



Universidad del Cauca
Facultad de Ingeniería Electrónica y Telecomunicaciones

Programas de Maestría y Doctorado en Ingeniería Telemática
Seminario de Investigación

Towards Automated Biomedical Ontology Harmonization

Gustavo Andrés Uribe Gómez

Estudiante de Doctorado

15 de Abril de 2014

1- Introducción

En esta sesión del seminario se presenta el artículo titulado “Towards Automated Biomedical Ontology Harmonization” presentado en la conferencia internacional pHealth2014 y publicado en la revista “Studies in health technology and informatics” homologada por Colciencias como A2. El presentador del seminario es el primer autor de este artículo. El segundo autor es el PhD. Diego Mauricio López de la Universidad del Cauca y el tercer autor es el PhD. Bernd Blobel de la Universidad de Regensburg, Alemania.

2- Problema

El artículo se centra en el problema de la falta de interoperabilidad o colaboración entre actores heterogéneos. Un ejemplo de este problema se presentó en la historia de la torre de babel donde no se logró alcanzar el objetivo de construir la torre hasta el cielo dado que cada actor no podía interpretar la comunicación de los otros actores. El problema de esta historia se repite en cada entorno de colaboración dado que el conocimiento, experiencias y las habilidades de cada actor son diferentes, lo que siempre dificulta el alcanzar los objetivos propuestos. En el dominio de la medicina este hecho es más visible dado la heterogeneidad de los actores en colaboración. Por ejemplo estos pueden ser de distintas especialidades médicas u enfoques. El artículo presentado muestra una serie de pasos que deben cumplirse con el fin de soportar la colaboración entre actores por medio de sistemas computacionales y propone la armonización automática de ontologías como una clave fundamental en el proceso.

3- Resultados

El primer paso para resolver el problema es resolver la colaboración entre dos actores. Estos tienen un objetivo en común que para poderlo alcanzar necesitan interpretar datos del entorno obteniendo información que les ayudará a realizar las acciones apropiadas. El conocimiento que tienen los actores es lo que les permite realizar todo este proceso. Si dos actores comparten el mismo conocimiento

entonces un sistema computacional solo deberá transmitir los datos desde un actor a otro. A este nivel de interoperabilidad se le llama interoperabilidad técnica, estructural y sintáctica. En el caso de que los actores no tengan el mismo conocimiento entonces el sistema computacional debe ayudar a interpretar los datos del entorno y del proceso. A este nivel de interoperabilidad se le conoce como interoperabilidad semántica. Adicionalmente, el sistema computacional debe soportar a los actores en el caso que no sean capaz de realizar la acción apropiada con la información derivada de los datos. Este nivel de interoperabilidad se le llama interoperabilidad organizacional. Un sistema computacional debe permitir todos estos niveles de interoperabilidad para poder lograr la colaboración entre dos actores. Sin embargo pretender un sistema que logre esto para cualquier pareja de actores no es viable, tal sistema sería muy complejo. Es por eso que se ve la necesidad de definir estándares o acuerdos que permitan disminuir la complejidad de tal sistema. Estos estándares disminuyen la heterogeneidad del sistema, sin embargo no la eliminan completamente pues no se puede garantizar que en un entorno de colaboración todos se ajusten a un estándar. No se puede obviar tampoco que los estándares se traslapan unos a otros y presentan diferentes versiones, lo que no permite simplificar más el sistema. De esto podemos concluir que un sistema computacional soportando la colaboración debe ser lo suficientemente flexible y adaptable que permita manejar diversos estándares sin la necesidad de realizar grandes esfuerzos en codificación y en el desarrollo de un nuevo sistema.

Como ya lo habíamos mencionado estos sistemas deben manejar el conocimiento de los actores involucrados y esto implica que tal conocimiento debe formalizarse. Las ontologías permiten formalizar el conocimiento ya establecido y dan la capacidad a los sistemas computacionales de realizar inferencias sobre tal conocimiento [1]. Las ontologías resultan también en un esfuerzo de estandarización y por ende encontramos los mismos problemas, es decir, variedad de ofertantes, traslapes y diferentes versiones. Para poder dar solución a esta problemática debemos realizar la armonización de las diversas ontologías.

Existen cuatro mecanismos de armonización de ontologías. El primero de ellos es la fusión de ontologías, en donde las clases o conceptos simplemente se unen en una sola ontología. El segundo método es el alineamiento de ontologías en donde se seleccionan clases y conceptos de las ontologías para formar una nueva que va a soportar la colaboración. Estos dos métodos dan como resultado una nueva ontología que soporta la colaboración, mientras que las ontologías fuente no intervienen en la colaboración. En los siguientes dos métodos las ontologías fuente se siguen usando y se crea una nueva ontología que muestra las equivalencias entre dichas ontologías (emparejamiento) o las semejanzas (mapeo).

Existen actualmente herramientas software que soportan estos mecanismos [2]. Estas herramientas presentan una serie de problemas que aún son brechas de investigación [3]. Se realizó un análisis de varias de estas herramientas. En la figura 1 se muestran parte de los resultados.

Tool	Architectural Viewpoint	Type of Matching	Input Format	Output Format	Functionality
MOMIS	No	Syntactic Structure-based	ER-S	XML Schema	Integration
Prompt	No	Syntactic Structure-based	OWL, RDF	OWL RDF	Compare Merging Matching
OntoBuilder	No	Syntactic Structure-based	XML Schema Web Form RDF, OWL, WSDL	XML Schema BizTalk RDF	Creation Matching Merging
COMA 3.0	No	Syntactic Structure-based	ER-S XML Schema RDF, OWL	Text File	Matching
S-Match	No	Syntactic Semantic	XML Schema OWL	Text File	Mapping
RiMOM	No	Syntactic Structure-based	OWL	RDF	Matching
Harmony (OpenII)	No	Syntactic Structure-based	ER-S, EEER XML Schema OWL, XMI	XML Schema	Integration Matching

Figure 1: Analisis herramientas de armonización de ontologías

Se pudieron encontrar varias falencias como el hecho de que ninguno de estos sistemas tiene en cuenta los niveles de granularidad de los ontologías y ni el semántica de las clases o conceptos de la ontologías. El sistema S-Match [4] es la única excepción, considerando algo de semántica sin embargo solo considera la semántica dada por la jerarquía de la ontología. Esto provoca que estos sistemas sean poco precisos y no puedan utilizarse para la automatización de la deseada armonización. La figure 2 muestra algunas pruebas de precisión realizadas con el sistema COMA 3.0 [5][6] usando ontologías del dominio biomédico como son ICD 10 [7], SNOMED CT [8] y LOINC [9].

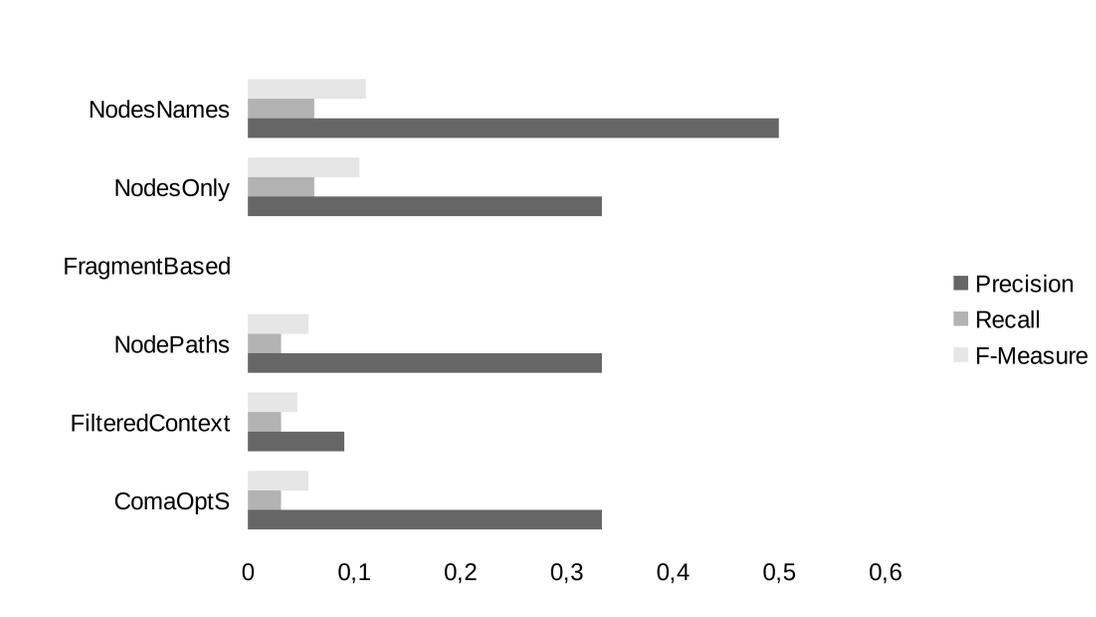


Figura 2: Pruebas de precisión realizadas sobre la herramienta COMA 3.0

5 Conclusiones

Dado que estos sistemas que ignoran la semántica no dan una solución completa al problema de la armonización de ontologías, se propone la implementación de un servicio que use ontologías de dominio para automatizar la armonización de ontologías mediante el razonamiento. Este proceso requiere que las ontologías fuente que representa el conocimiento de cada actor sean armonizadas con la ontología de dominio. Se espera que a futuro las ontologías que representan el conocimiento de cada actor se creen desde un inicio armonizadas con la ontología de dominio. Adicionalmente para conectar diferentes dominio se requiere que las ontologías de dominio se encuentren derivadas de una ontologías de nivel superior y que el conocimiento pueda ser descrito mediante un marco arquitectónico como el Modelo Generico de Componentes que restrinja el proceso y garantiza su coherencia [10], [11].

Bibliografía

- [1] B. Blobel and F. Oemig, "Ontology-driven health information systems architectures," *This Vol.*, 2009.
- [2] J. Euzenat and P. Shvaiko, *Ontology matching*. Springer, 2007.
- [3] P. Shvaiko and J. Euzenat, "Ontology matching: state of the art and future challenges," *Knowl. Data Eng. IEEE Trans.*, vol. 25, no. 1, p. 158.172, 2012.
- [4] F. Giunchiglia, P. Shvaiko, and M. Yatskevich, "S-Match: an algorithm and an implementation of semantic matching," *Semantic Web Res. Appl.*, pp. 61–75, 2004.
- [5] H. H. Do and E. Rahm, "COMA: a system for flexible combination of schema matching approaches," in *Proceedings of the 28th international conference on Very Large Data Bases*, 2002, pp. 610–621.
- [6] E. Rahm, P. Arnold, H.-H. Do, and D. Aumüller, "COMA 3.0," *COMA 3.0 | Abteilung Datenbanken Leipzig*. [Online]. Available: <http://dbs.uni-leipzig.de/Research/coma.html>. [Accessed: 05-Feb-2013].
- [7] World Health Organization, "International classification of diseases (ICD)," 2012.
- [8] IHTSDO, "SNOMED Clinical Terms Overview." Sep-2008.
- [9] L. O. I. Names, *Codes (LOINC)*. 2003.
- [10] B. Blobel and P. Pharow, "Architectural Approaches to Health Information Systems for Empowering the Subject of Care," *Med. Care N.a* 5, vol. 1, p. 355, 2008.
- [11] B. Blobel, "Knowledge Representation and Management Enabling Intelligent Interoperability–Principles and Standards," *Stud. Health Technol. Inform.*, no. 186, pp. 3–18, 2013.