

Agentes Inteligentes y Web Semántica

Christian Gimenez¹ Germán Braun^{1,2} Angela Oyarzun¹
Gaston Michelan¹ Emiliano Rios Gavagnin¹ Giuliano Marinelli¹
Sandra Roger¹ Mario Moya¹ Claudio Vaucheret¹ Laura Cecchi¹

email: {christian.gimenez, german.braun, gaston.michelan, giuliano.marinelli,
roger, mario.moya, cv, lcecchi}@fi.uncoma.edu.ar
{angela.oyarzun, emiliano.rios}@est.fi.uncoma.edu.ar

¹*Grupo de Investigación en Lenguajes e Inteligencia Artificial*
Departamento de Teoría de la Computación - Facultad de Informática
UNIVERSIDAD NACIONAL DEL COMAHUE
²*Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)*

Resumen

En este trabajo se presentan las líneas de investigación desarrolladas en el Proyecto de Investigación Agentes Inteligentes y Web Semántica, de la Facultad de Informática, UNCo. En el proyecto, se estudian técnicas de representación de conocimiento y razonamiento, metodologías de modelado conceptual y mecanismos para la interoperabilidad de aplicaciones, tanto a nivel de procesos como de datos, para entre otros, dar soporte a comunidades de desarrollo de ontologías.

En este sentido, se ha desarrollado una herramienta Web denominada *crowd* que da soporte a la ingeniería ontológica asistida con razonamiento automático y que está siendo extendida con nuevas funcionalidades estudiadas en las diferentes líneas de investigación.

Palabras Clave: Ingeniería de Software basada en Conocimiento, Lógicas Descriptivas, Ontologías. Datos Enlazados. OVM

Contexto

Este trabajo está parcialmente financiado por la UNCo, en el marco del proyecto de investigación Agentes Inteligentes y Web Semántica

(04/F014) y de una Beca de Investigación de en la categoría Graduado de Iniciación y de una Beca de Iniciación en la Investigación para Estudiantes. Asimismo, lo financia parcialmente el Consejo Interuniversitario Nacional (CIN) con dos Becas de Estímulo a las Vocaciones Científicas 2018 y 2019 y por el Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), en el contexto de una beca interna doctoral. El proyecto de investigación tiene una duración de cuatro años y ha comenzado en 2017. El proyecto se desarrolla en forma colaborativa con docentes-investigadores de la UNS.

1. Introducción

Actualmente, es esperable que las aplicaciones accedan a grandes cantidades de datos que se modifican en tiempo real, de manera constante y que se presentan en diferentes formatos, especialmente disponible a través de la Web. Muchas de estas aplicaciones ocurren bajo restricciones de tiempo críticas y en intensa interacción con el usuario. En este contexto, la representación conceptual de los dominios para la generación y extracción de información y conocimiento, es central en la toma de decisiones. Bajo este aspecto, el objetivo general del proyecto de investigación es el estudio y desarrollo

de técnicas de Inteligencia Artificial para dotar de inteligencia, conocimiento e interoperabilidad semántica a agentes inmersos en ambientes complejos.

Asimismo, la Web es un ambiente ideal para estos agentes que, junto con la generación y representación de conocimiento, las ontologías y el razonamiento automático, son fundamentales en la evolución de dicha Web, denominada Web Semántica. La meta de la Web Semántica es crear una Web de conocimiento en la cual la semántica del contenido es explicitada, permitiendo novedosas aplicaciones que combinan datos de sitios heterogéneos para, entre otros objetivos, mejorar la experiencia de los usuarios de acuerdo a sus necesidades. Esta red de funcionalidades basadas en conocimiento hará a dicho conocimiento procesable por una computadora, para soportar conductas inteligentes por parte de dichas máquinas.

Así en el proyecto, se estudian técnicas de representación de conocimiento y razonamiento, metodologías de modelado conceptual y mecanismos para la interoperabilidad de aplicaciones, tanto a nivel de procesos como de datos, para entre otros, dar soporte a comunidades de desarrollo de ontologías.

En esta dirección, se ha desarrollado *crowd*[1, 2], una herramienta Web, creada en respuesta a la complejidad inherente al modelado conceptual y ontológico.

En este trabajo se presentan las líneas de investigación desarrolladas en el Proyecto de Investigación Agentes Inteligentes y Web Semántica, de la Facultad de Informática. Se describen los resultados parciales alcanzados, avalados con publicaciones y los trabajos en desarrollo.

Agentes Inteligentes y Web Semántica tiene asociado el Proyecto de Extensión *Agentes Robots: Divulgando Computación en la Escuela Media*, cuyo objetivo es divulgar temas de la disciplina y enseñar a programar utilizando como eje motivador robots físicos, a través de la *Copa de Fútbol FAI: la definición es por penales*[3].

2. Línea de Investigación y Desarrollo

Motivados por el objetivo del proyecto, se delinearon varias líneas de investigación que confluyen explorando, entre otros, temas afines a la representación del conocimiento, las Lógicas Descriptivas (DLs), las Ontologías, la Ingeniería de Software basada en Conocimiento, la Ingeniería de Conocimiento y los Datos Enlazados y Abiertos. A continuación se las describen brevemente.

(L1) Formalización de Sistemas de Visualización Ontológicos con soporte de Razonamiento: Durante varias décadas, un esfuerzo considerable ha sido invertido en ampliar el horizonte teórico de las DLs y los lenguajes de ontologías. Sin embargo, se ha relegado a un segundo plano el desarrollo de ambientes de ingeniería ontológica para simplificar la formalización y la validación de las propiedades de los modelos conceptuales sobre dominios de aplicación.

En esta línea se trabaja en la integración de lenguajes de modelado conceptual visuales con razonamiento basado en lógica y su utilidad en cuanto a tareas de ingeniería ontológica automáticas, y la validación de modelos. En este contexto, se formalizó un sistema teórico para la manipulación de modelos visuales, dando lugar a la definición de ontologías gráficas y se implementó *crowd* como instancia de esta formalización.

(L2) Lenguaje Gráfico para Consultar Ontologías: En esta línea de investigación se propone desarrollar un lenguaje gráfico que represente las consultas SPARQL-DL a realizarse sobre una ontología. Dicho lenguaje se pretende que posea primitivas que sean similares a los lenguajes de modelado ampliamente usados en la ingeniería de Software para minimizar la curva de aprendizaje. De esta manera, se espera poder buscar patrones que se presenten en la ontología mezclando consultas de nivel conceptual (TBox) como del nivel de datos (ABox).

Inicialmente, se describirán primitivas visuales para consultar las estructuras del TBox. Las

consultas visuales se traducirán a su versión textual, las cuales en conjunto con una ontología serán ejecutadas sobre un motor SPARQL-DL para obtener las respuestas. Las respuestas obtenidas, son mostradas al usuario en forma textual, pudiendo ser interpretadas gráficamente reemplazando las variables por los nombres de clases encontrados.

Para implementar este lenguaje, se continuará con el desarrollo de la herramienta Web denominada *crowd*, la que asiste al modelador en el diseño y la visualización de ontologías, actualmente a través del lenguaje de modelado conceptual UML y que soporta razonamiento automático para validar los modelos.

(L3) Alineamiento de Ontologías en un Ambiente Gráfico Web de Ingeniería Ontológica: En la actualidad, los datos presentes en sistemas informáticos son muy abundante y están almacenados en distintos formatos. Las ontologías son indispensables en este escenario, ya que permiten una descripción formal de los datos y de su dominio. Por esta razón, es importante trabajar sobre la heterogeneidad presente en los modelos de información mediante el alineamiento de ontologías, el cual tiene como objetivo encontrar correspondencias entre entidades de modelos diferentes y, así, reducir la brecha semántica entre representaciones de un mismo dominio.

Esta línea de investigación propone trabajar sobre las técnicas de alineamiento de ontologías para integrarlas en la herramienta *crowd*. Se pretende desarrollar un diseño que permita la integración de ontologías diferentes en un entorno gráfico, mediante la identificación de correspondencias sintácticas y semánticas entre sus entidades.

(L4) Verbalización multilingüe del metamodelo UML: Esta línea de investigación estudia la verbalización multilingüe del metamodelo UML, orientado hacia un subconjunto de primitivas de los diagramas de clases. Se utiliza una representación intermedia de los diagramas de clases en sentencias concretas de lógica de primer orden (FOL) para favorecer las verbalizaciones en diversos idiomas.

La implementación de esta formalización se

realiza sobre *crowd* mediante una extensión que permita la incorporación de esta nueva funcionalidad. Como principal novedad con respecto a las aplicaciones vigentes, se puede destacar que los usuarios podrán visualizar y editar de forma *on-line* diagramas de clases UML y posteriormente, realizar su verbalización a lenguaje natural.

Un aspecto relevante, es que la traducción a lenguaje natural está destinada a ser multilingüe, a diferencia de la mayoría de las herramientas existentes, donde el lenguaje natural de salida es generalmente el inglés.

(L5) Algoritmos de Layout para una Herramienta Web de Modelado Ontológico: En la actualidad disponemos de herramientas que asisten a los modeladores para la realización de tareas de modelado conceptual y ontológico, las cuales brindan un soporte gráfico de manera que se simplifique la visualización del contenido y facilite dicha tarea.

Debido a la complejidad inherente de modelos y ontologías de dominio en cuanto al tamaño de los diagramas y a los requerimientos de edición de los modeladores, algoritmos efectivos y eficientes de layout deben ser desarrollados para garantizar la legibilidad de los modelos. El objetivo de esta línea de investigación es investigar y diseñar algoritmos de layout. Para esto se profundizará en técnicas de visualización basadas en la teoría de *Crossing Number* y sus variantes, aplicado a los lenguajes clásicos de modelado conceptual como UML, EER y ORM.

Utilizaremos la herramienta *crowd* para la implementación de nuevos prototipos y simultáneamente este trabajo permitirá la ampliación de la herramienta.

(L6) Formalización de Modelos de Variabilidad Ortogonal: En esta línea de investigación se trabaja en la formalización de Modelos de Variabilidad Ortogonal (OVM) en lenguajes formales como FOL y DL. Asimismo, tiene como objetivo el desarrollo de una herramienta Web, basada en un modelo cliente-servidor, que permita integrar el soporte gráfico para el diseño de OVM y el razonamiento automático

para validar dichos modelos. El objetivo de esta herramienta es solventar las limitaciones presentadas en las propuestas existentes basadas en razonadores SAT a través de la utilización de DLs. De igual manera, busca evitar la ocurrencia de errores en el diseño de los modelos al proveer una interfaz de usuario gráfica con un conjunto de primitivas del lenguaje OVM y las funcionalidades necesarias para la construcción y edición de los diagramas.

(L7) Datos Enlazados en el dominio de publicaciones científico: Debido a la promulgación de las leyes en Argentina que exigen la publicación de datos abiertos, (Ley Nacional N° 26.899), existen sitios nacionales en los cuales se expone información en formatos de uso libre tales como JSON y .CSV. Teniendo en cuenta la importancia que comienza a tener el paradigma de Web Semántica en conjunto al concepto de Datos Enlazados, este trabajo propone el diseño de una herramienta que relacione mediante enlaces los datos ya existentes en esos formatos y permita enriquecer la información mediante la colaboración de diversas instituciones que unen sus bases de datos de información científica, manteniendo el estándar que se propone de enlaces y relaciones bajo el concepto de Web Semántica, promulgando la unicidad, consistencia y el libre acceso a la información.

3. Resultados Obtenidos y Trabajo Futuro

Actualmente, *crowd* permite crear modelos basados en un subconjunto de primitivas de UML. Soporta el razonamiento automático para determinar inconsistencias del modelo. Asimismo, es posible obtener la representación textual del modelo formalizado en OWL2 para ser utilizado con otras herramientas. Un prototipo está disponible en crowd.fi.uncoma.edu.ar.

Respecto de (L1), se está trabajando en la expansión de los conceptos teóricos para lograr mayor interacción gráfica y lógica y, también, en el desarrollo de nuevas funcionalidades de

crowd: importación de ontologías OWL 2 y soporte a múltiples lenguajes gráficos de modelado conceptual. Asimismo, se ha comenzado con el desarrollo de la línea (L3) para abarcar más tareas asociadas a la ingeniería ontológica. En cuanto al layout automático (L5), se desarrolló un algoritmo genético [4] que permite disponer visualmente al grafo que representa al modelo con un crossing number adecuado. Actualmente, se está estudiando la adaptación de este algoritmo para que se comporte apropiadamente de acuerdo a los diferentes lenguajes de modelado conceptual soportados por *crowd*.

Definido el lenguaje visual propuesto en la línea (L2)[5], se está implementando una extensión de la arquitectura de *crowd*, que consiste en un diseño de interfaz para el modelado de consultas, utilizando dicho lenguaje gráfico. En el servidor, se llevará a cabo la traducción y su ejecución sobre una ontología ofrecida por el usuario, mostrando los resultados en el cliente.

En cuanto a los resultados de la línea (L4), se diseñaron los esquemas de verbalización en lenguaje español, basados en *templates*[6]. Además, se integró al módulo de verbalización una herramienta que permite obtener la morfología gramatical (género y número) de cada uno de los nombres de clases y atributos. Asimismo, se estableció la conexión con un servicio web que realiza la conjugación verbal de los nombres de las asociaciones. Actualmente se está finalizando la fase de implementación, extendiendo a *crowd*. A futuro, se prevee la extensión a otros idiomas, mediante la creación de nuevos *templates* de verbalización, y la correspondiente validación.

En esta línea de investigación (L6) se formalizó OVM utilizando lenguajes formales como FOL y DL[7]. Actualmente, se está desarrollando la herramienta Web, que permita integrar el soporte gráfico para el diseño de OVM y el razonamiento automático para validar dichos modelos[8].

Finalmente, en el marco de la línea (L7) se ha trabajado en procesar los datos existentes y convertirlos en un formato estándar (por convención RDF)[9]. En este sentido, es necesario además identificarlos unívocamente utilizando

URIs para mantener consistencia, reconciliarlos, enlazarlos y enriquecerlos con información relacionada mediante lenguajes de conocimiento como OWL, trabajo que se está desarrollando. Se plantea como trabajo futuro desarrollar una interfaz amigable para ser presentados y utilizados por los usuarios.

4. Formación de Recursos Humanos

Uno de los integrantes está finalizando el Doctorado en Ciencias de la Computación en la Universidad Nacional del Sur (beca interna doctoral CONICET). Otro de los miembros del proyecto, goza de una Becas de Investigación en la categorías de Graduado de Iniciación, UNCo y en el transcurso del 2019 comenzará a cursar un posgrado acreditado. Dos de los integrantes están cursando actualmente posgrados: Magíster en Ciencia, tecnología e Innovación de la UNRN y el Magíster en Ciencias de la Computación de la UNS.

Finalmente, los integrantes alumnos se encuentran terminando la carrera Licenciatura en Ciencias de la Computación. Tres de ellos están actualmente desarrollando su tesis. Uno de estos estudiantes goza de la Beca de Iniciación en la Investigación para Estudiantes de la UNCo y dos de ellos Becas de Estímulo a las Vocaciones Científicas 2018 y 2019 (CIN).

Referencias

- [1] Christian Gimenez, Germán Braun, Laura Cecchi, and Laura Fillotrani. crowd: A Tool for Conceptual Modelling assisted by Automated Reasoning - Preliminary Report. In *the 2nd Simposio Argentino de Ontologías y sus Aplicaciones SAOA '16 JAIIO '16*, 2016.
- [2] Germán Braun, Christian Gimenez, Pablo Fillotrani, and Laura Cecchi. Towards conceptual modelling interoperability in a web tool for ontology engineering. In *the 3rd Simposio Argentino de Ontologías y sus Aplicaciones SAOA '17 JAIIO '17*, 2017.
- [3] Jorge Rodriguez, Guillermo Grosso, Rafael Zurita, and Cecchi Laura. Intervención de la Facultad de Informática en la enseñanza de Ciencias de la Computación en la Escuela Media basada en Robótica Educativa. In *XI Congreso de Tecnología en Educación y Educación en Tecnología*, 2016.
- [4] Giuliano Marinelli, Germán Braun, Laura Cecchi, and Pablo Rubén Fillotrani. Algoritmos de layout automático para una herramienta multi-vistas de modelado ontológico. In *XX Workshop de Investigadores en Ciencias de la Computación (WICC 2018, Universidad Nacional del Nordeste)*, 2018.
- [5] Christian Gimenez, Germán Braun, Laura Cecchi, and Pablo Rubén Fillotrani. Towards a visual sparql-dl query builder. In *XXIV Congreso Argentino de Ciencias de la Computación (CACIC 2018)*, Facultad de Ciencias Exactas - Universidad Nacional del Centro de la Provincia de Buenos Aires, 2018.
- [6] Matías Garrido, Germán Braun, and Sandra Roger. Agentes inteligentes y web semántica: hacia la verbalización de un subconjunto de uml en una herramienta gráfica web. In *XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017, ITBA, Buenos Aires)*, 2017.
- [7] Germán Alejandro Braun, Matias Pol'la, Laura Andrea Cecchi, Agustina Buccella, Pablo R. Fillotrani, and Alejandra Cechich. A DL semantics for reasoning over ovm-based variability models. In *Proceedings of the 30th International Workshop on Description Logics, Montpellier, France, July 18-21, 2017.*, 2017.
- [8] Angela Oyarzun, Germán Braun, Laura Cecchi, and Pablo Rubén Fillotrani. A graphical web tool with dl-based reasoning support over orthogonal variability models. In *XXIV Congreso Argentino de Ciencias de la Computación (CACIC 2018)*, Facultad de Ciencias Exactas - Universidad Nacional del Centro de la Provincia de Buenos Aires, 2018.
- [9] Gaston Michelin, Germán Braun, Laura Cecchi, and Pablo Rubén Fillotrani. Hacia la integración semántica de repositorios digitales abiertos en argentina. In *5to Congreso Nacional de Ingeniería Informática- Sistemas de Información (CONAISI 2017)*, UTN - Facultad Regional Santa Fe, 2017.