

VALIDACIÓN Y VERIFICACIÓN CON ALLOY DE ONTOLOGÍAS EN OWL QUE UNIFICAN MODELOS DE DATOS HETEROGÉNEOS

Mario Luis Accattoli ¹, Ana Garis ¹, Daniel Riesco ¹

¹ Departamento de Informática - Facultad de Ciencias Físico-Matemáticas y Naturales (FCFMyN) Universidad Nacional de San Luis

accattolimario@gmail.com, agaris@unsl.edu.ar, driesco@unsl.edu.ar

RESUMEN

Ante el requerimiento de unificar sistemas de información preexistentes, las ontologías se presentan frecuentemente como la mejor opción para resolver interoperabilidad semántica e integrar modelos de datos de fuentes heterogéneas. Entre los lenguajes más populares para especificar dichas ontologías se encuentra OWL, un lenguaje basado en lógica descriptiva y soportado por un razonador que permite chequear la consistencia lógica. Sin embargo, tiene ciertas limitaciones para verificar y validar algunos aspectos relevantes. Por ejemplo, carece de características que posibiliten chequear si la ontología es coherente con el dominio del conocimiento, si es compatible para recuperar la información deseada, o si es completa y consistente de acuerdo a los requerimientos. Alloy es un lenguaje formal soportado por un analizador automático SAT, que habilita la Verificación y Validación (V&V) de modelos.

La presente línea de investigación tiene como objetivo estudiar el uso de Alloy para la V&V de ontologías en OWL desarrolladas para unificar modelos de datos heterogéneos. Como caso de estudio se plantea la V&V con Alloy, de una ontología genérica a nivel nacional especificada para unificar el modelo de datos utilizado en de las distintas jurisdicciones dentro los sistemas de información de los Poderes Judiciales de Argentina.

Palabras Claves: Ontologías, OWL, Alloy, Verificación y Validación de Modelos, Unificación Modelo de Datos Poder Judicial.

CONTEXTO

La línea de investigación se enmarca en la continuación del Proyecto de Investigación “Ingeniería de Software: Conceptos, Prácticas y Herramientas para el desarrollo de Software con Calidad” – Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis. Dicho proyecto fue acreditado con evaluación externa y contó con el financiamiento de la UNSL.

1. INTRODUCCIÓN

Dentro del ámbito de la Ingeniería de Software las ontologías son utilizadas para múltiples propósitos, tales como el modelado conceptual, la interoperabilidad semántica, y también para facilitar el intercambio e integración de los datos [1]. Los lenguajes ontológicos permiten definir la semántica de los conceptos, habilitando no solo el almacenamiento de la información sino también la búsqueda y recuperación desde modelos de datos que provienen de fuentes heterogéneas. Sin embargo, los beneficios estarían limitados si la definición de la ontología se llevara a cabo de manera informal, generando un modelo ambiguo e impreciso.

Los métodos formales, proveen técnicas, métodos y herramientas para asegurar la correctitud de los modelos. Típicamente, incluyen un lenguaje de especificación soportado por herramientas de verificación y validación utilizados para eliminar inconsistencias y chequear propiedades deseadas. OWL [2] es un lenguaje formal para ontologías, aplicado principalmente

para representar la semántica Web. Si bien, el lenguaje está basado en lógica descriptiva y es soportado por un razonador que permite chequear la consistencia lógica de la ontología, no es posible verificar y validar algunos aspectos relevantes, tales como si es coherente con el dominio del conocimiento, o compatible para recuperar la información deseada.

Alloy [3] es un lenguaje formal soportado por un analizador automático SAT, que habilita la Verificación y Validación (V&V) de modelos, y permite realizar especificaciones no ambiguas, así como identificar y demostrar propiedades del sistema. El potencial de Alloy podría ser aprovechado para la V&V de ontologías OWL.

La presente línea de investigación se enfoca en el estudio de Alloy para V&V de ontologías en OWL desarrolladas para unificar modelos de datos heterogéneos. En particular, la propuesta procura resolver un problema que se plantea en los sistemas de los Poderes Judiciales de Argentina relacionado a la dificultad de unificar información y homogeneizar conceptos, de tal manera que también se consideren las particularidades de los distintos distritos provinciales.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Los ejes de la investigación se basan en los siguientes puntos:

- La evaluación de los mecanismos para llevar a cabo la unificación de modelos de datos de fuentes heterogéneas mediante OWL y la identificación de problemas de inconsistencias que deberían ser luego verificados y validados para chequear la correctitud del modelo.
- El análisis de antecedentes en la adopción de Alloy para la V&V de ontologías OWL y enfoques de transformación entre OWL y Alloy.

- La especificación en OWL de la ontología que unifica modelos de datos heterogéneos en el dominio de la Justicia. En este sentido, es necesario estudiar la existencia de ontologías ya definidas en dicho dominio y evaluar los elementos que deben ser verificados y validados.

3. RESULTADOS OBTENIDOS/ESPERADOS

Con respecto a los mecanismos para realizar la unificación de modelos de datos de fuentes heterogéneas, se han analizado enfoques generales presentados en la literatura, esto es, el de ontología global, múltiple e híbrida [1,4,5,6]. El estudio preliminar indica que se debería partir de un proceso de ingeniería inversa de los modelos actuales [7], para luego armar una ontología unificada del modelo de datos. Es necesario evaluar aún si el proceso de ingeniería inversa puede ser ejecutado en forma automático o semiautomático [8,9].

Resta analizar cómo se lleva a cabo la unificación con OWL en el caso de estudio planteado en el trabajo: Modelos de Datos de Sistemas de Gestión Judicial en Argentina. Se espera poder identificar problemas de inconsistencia mientras se desarrolla esta actividad, que permitan definir reglas de V&V generales a chequear con Alloy.

La adopción de Alloy para la V&V de estas ontologías OWL requiere investigar propuestas similares y enfoques de transformación entre OWL y Alloy, tales como el presentado en [10].

Para chequear la propuesta se utilizará un caso de estudio real. Es decir se planteará la definición de un modelo de datos unificado de los Sistemas de Gestión Judicial en Argentina a través de ontologías validado y verificado formalmente. La unificación no solo mejoraría la comparación entre provincias, sino que se convertiría en un base de datos de consultas de antecedentes, facilitando la recopilación de casos similares para una mejor resolución de los

actuales, así como el registro de la trazabilidad de comportamientos de personas relacionadas con causas judiciales.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo de la línea de investigación está integrado por investigadores de la UNSL, asesores externos y alumnos de grado y posgrado. En el grupo se llevan a cabo numerosas tesis de especialización, maestría y doctorado, así como trabajos finales de carreras de grado. En particular, el presente trabajo se desarrolla en el marco de la tesis de uno de los autores para optar al grado de Doctor en Ingeniería Informática en la UNSL.

Los avances logrados en esta línea de investigación, en el marco del Proyecto de Investigación mencionado en la Sección Contexto, establecen una base para el inicio de otras tesis de posgrado, tanto en el doctorado como en la Maestría en Ingeniería de Software, la Maestría en Calidad del Software, la Especialización en Ingeniería de Software de la UNSL. De igual forma, el presente trabajo, brinda un aporte para desarrollar trabajos finales en las carreras de grado de la UNSL, específicamente en la Licenciatura en Ciencias de la Computación, Ingeniería en Informática e Ingeniería en Computación.

5. BIBLIOGRAFÍA

[1] I.F. Cruz, H. Xiao, "The role of ontologies in data integration". *Engineering Intelligent Systems for Electrical Engineering and Communications*, 13 (4), p. 245. 2005.

[2] W3C, OWL Web Ontology Language Reference. Disponible en <https://www.w3.org/TR/owl-ref/> (última visita marzo de 2020)

[3] D. Jackson. *Software Abstractions: Logic, Language, and Analysis*. MIT Press, revised edition, 2012.

[4] D. Skoutas and A. Simitsis, "Ontology-based Conceptual Design of ETL Processes for both Structured and Semi-structured Data", *International Journal on Semantic Web and Information Systems, Special Issue on Semantic Web and Data Warehousing*, 2007.

[5] H. Wache, T. Vogeles, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann and S. Hubner, "Ontology-Based Integration of Information. A Survey of Existing Approaches," In Proc. of the IJCAI workshop on Ontologies and Information Sharing, 2001.

[6] A. Buccella, A. Cechich, N. Brisaboa, "Ontology-Based Data Integration Methods: A Framework for Comparison", *Revista Colombiana de Computación*, 2005.

[7] Justas Trinkunas, Olegas Vasilecas: *Building Ontologies from Relational Databases Using Reverse Engineering Methods (2007)(CompSysTech'07)*

[8] H. Afzal, M. Waqas, T. Naz, "OWLMap: Fully Automatic Mapping of Ontology into Relational Database Schema", *(IJACSA) International Journal of Advanced Computer Science and Applications*, Vol. 7, No. 11, 2016.

[9] L. Zemmouchi-Ghomari, "Cohabitation of Relational Databases and Domain Ontologies in the Semantic Web Context", *Journal of Systems Integration*, 2018.

[10] Y. Song, R. Chen, Y. Liu, "A Non-Standard Approach for the OWL Ontologies Checking and Reasoning", *Journal of Computers*, Vol.7(10): 2454-2461, 2012.