

Enriquecimiento Automático de Ontologías Biomédicas mediante el uso de Mappings

María Jesús García-Godoy, Esteban López-Camacho, María del Mar Roldán-García, José F. Aldana-Montes

Khaos Research Group, Departamento de Lenguajes y Ciencias de la Computación, Universidad de Málaga

Resumen Dione es una representación en OWL del ICD-10-CM, consistente lógicamente, cuyos axiomas definen las inclusiones y exclusiones del ICD-10-CM mediante una metodología basada en los *mappings* ICD-10-CM/SNOMED-CT, proporcionados por UMLS y BioPortal, y que han sido validados por una comunidad de expertos en el ámbito biomédico. En este artículo se presenta una metodología automática que permite la población con axiomas en Dione a partir de los *mappings* establecidos entre ICD-10-CM y otra ontología biomédica proporcionados por BioPortal. Para mostrar su funcionamiento, se han utilizado los *mappings* entre Dione y ORDO. Esta última es una ontología que incluye el conjunto de enfermedades raras, genes y otras características para la población de nuevos axiomas en Dione. Una vez que estos axiomas se incluyeron en Dione, se comprobó su consistencia utilizando el razonador ELK y se mostró con un caso de uso que las clases equivalentes entre las ontologías DIONE-ORDO permitían la inferencia de axiomas que definen una clase ICD-10-CM en DIONE a una clase que representa una enfermedad rara en ORDO y, viceversa. Esta metodología es aplicable a dos ontologías biomédicas cualquiera cuyos *mappings* estén definidos en BioPortal.

1. Introducción

Dione es una representación en OWL del ICD-10-CM (Clasificación Internacional de Enfermedades, versión 10, Modificaciones Clínicas), consistente lógicamente, cuyos axiomas definen las inclusiones y exclusiones del ICD-10-CM mediante el uso de una metodología basada en los *mappings* SNOMED-CT/ICD-10-CM [6], [3], [8]. Dione está compuesta de 391.669 clases, 391.720 axiomas de anotación y 11.795 axiomas de tipo owl:equivalentClass, las cuales se han generado mediante 104.646 relaciones extraídas de los *mappings* oficiales SNOMED CT/ICD-10-CM proporcionados por UMLS [5] y BioPortal [2]. En esta línea de investigación desarrollada por los autores se propuso como trabajo futuro la inclusión de *mappings* a partir de otras ontologías biomédicas con el objetivo de relacionar Dione con éstas.

En este trabajo emergente, se propone una metodología para la inclusión de forma automática de *mappings* entre Dione y una ontología biomédica proporcionada por BioPortal. Como ejemplo se ha utilizado la ontología ORDO

(Orphanet Rare Disease Ontology) que ha sido desarrollada por Orphanet y el Instituto de Bioinformática Europeo y que proporciona un vocabulario estructurado en enfermedades raras, genes y otras características que son una fuente importante para el análisis biocomputacional de enfermedades raras [4]. Se han incluido los *mappings* en axiomas de tipo owl:EquivalentClass relacionando una clase perteneciente a Dione con la clase equivalente en ORDO. Se ha determinado que Dione es consistente y se ha presentado la utilidad de este método en la enfermedad del virus Zika [7] y en la actinomicosis [9]. Los axiomas previamente incluidos en Dione que definen cada categoría del ICD-10-CM son inferidos en la clase equivalente en la ontología ORDO y viceversa.

2. Metodología

BioPortal es un repositorio de ontologías biomédicas que ofrece una aplicación y varios servicios Web para explorar y obtener información de las ontologías almacenadas. BioPortal ofrece una API REST documentada [1] mediante la cual los usuarios pueden acceder a su información, siendo posible extraer los *mappings* entre una ontología objetivo y el resto de ontologías biomédicas del repositorio. Los *mappings* incluidos en BioPortal son validados de forma semiautomática por el algoritmo LOOM y la comunidad de usuarios.

Dado que Dione es una representación en OWL del ICD-10-CM, haciendo uso del servicio REST de BioPortal, se obtuvieron todos los *mappings* de las ontología ICD-10-CM (una jerarquía sobre las categorías del ICD-10-CM que no están formalmente definidas a diferencia de Dione cuyas clases están definidas mediante los *mappings* ICD-10-CM/SNOMED-CT) y la ontología de enfermedades raras ORDO. El número de *mappings* entre ambas ontologías que se obtuvo fue de 349. Haciendo uso de la librería Java OWL-API, se cargaron ambas ontologías y se combinaron en una sola. Posteriormente, se creó un axioma tipo owl:equivalentClass usando cada par de clases (correspondientes a las ontologías de DIONE y ORDO) que constituían los 349 mappings obtenidos de BioPortal. El resultado fue exportado a una nueva ontología que reuniera la información de cada una de las ontologías por separado y las equivalencias creadas a partir de los *mappings*¹.

Dado que la cantidad de información de Bioportal se incrementa en el tiempo, y continuamente se incluyen nuevos *mappings* entre los conceptos de las ontologías disponibles, todo este proceso se desarrolló en una aplicación Java ² para ser usado con distintos pares de ontologías, o para actualizar con nuevos *mappings* las ontologías ya creadas. Esta metodología automática contribuye a la facilidad de creación de nuevas ontologías con un mayor número de interconexiones entre conceptos, permitiendo la realización de razonamientos más complejos.

¹ <http://khaos.uma.es/dione/PublicDownload?file=DIONE-ORDO.zip>

² Código disponible para usuario y desarrolladores (previo registro a la plataforma Bitbucket.org): <https://mjggodoy@bitbucket.org/estebanpua/bioportal-mappings.git>

3. Caso de Uso

Para la realización del caso de uso fue necesario la comprobación de la consistencia de la ontología Dione poblada con los *mappings* de ORDO. Para ello, se utilizó el razonador ELK mediante la OWL-API. Una vez que se comprobó que la ontología es consistente, se procedió a la clasificación de Dione mediante uso del mismo razonador en un PC Intel(R) Core (TM) i7-2600 CPU con 3.39 GHz y 16 GB de RAM.

Tras la clasificación, se observó la creación de nuevas clases que no existían en Dione presentes en los *mappings* de ORDO-ICD-10-CM como es el caso de la clase A92.5 que corresponde a la “Enfermedad del virus del Zika”. La clase A92.5 de Dione corresponde a la clase Orphanet_448237 (enfermedad del virus del Zika) de la ontología ORDO. En este caso, no sólo se ha añadido una clase nueva en Dione sino que los axiomas de la clase de ORDO, han sido heredados por la clase A92.5. Algunos de estos axiomas que definen estas dos clases equivalentes y proceden de la ontología ORDO son los siguientes: *parte de* “Enfermedad viral rara”, *tiene un rango de incidencia de* “1-5/10000” y *está presente en* “Brasil” y *tiene una incidencia anual de* 58.8f.

Otro ejemplo destacable en este trabajo es la enfermedad de la Actinomicosis. En Dione, esta enfermedad se corresponde con la categoría A42 (rdf:label corresponde a Actinomicosis) y en ORDO con la categoría Orphanet_457095. En este experimento preliminar se ha creado una instancia “Enfermedad” en la que se definieron dos aserciones propiedad objeto: *agente causativo* “Genus Actinomyces”, *cursa* “Infección” y *afecta* “a todas las edades”. El razonador ELK clasificó la instancia en la clase A42 de Dione y, además, al definir los *mappings* Dione-ORDO se infirió que esa misma instancia clasificada en A42 también se clasificó en la clase Orphanet_457095. La instancia se clasificó en ambas clases cuyos axiomas que las definen se complementan (Fig 1).

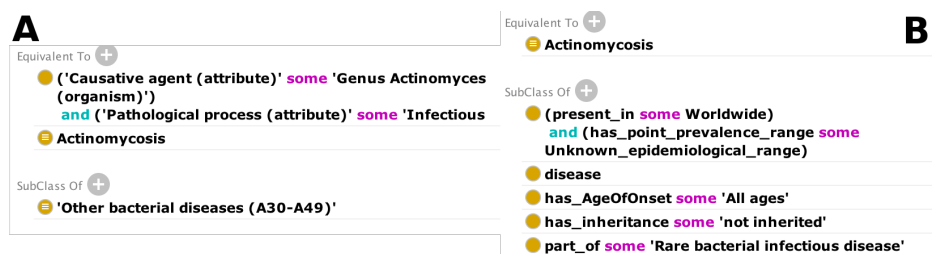


Figura 1. Fig.1A muestra los axiomas que definen la clase A42 de Dione junto con el *mapping* a la enfermedad rara Orphanet_457095 en ORDO. Fig.1B muestra los axiomas pertenecientes a la clase _457095. Los axiomas de ambas clases complementan la definición de la enfermedad Actinomicosis

4. Conclusión

En este trabajo se ha presentado un procedimiento automático de enriquecimiento de ontologías usando los *mappings* proporcionados por BioPortal. Estos *mappings* disponibles son validados y además crecen en número periódicamente, lo que hace útil su uso para aumentar las interconexiones entre conceptos entre ontologías. Mediante nuestro proceso automático es posible mantener actualizadas las ontologías con los últimos *mappings* incluidos en Bioportal. Las ontologías actualizadas permiten un conjunto de razonamientos más complejos que serían imposibles sin la inclusión de estos *mappings*.

Como trabajo futuro, se pretende automatizar aún más el proceso para que la actualización de una ontología se realice periódicamente con una frecuencia establecida por un usuario. Se prevé también extender nuestro caso de uso interconectando más de dos ontologías entre sí y estudiar los resultados de los razonamientos que son posibles a realizar en casos más complejos. En adición, se plantea analizar el impacto del algoritmo LOOM proporcionado por Biportal en el emparejamiento léxico entre conceptos de ontologías biomédicas y detectar qué efecto tiene en los axiomas generados a partir de los *mappings*.

Agradecimientos

Este trabajo ha sido parcialmente financiado por los Proyectos TIN2014-58304 y TIN2017-86049-R (Ministerio de Economía, Industria y Competividad), y P12-TIC-1519 (Plan Andaluz de Investigación, Desarrollo e Innovación).

Referencias

1. Bioportal API documentation. <http://data.bioontology.org/documentation>, accessed: 13-March-2018
2. Bioportal website. <http://bioportal.bioontology.org/>, accessed: 13-March-2018
3. ICD-10-CM/SNOMED-CT, Mappings Technical Specifications. http://www.nlm.nih.gov/research/umls/mapping_projects/snomedct_to_icd10cm_tech_spec_201312.pdf, accessed: 13-March-2018
4. ORDO ontology. <https://www.ebi.ac.uk/ols/ontologies/ordo>, accessed: 13-March-2018
5. SNOMED CT release files. <http://www.nlm.nih.gov/research/umls/>, accessed: 13-March-2018
6. SNOMED CT to ICD-10-CM Map. [http://www.orpha.net/consor/cgi\\$-\\$bin/index.php](http://www.orpha.net/consor/cgi$-$bin/index.php), accessed: 13-March-2018
7. Hills, S.L., Fischer, M., Petersen, L.R.: Epidemiology of zika virus infection. *The Journal of Infectious Diseases* **216**(10), 868–874 (2017). <https://doi.org/10.1093/infdis/jix434>
8. Roldán-García, M.d.M., García-Godoy, M.J., Aldana-Montes, J.F.: Dione: An owl representation of icd-10-cm for classifying patients' diseases. *Journal of Biomedical Semantics* **7**(1), 62 (2016). <https://doi.org/10.1186/s13326-016-0105-x>
9. Valour, F., Sénéchal, A., Dupieux, C., Karsenty, J., Lustig, S., Breton, P., Gleizal, A., Bousset, L., Laurent, F., Braun, E., Chidiac, C., Ader, F., Ferry, T.: Actinomycosis: Etiology, clinical features, diagnosis, treatment, and management **7**, 183–97 (2014)