

ADRSpanishTool: una herramienta para la detección de efectos adversos e indicaciones

ADRSpanishTool: a tool for extracting adverse drug reactions and indications

Santiago de la Peña*, Isabel Segura-Bedmar*, Paloma Martínez*, José Luis Martínez**

*Universidad Carlos III de Madrid, Av. Universidad, 30, 28911, Madrid

**Daedalus SA, Avda. de la Albufera 321, 28031, Madrid

spena@pa.uc3m.es, {isegura,pmf}@inf.uc3m.es, jmartinez@daedalus.es

Resumen: Presentamos una herramienta basada en coocurrencias de fármaco-efecto para la detección de reacciones adversas e indicaciones en comentarios de usuarios procedentes de un foro médico en español. Además, se describe la construcción automática de la primera base de datos en español sobre indicaciones y efectos adversos de fármacos.

Palabras clave: Extracción de Información, Medios Sociales

Abstract: We present a tool based on co-occurrences of drug-effect pairs to detect adverse drug reactions and drug indications from user messages that were collected from an online Spanish health forum. In addition, we also describe the automatic construction of the first Spanish database for drug indications and adverse drug reactions.

Keywords: Information Extraction, Social Media

1 Introducción

El objetivo principal de la Farmacovigilancia consiste en estudiar el uso y los efectos de los medicamentos en los pacientes. Además, se encarga de generar alertas sobre posibles reacciones adversas a un medicamento, intentando evaluar el riesgo, para finalmente informar a los profesionales sanitarios y a los pacientes con el fin de evitar estas reacciones adversas. En los últimos años, la Farmacovigilancia ha ganado cada vez más interés debido al creciente número de reacciones adversas (Bond y Raehl, 2006) y el coste asociado de estas reacciones (van Der Hooft et al., 2006).

Hoy en día, los principales organismos regulatorios de medicamentos, como son la “Food and Drug Administration” (FDA) en EE.UU o la Agencia de Medicina Europea (EMA), trabajan activamente en el diseño de políticas e iniciativas dirigidas a facilitar la notificación de reacciones adversas a medicamentos por los profesionales sanitarios y por los pacientes. Sin embargo, varios estudios han demostrado que la notificación de estas reacciones adversas es aún poco frecuente debido a que muchos profesionales de la salud no tienen tiempo suficiente para utilizar los

sistemas de notificación de reacciones adversas (Bates et al., 2003; van Der Hooft et al., 2006; McClellan, 2007). Además, los profesionales sanitarios sólo informan de aquellas reacciones de las que tienen certeza absoluta de su existencia (Herxheimer, Crombag, y Alves, 2010). A diferencia de los informes de los profesionales de la salud, los informes de los pacientes a menudo proporcionan información más detallada y explícita sobre las reacciones adversas. Sin embargo, la tasa de reacciones notificadas por los pacientes es todavía muy baja, probablemente debido a que muchos pacientes todavía no son conscientes de la existencia de estos sistemas de notificación.

Nuestra hipótesis principal es que la información que los pacientes escriben en distintos medios sociales, como por ejemplo los foros sobre temas de salud, pueden ser un recurso complementario para los sistemas de notificación de reacciones adversas.

En los últimos años, se han desarrollado varios sistemas dedicados a la detección de reacciones adversas en medios sociales (Leaman et al., 2010; Nikfarjam y Gonzalez, 2011), sin embargo ninguno de estos trabajos ha tratado el problema en español.

En este artículo, presentamos una herramienta para la detección automática de efectos adversos e indicaciones de fármacos en mensajes de usuarios de ForumClinic¹, una plataforma social donde los pacientes intercambian información sobre sus enfermedades y sus tratamientos.

2 Descripción de la herramienta ADRSpanishTool

La infraestructura para desarrollar y desplegar los componentes de la herramienta que se ha usado ha sido GATE². Para procesar los mensajes, hemos utilizado la herramienta de Textalytics³, que sigue un enfoque basado en diccionario para identificar los fármacos y sus efectos. Para la construcción del diccionario, utilizamos los siguientes recursos: CIMA⁴ y MedDRA⁵. CIMA es una base de datos online administrada por la Agencia Española de Medicamento y productos sanitarios (AEMPS) que contiene información sobre todos los fármacos aprobados en España. MedDRA es un recurso multilingüe que proporciona información sobre eventos adversos de medicamentos. El diccionario contiene un total de 5.800 fármacos y 13.246 efectos adversos con 48,632 sinónimos distintos. Además, se han creado varios gazetteers para aumentar la cobertura en la detección de fármacos y efectos proporcionada por el diccionario. En concreto, hemos desarrollado varios crawlers para buscar y descargar páginas relacionadas con fármacos de sitios webs como MedLinePlus⁶ y Vademecum⁷. Mediante el uso de expresiones regulares aplicadas sobre las páginas descargadas, conseguimos obtener de forma automática una lista de fármacos y de efectos. El sistema ATC⁸, para la clasificación de fármacos, también fue utilizado y volcado en un gazetteer para detectar nombre de grupos de fármacos.

GATE proporciona una herramienta de anotación de patrones denominada JAPE (Java Annotation Patterns Engine). Es una versión del Common Pattern Specification Language (CPSL) que mediante distintas fa-

ses, cada una de las cuales conteniendo unas reglas patrón/acción, constituyen una cascada de transductores de estados finitos que actúan sobre las anotaciones. En nuestro sistema se usa esta herramienta para filtrar las anotaciones procedentes del diccionario de Textalytics y separar las procedentes de los Gazetteers.

Aunque hay varias bases de datos con información sobre fármacos y sus efectos, como SIDER⁹ o MedEffect¹⁰, ninguna está disponible en español. Además, estos recursos no incluyen indicaciones. Así, hemos construido de manera automática la primera base de datos disponible en español, *SpanishDrugEffectBD*, con información sobre fármacos, sus indicaciones y sus reacciones adversas. Esta base de datos se puede usar para identificar de manera automática indicaciones y reacciones adversas en textos. La figura 1 muestra el esquema de la base de datos. El primer paso fue poblar la base de datos con los fármacos y efectos de nuestro diccionario. Los ingredientes activos se almacenan en la tabla *Drug*, mientras que sus sinónimos y nombres comerciales en la tabla *DrugSynset*. Igualmente, se almacenan los conceptos obtenidos de MedDRA en la tabla *Effect* y sus sinónimos en la tabla *EffectSynset*.

Para obtener las relaciones entre los fármacos y sus efectos, desarrollamos varios crawlers para descargar los apartados sobre indicaciones y reacciones adversas de prospectos contenidos en las siguientes webs: MedLinePlus, Prospectos.Net¹¹ y Prospectos.org¹². Una vez descargados estos apartados, fueron procesados usando la herramienta Textalytics para etiquetar los fármacos y sus efectos. Los efectos descritos en el apartado de indicaciones de un fármaco fueron almacenados como relaciones de tipo indicación. De forma similar, los efectos descritos en los apartados de reacciones adversas fueron almacenados como relaciones de tipo reacción adversa.

La herramienta utiliza un enfoque basado en coocurrencia de entidades para extraer las relaciones. Además, mediante la consulta a la base de datos se comprueba si una relación corresponde a una indicación o a un efecto

¹<http://www.forumclinic.org/>

²<http://gate.ac.uk/>

³<https://textalytics.com/>

⁴<http://www.aemps.gob.es/cima/>

⁵<http://www.meddra.org/>

⁶<http://www.nlm.nih.gov/medlineplus/spanish/>

⁷www.vademecum.es

⁸<http://www.whocc.no/atc/>

⁹<http://sideeffects.embl.de/>

¹⁰<http://www.hc-sc.gc.ca/dhp-mps/medeff/index-eng.php>

¹¹<http://www.prospectos.net/>

¹²<http://prospectos.org/>

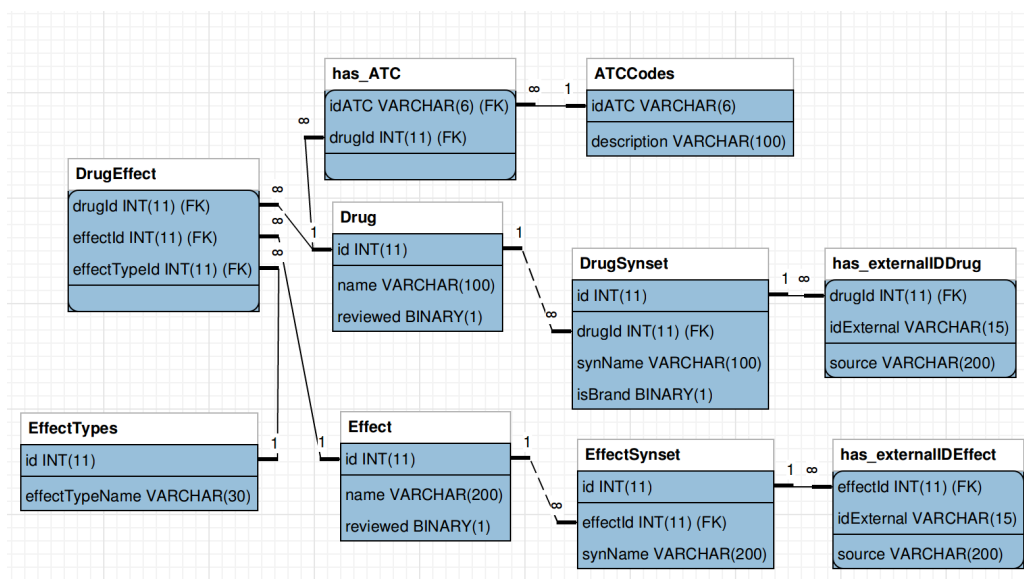


Figura 1: Esquema de la base de datos SpanishDrugEffectBD

adverso. Creamos una aplicación de Gate¹³ en la que integramos el módulo de Textalytics, los gazetteers y la extracción de relaciones fármaco-efecto usando la base de datos. En la figura 2 puede verse una representación de la aplicación.

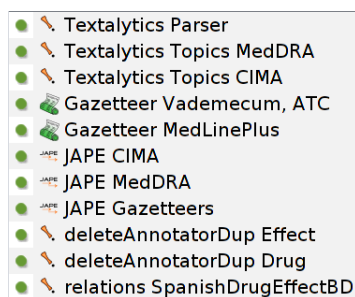


Figura 2: Aplicación completa para anotación de relaciones fármaco-efecto en GATE

El lector puede encontrar una descripción más detallada de los recursos y procesos utilizados por la herramienta en el trabajo (Segura-Bedmar, Revert, y Martínez, 2014).

Mediante la tecnología Jetty¹⁴ implementada en el GATE se desplegó un servidor web para ejecutar la aplicación y poder visualizar los resultados. En la figura 3 se puede ver un ejemplo de salida de la aplicación.

La herramienta ha sido evaluada sobre el corpus SpanishADR¹⁵, el primer corpus en Español formado por 400 comentarios. Estos

comentarios fueron obtenidos de ForumClinic y anotados por dos anotadores. El corpus contiene un total de 188 fármacos y 545 efectos. Además, se han anotado 61 indicaciones y 103 reacciones adversas. La herramienta muestra una precisión del 86 % y una cobertura del 81 % en la detección de fármacos. La detección de efectos es una tarea más difícil debido al gran número de expresiones idiomáticas que los pacientes utilizan para describir sus experiencias con los medicamentos. La herramienta sólo consigue un 63 % de precisión y un 51 % en la detección de entidades de tipo efecto. Respecto a los resultados en la extracción de relaciones fármaco-efecto, la herramienta consigue una buena precisión de un 83 % pero con muy baja cobertura 15 %.

3 Conclusión y Trabajo Futuro

En este artículo presentamos la primera herramienta dedicada a la detección de indicaciones y reacciones adversas en mensajes de usuario obtenidos de un foro sobre salud en español. Una demo de la herramienta está disponible en el sitio web: <http://163.117.129.57:8090/gate/>.

También se describe en el trabajo la creación de una base de datos de indicaciones y reacciones adversas obtenidos de forma automática de prospecto sobre fármacos. Esta es la primera base de datos disponible en español. Aunque su uso no mejora los resultados por su limitada cobertura, pensamos que puede ser un recurso valioso para futuros

¹³<http://gate.ac.uk/>

¹⁴<http://www.eclipse.org/jetty/>

¹⁵<http://labda.inf.uc3m.es/SpanishADRCorpus>

Results - Drugs: Indications and Adverse Effects

Text Annotated

DEPR357 segun muchos prospectos no se deben pasar 21 días (3 semanas). . el **alprazolam** es precisamente el ansiolítico que mas dependencia genera por su alta potencia y vida media corta. . pero si tuviera que darte mi opinion, te diria que depende de la "adiccion a la **ansiedad**" que tengas. si andas medio tranquilo y bien, la dependencia no es tanta y los puedes soltar con facilidad; si tu ansiedad es altísima no los vas a poder dejar. . un saludo vecino

TRBI1122 hola anais. . finalmente conseguí hablar con mi psiquiatra habitual y me dijo que el **vertigo** era un efecto secundario de la **lamotrigina**. le comente que lo había suspendido. así esta informado de mi situacion

Drug **Effect** **Indication** **AdverseEffect**

Figura 3: Ejemplo de la salida producida por la aplicación

desarrollos.

Este trabajo se enmarca dentro del proyecto europeo TrendMiner¹⁶, siendo uno de sus principales objetivos incrementar la cobertura multilingüe de herramientas y ontologías en el dominio médico.

En un futuro nos planteamos aumentar el tamaño del corpus para poder aplicar técnicas de aprendizaje automático. También nos gustaría crear un diccionario con vocabulario y expresiones idiomáticas que los pacientes utilizan a la hora de describir sus tratamientos y efectos. Este tipo de recursos puede ser muy útil a la hora de desarrollar sistemas de detección de efectos adversos. También nos planteamos transformar la base de datos en una ontología y poblarla con más conceptos y relaciones.

Agradecimientos

Este trabajo se ha financiado por el proyecto europeo TrendMiner [FP7-ICT287863], por el proyecto MULTIMEDICA [TIN2010-20644-C03-01], y por la Red de Investigación MA2VICMR [S2009/TIC-1542].

Bibliografía

Bates, DW., RS. Evans, H. Murff, PD. Stetson, L. Pizziferri, y G. Hripcsak. 2003. Detecting adverse events using information technology. *Journal of the American Medical Informatics Association*, 10(2):115–128.

Bond, CA. y CL. Raehl. 2006. Adverse drug reactions in United States hospitals. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*, 26(5):601–608.

Herxheimer, A., MR. Crombag, y TL. Alves. 2010. Direct patient reporting of adverse drug reactions. a twelve-country survey & literature review. *Health Action International (HAI)(Europe)*. Amsterdam.

Leaman, R., L. Wojtulewicz, R. Sullivan, A. Skariah, J. Yang, y G. Gonzalez. 2010. Towards internet-age pharmacovigilance: extracting adverse drug reactions from user posts to health-related social networks. En *Proceedings of the 2010 workshop on biomedical natural language processing*, páginas 117–125. Association for Computational Linguistics.

McClellan, M. 2007. Drug safety reform at the fda—pendulum swing or systematic improvement? *New England Journal of Medicine*, 356(17):1700–1702.

Nikfarjam, A. y GH. Gonzalez. 2011. Pattern mining for extraction of mentions of adverse drug reactions from user comments. En *AMIA Annual Symposium Proceedings*, volumen 2011, página 1019. American Medical Informatics Association.

Segura-Bedmar, I., R. Revert, y P. Martínez. 2014. Detecting drugs and adverse events from Spanish social media streams. En *Proceedings of the 5th International Louhi Workshop on Health Document Text Mining and Information Analysis (Louhi 2014)*.

van Der Hooft, CS., MCJM. Sturkenboom, K. van Grootheest, HJ. Kingma, y BHCh. Stricker. 2006. Adverse drug reaction-related hospitalisations. *Drug Safety*, 29(2):161–168.

¹⁶<http://www.trendminer-project.eu/>