialmente difíciles de reconocer por medios tradicionales ya que, si bien pueden simpatizar con ellos, no suelen pertenecer o apoyar activamente a movimientos radicales. Sin embargo, estos sujetos sí suelen dejar un *rastros digital* detectable. Por una parte, la formación terrorista de estos individuos suele ser de tipo autodidacta, empleando Internet para, por ejemplo, acceder a webs sobre cómo fabricar explosivos. Asimismo, suelen dar indicios de sus intenciones (sino anunciarlas directamente) en blogs, foros, etc., información toda ella que podría ser monitorizada empleando herramientas especializadas de NLP y MO con objeto de detectar en las publicaciones *marcadores* (indicios) que denoten conductas peligrosas o estados emocionales del autor conducentes a un comportamiento antisocial, filtrando y ordenando a los sospechosos de acuerdo a su grado de radicalización o amenaza [5,20,22,23].

2.3. Dificultades añadidas

Por otra parte, existen factores que complican la tarea. En primer lugar, monitorizar el tráfico de red de los servidores en los que se publica no suele ser efectivo debido a los cambios de IPs y URLs, los mecanismos de anonimización, etc., además de la cuestión de la privacidad de las comunicaciones [12]. Por otra parte, estos usuarios suelen emplear varios *alias* diferentes a la vez, lo que dificulta su identificación, aspecto éste en el que se ha venido trabajando en base a IPs, técnicas de PLN (p.ej. técnicas de identificación de autoría) y ARS, análisis temporal, etc. [4,12].

En segundo lugar, si bien las técnicas de PLN actuales son extrapolables de forma natural a los entornos web de los medios de comunicación tradicionales, en el caso de las redes sociales nos topamos con serios problemas, ya que este tipo de textos no suele adherirse al uso estándar de la lengua. A este respecto, podemos resumir en dos los principales desafíos a los que hemos de enfrentarnos. Primero, el empleo de textos extremadamente cortos o *microtextos* —p.ej. en Twitter—, a menudo carentes de contexto lingüístico y que necesitan, para su análisis, de un refinamiento y actualización de las técnicas tradicionales de PLN. Segundo, la gran cantidad de *ruido* presente en dichos textos a diversos niveles: escritura no convencional, abreviaciones idiosincrásicas, eliminación de determinantes y otras partículas, el uso de emoticonos a modo de contexto, etc. Se hace también preciso en estos casos bien la adaptación de las técnicas y herramientas de PLN existentes, bien la *normalización* de los textos (i.e. transformarlos en lenguaje estándar sí procesable mediante técnicas clásicas).

El tercer factor es el del *idioma*, pues la gran mayoría de las técnicas de NLP y MO a emplear son dependientes del idioma. Se hace necesario, pues, como paso previo a dicho procesamiento, identificar el idioma en el que están escritos los textos a analizar, cuestión en la que, si bien se han realizado notables avances en el caso de texto “normal”, ciertos dominios como el de los textos no estándar presentes en las redes sociales siguen presentando un desafío debido a las características anteriormente comentadas [1].

3. Trabajo en el ámbito de la Minería de Opiniones

Habiendo descrito ya el contexto de discusión, presentamos a continuación nuestro enfoque para tareas de MO sobre microtextos, uno de los escenarios más complicados sobre los que operar.

3.1. Minería de Opiniones: trabajo de terceros

Existe una cierta experiencia en la aplicación de MO en el entorno web, principalmente sobre blogs y recopilaciones de críticas de productos [17]. Las técnicas más extendidas se basan en explotar aspectos léxicos del texto, aún cuando el empleo de información sintáctica es vital para tratar aspectos como la negación, los *intensificadores* (formas lingüísticas que aumentan la intensidad de una opinión) y las frases adversativas (estructuras sintácticas que matizan o contradicen una opinión). Trabajos como el de Greene y Resnik [10] defienden la conexión entre la estructura del texto y el sentimiento implícito, siendo preciso enlazar adecuadamente la sintaxis y la semántica léxica para extraer el sentimiento de un texto. Por su parte, Guo y Wan [11] comentan que usar información sintáctica permite obtener mejor rendimiento a la hora de recuperar documentos de opinión con respecto a aquellos sistemas que sólo explotan la información léxica y la proximidad entre términos. Cruz et al. [6] muestran también que, para extraer opiniones estructuradas de artículos a partir de textos subjetivos con críticas de los mismos, es preciso capturar el contexto mediante técnicas de PLN como la lematización y la etiquetación morfosintáctica y, sobre todo, el análisis sintáctico basado en dependencias. Asimismo, las relaciones sintácticas también son útiles para determinar los conceptos contenidos en un microtexto para luego enlazarlos a fuentes de información relacionadas como, por ejemplo, artículos de la Wikipedia [19].

Por otra parte, hasta hace muy poco no se abordó la MO sobre microtextos, centrándose hasta ahora en determinar si un texto es de opinión o no, u obtener el nivel global de opinión de una expresión [18,24].

3.2. Nuestro enfoque

En nuestro caso, hemos tratado de sobreponernos a los inconvenientes, ya comentados, que presentan este tipo de tareas planteando un sistema de PLN sobre microtextos que consta de las siguiente fases:

1. Un *identificador del idioma* que reconoce el idiomas(s) en el que está escrita la entrada.
2. Según el acercamiento: (a) un *preprocesador* que trate los aspectos léxicos no estándar presentes en los textos; o (b) un *normalizador* que convierta el texto no estándar de entrada en estándar.
3. Un *segmentador* que divide el texto en frases y *tokens* (palabras, números, emoticonos, etc).
4. Un *etiquetador morfosintáctico* que permite obtener la información morfológica de cada palabra (género, número, tiempo, etc.) así como su forma canónica o *lema*.

5. Un *análizador sintáctico de dependencias* mediante el cual se identifican relaciones binarias *padre-dependiente* entre los términos de una frase (p.ej. entre el verbo y el objeto de la acción).

La información obtenida durante todo este proceso sirve para alimentar una serie de *clasificadores automáticos* para cada una de las tareas a realizar. Una vez obtenido el resultado, la información léxica y sintáctica nos permite explicar el resultado obtenido.

4. Experiencia adquirida

En los últimos años nuestro grupo de investigación ha participado en diversos foros, muchos de carácter competitivo, del ámbito de la MO, sobre todo en tareas sobre microtextos. Repasaremos brevemente a continuación parte de este trabajo, agrupándolo por temática:

4.1. Clasificación de la polaridad del sentimiento

Se trata de determinar si un tuit contiene opinión o no y, de ser así, si ésta es positiva, negativa o neutra. Nuestro enfoque se situó en los puestos de cabeza en la evaluación en español del TASS 2013 [25] con el corpus TASS General. El sistema mostró un comportamiento muy equilibrado para todas las categorías, aspecto clave en situaciones reales en las que no se conoce el sesgo de los datos.

En cuanto al inglés, nuestro sistema fue el mejor competidor español en el SemEval 2014 [26], que incluía tareas diferenciadas para tuits, SMSs y blogs, obteniendo resultados especialmente buenos para éstos últimos al clasificarse entre los primeros. Adicionalmente, en la denominada “tarea sorpresa” de detección de sarcasmo, obtuvimos unos resultados notables sin hacer ninguna adaptación específica.

4.2. Clasificación temática

De nuevo en la competición TASS 2013 [25], la tarea mencionada, de notable dificultad, consistió en determinar de qué tema(s) trataba cada uno de los tuits del corpus TASS General: cine, fútbol, economía, entretenimiento, literatura, música, política, deportes, tecnología y “otros”. En los resultados finales nuestro sistema obtuvo el mejor rendimiento.

4.3. Análisis político

La tarea de análisis político del, otra vez, TASS 2013 [25], consistió en determinar la orientación política de una lista de usuarios de Twitter, utilizando para ello el corpus TASS Political, compuesto por tuits en español publicados durante la campaña de las elecciones generales del 2011. Nótese que no se trataba de asociar cada usuario con un partido, sino de determinar si su ideario político se encuadraba en Izquierda, Derecho, Centro o Indefinido. Los resultados obtenidos nos situaron nuevamente entre los primeros.

4.4. Análisis de reputación

La tarea de *Reputation Dimensions Classification*, encuadrada en la competición RepLab 2014 [27], consistía en clasificar una serie de tuits relacionados con empresas del sector bancario y de la automoción, y escritos en inglés o español, en alguna de las siguientes dimensiones: rendimiento,

liderazgo, ciudadanía, gobernanza, lugar de trabajo, innovación, productos y servicios, u “otros”. Nuestros resultados nos situaron nuevamente entre los primeros.

4.5. *Ranking de autores*

También dentro del RepLab 2014 [27], se desarrolló una tarea de *Author Profiling* consistente en crear una lista ordenada de usuarios influyentes de los sectores de automoción y banca partiendo de sus perfiles de Twitter —tanto en español como inglés. Nuestro sistema nos situó en cabeza al tomar en consideración la descripción que el propio usuario proporcionaba en su perfil, ya que la información extraída del análisis lingüístico de dichos perfiles demostró tener un gran valor discriminativo.

4.6. *Normalización de tuits*

Como ya se apuntó, la *normalización* (de microtextos) consiste en tomar un texto escrito en lenguaje no estándar (con repeticiones de caracteres, emoticonos, ortografía obviada, eliminación de partículas y signos de puntuación, etc.) y convertirlo a lenguaje estándar. Tras haber participado con nuestro prototipo inicial en la tarea de normalización para español del Tweet-Norm 2013 [28] y haberlo rediseñado por completo para incrementar su flexibilidad, acabamos de realizar una primera adaptación del mismo al inglés de cara a nuestra participación en la competición ACL W-NUT 2015 [8].

4.7. *Identificación del idioma en Twitter*

Por último, nuestro grupo participó en la competición TweetLID 2014 [7] de identificación del idioma en tuits, centrada en los idiomas de mayor presencia en la Península: español, gallego, catalán, euskera, portugués e inglés. Para ello estudiamos la adaptación de herramientas ya existentes al nuevo contexto, un contexto especialmente complejo debido no sólo a que se trabajaba con tuits —textos extremadamente cortos y a menudo desnormalizados—, sino que había que sumarle la presencia de fenómenos de *code mixing* (mezcla de idiomas en una misma entrada), las influencias mutuas entre lenguas coexistentes en una misma población y la proximidad entre algunas de ellas (p.ej. gallego y portugués).

5. Conclusiones

Hemos presentado en este trabajo, por una parte, una introducción al uso que de los denominados *social media* hacen los extremistas y, por otra, una introducción a cómo se emplean técnicas de Procesamiento del Lenguaje Natural y Minería de Opiniones en la lucha contra tal amenaza.

A continuación, prestando especial atención al caso de Twitter, hemos repasado nuestra experiencia en numerosas tareas relacionadas con el análisis de contenidos en redes sociales empleando este tipo de técnicas. En ese contexto, nuestros sistemas se han comportado de manera muy satisfactoria en diversas competiciones tanto nacionales como internacionales.

Si bien nuestro trabajo realizado hasta la fecha no está aplicado directamente a tareas de Defensa y Seguridad, las técnicas y recursos desarrollados pueden ser adaptados, de una forma relativamente sencilla, para su empleo, entre otras, en las tareas de inteligencia descritas anteriormente. Asimismo,

aunque nuestra experiencia se ha centrado, sobre todo, en el español y el inglés, dicho trabajo es adaptable a otros idiomas como, por ejemplo, francés o árabe, siempre y cuando se pueda disponer de los recursos necesarios para dichos idiomas, tanto lingüístico-computacionales (p.ej. etiquetadores, analizadores sintácticos, etc.) como humanos (conocedores de la lengua).

Agradecimientos

Investigación financiada por Ministerio de Economía y Competitividad (FFI2014-51978-C2-2-R), Xunta de Galicia (R2014/034) y Ministerio de Educación, Cultura y Deporte (FPU13/01180).

Referencias

1. S. Bergsma, P. McNamee, M. Bagdouri, C. Fink, and T. Wilson. Language identification for creating language-specific Twitter collections. In *Proc. of 2nd Workshop on Language in Social Media*, pp. 65–74, 2012. ACL.
2. O. Biran, S. Rosenthal, J. Andreas, K. McKeown, and O. Rambow. Detecting influencers in written online conversations. In *Proc. 2nd Workshop on Language in Social Media*, pp. 37–45, 2012. ACL
3. J.P. Bjelopera and M.A. Randol. American Jihadist terrorism: Combating a complex threat. CRS Report for Congress, Congressional Research Service, Washington, DC, 2010.
4. J. Brynielsson, A. Horndahl, F. Johansson, L. Kaati, C. Martenson, and P. Svenson. Analysis of Weak Signals for Detecting Lone Wolf Terrorists. In *Proc. of 2012 European Intelligence and Security Informatics Conference (EISIC 2012)*, pp. 197–204, 2012.
5. K. Cohen, F. Johansson, L. Kaati, and J.C. Mork. Detecting linguistic markers for radical violence in social media. *Terrorism and Political Violence*, 26(1):246–256, 2014.
6. F. Cruz, J. Troyano, F. Enríquez, J. Ortega, and C.G. Vallejo. A knowledge-rich approach to feature-based opinion extraction from product reviews. In *Proc. of 2nd Int. Workshop on Search and Mining User-generated Contents (SMUC'10)*, pp. 13–20, 2010. ACM.
7. Y. Doval, D. Vilares and J. Vilares. Identificación Automática del Idioma en Twitter: Adaptación de Identificadores del Estado del Arte al Contexto Ibérico. in A. Zubiaga *et al.* (eds.), *TweetLID 2014. Tweet Language Identification Workshop 2014*. Vol 1228 of *CEUR Workshop Proceedings*, pp. 39–43. CEUR-WS.org, 2014.
8. Y. Doval, J. Vilares and C. Gómez-Rodríguez. LYSGROUP: Adapting a Spanish microtext normalization system to English. in *ACL 2015 Workshop on Noisy User-generated Text. Proceedings of the Workshop*, pp. 99–105, 2015. ACL.
9. E. Ferrara. Manipulation and abuse on social media. *ACM SIGWEB Newsletter*, 4:1–4:9, 2015.
10. S. Greene and P. Resnik. More than words: Syntactic packaging and implicit sentiment. In *Proc. of NAACL-HLT 2009*, pp. 503–511, 2009. ACL.
11. L. Guo and X. Wan. Exploiting syntactic and semantic relationships between terms for opinion retrieval. *JASIST*, 63(11):2269–2282, 2012.
12. F. Johansson, L. Kaati, and A. Shrestha. Detecting multiple aliases in social media. In *Proc. of Int. Conf. on Advances in Social Networks Analysis and Mining*, pp. 1004–1011. ACM, 2013.

13. J. Jordán. Procesos de radicalización yihadista en España: Análisis sociopolítico en tres niveles. *Revista de Psicología Social*, 24(2):197–216, 2009.
14. D. Jurafsky and J.H. Martin. *Speech and Language Processing (2nd Ed.)*. Pearson, 2009.
15. G. Katsimpras, D. Vogiatzis, and G. Paliouras. Determining influential users with supervised random walks. In *Proc. of 24th Int. Conf. on World Wide Web Companion*, pp. 787–792, 2015.
16. G. L’Huillier, S.A. Ríos, H. Alvarez, and F. Aguilera. Topic-based social network analysis for virtual communities of interests in the Dark Web. In *Proc. of ACM SIGKDD Workshop on Intelligence and Security Informatics (ISI-KDD’10)*, pp. 9:1–9:9. ACM, 2010.
17. B. Liu. *Sentiment Analysis and Opinion Mining*, vol. 16 of *Synthesis Lectures on Human Language Technologies*. Morgan & Claypool Publishers, 2012.
18. E. Martínez-Cámara, M.T. Martín-Valdivia, L. Alfonso Ureña López, and A. Montejo-Ráez. Sentiment analysis in Twitter. *Natural Language Engineering*, 20(1):1–28, 2014.
19. E. Meij, W. Weerkamp, and M. de Rijke. Adding semantics to microblog posts. In *WSDM 2012: 5th ACM Int. Conf. on Web Search and Data Mining*, 2012. ACM.
20. M. Munezero, T. Kakkonen, and C. Montero. Towards automatic detection of antisocial behavior from texts. In *Proc. of IJCNLP Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011)*, pp. 20–27, 2011. Asian Federation of NLP.
21. S.A. Ríos and R. Muñoz. Dark Web portal overlapping community detection based on topic models. In *Proc. of ACM SIGKDD Workshop on Intelligence and Security Informatics (ISI-KDD’12)*, pp. 2:1–2:7, 2012. ACM.
22. A. Semenov, A. Nikolaev, and J. Veijalainen. Online activity traces around a Boston Bomber. In *Proc. of 2013 IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 1050–1053, 2013.
23. A. Semenov, J. Veijalainen, and A. Boukhanovsky. A generic architecture for a social network monitoring and analysis system. In *Proc. of 14th Int. Conf. on Network-Based Information Systems (NBIS 2011)*, pp. 178–185. IEEE, 2011.
24. M. Thelwall, K. Buckley, and G. Paltoglou. Sentiment strength detection for the social web. *JASIST*, 63(1):163–173, 2012.
25. D. Vilares, M. A. Alonso, and C. Gómez-Rodríguez. LyS at TASS 2013: Analysing Spanish tweets by means of dependency parsing, semantic-oriented lexicons and psychometric word-properties. *TASS 2013 — Workshop on Sentiment Analysis at SEPLN 2013*, pp. 179–186. 2013.
26. D. Vilares, M. Hermo, M. A. Alonso, C. Gómez-Rodríguez and Y. Doval. LyS: Porting a Twitter Sentiment Analysis Approach from Spanish to English. *The 8th International Workshop on Semantic Evaluation (SemEval 2014). Proceedings of the Workshop*, pp. 411–415. 2014.
27. D. Vilares, M. Hermo, M. A. Alonso, C. Gómez-Rodríguez and J. Vilares. LyS at CLEF RepLab 2014: Creating the State of the Art in Author Influence Ranking and Reputation Classification on Twitter. in L. Cappellato, N. Ferro, M. Halvey, W. Kraaij (eds.), *CLEF 2014 Working Notes*. Vol 1180 of *CEUR Workshop Proceedings*, pp. 1468–1478. CEUR-WS.org, 2014.
28. J. Vilares, M. A. Alonso and D. Vilares. Prototipado Rápido de un Sistema de Normalización de Tuits: Una Aproximación Léxica. in I. Alegria et al. (eds.), *Tweet-Norm 2013. Tweet Normalization Workshop 2013*. Vol. 1086 of *CEUR Workshop Proceedings*, pp. 39–43. CEUR-WS.org, 2014.