

Linked Data y Ontologías en una Herramienta Gráfica Web

Gaston Michelin¹ Germán Braun^{1;2;3} Laura Cecchi¹
 Pablo Fillottrani^{2;4}

email: {gaston.michelan,german.braun,lcecchig}@fi.uncoma.edu.ar, prf@cs.uns.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*Laboratorio de I&D en Ingeniería de Software y Sistemas de Información*
 Departamento de Ciencias e Ingeniería de la Computación
 UNIVERSIDAD NACIONAL DEL SUR

³*Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)*

⁴Comisión de Investigaciones Científicas de la provincia de Buenos Aires (CIC)

Resumen

Esta línea de investigación se desarrolla en forma colaborativa entre docentes-investigadores de la Universidad Nacional del Comahue y de la Universidad Nacional del Sur, en el marco de proyectos de investigación financiados por las universidades antes mencionadas.

El objetivo general del trabajo de investigación es permitir la interacción entre fuentes de datos enlazados disponibles en la Web y la herramienta cliente-servidor para el modelado conceptual gráfico con soporte de razonamiento: *crowd*. De este modo, se espera poder navegar cualquier ontología asociada a los datos y observar sus relaciones de una manera gráfica, esto último, con el fin de facilitar su interpretación al usuario convencional.

Palabras Clave: Linked Open Data, Web Semántica, Ingeniería de Software.

Contexto

Este trabajo está parcialmente financiado por la Universidad Nacional del Integración de fuentes de datos, su representación, la extracción de información a partir de ellos y para establecer sus relaciones y correctitud. Para los primeros casos tenemos, como área de estudio, lo que se llama Linked

Comahue, en el marco del proyecto de investigación Agentes Inteligentes y Web Semántica (04/F014) y a través de una beca de Iniciación a la Investigación para alumnos; por la Universidad Nacional del Sur a través del proyecto de investigación Integración de Información y Servicios en la Web (24/N027), y por el Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), en el contexto de una beca interna doctoral. Los proyectos de investigación tienen una duración de cuatro años y la beca doctoral una duración de 5 años, finalizando esta última en abril de 2019. La beca de Iniciación a la Investigación tiene una duración de un año.

1. Introducción

En la actualidad, el rol de la Tecnología de la Información (TI) es indispensable en el éxito de los procesos intensivos de datos tales como la inteligencia de negocios y la toma de decisiones [1]. Sin embargo, existe una importante brecha entre las personas que interpretan y usan los datos, y los sistemas, en los cuales dichos datos son almacenados y procesados. Esto redundando en la necesidad de nuevas tecnologías y herramientas para la Data [2, 3, 4], movimiento propuesto por Tim Berners Lee, cuyo principal objetivo es publicar y conectar datos estructurados en la Web [5, 6], a partir de un conjunto de prácticas para tal fin. Para el segundo caso podemos encontrar

herramientas de modelado conceptual, las cuales nos permiten, a través de relaciones entre conceptos u ontologías, y un razonador[7, 8, 9], evaluar la correctitud de esos datos en cuanto a sus relaciones y a la vez, proponer nuevas interacciones implícitas entre ellos[10].

En este sentido y como un primer enfoque para la solución de estos problemas, hemos desarrollado en nuestros grupos de investigación, dos herramientas: *crowd* [11, 12, 13] y LinkedGILIA [14, 15]. *Crowd* es una herramienta cliente-servidor para el modelado conceptual gráfico y diseño de ontologías con soporte de razonamiento para asistir a usuarios en esos procesos. Además, permite ejecutar chequeos de satisfacibilidad e inferencia de restricciones implícitas, cuyos resultados se muestran utilizando el mismo lenguaje gráfico. Por otro lado, LinkedGILIA es una herramienta que nos permite convertir bases de datos relacionales a RDF [16], consultarlas a través de SPARQL [17] y enlazar esos datos con otras fuentes RDF existentes en la Web, aplicando el concepto de LinkedData y haciendo que estos sean de público acceso. Esta herramienta está orientada particularmente al dominio académico y de información científica.

En el ámbito de este trabajo proponemos, como primera instancia, agregar datos enlazados a *crowd* para permitir navegar cualquier ontología asociada a ellos y poder observar sus relaciones gráficamente, esto también nos habilitará a realizar el proceso inverso, es decir, a partir de una ontología obtener sus datos relacionados. A su vez, se planea colaborar con el modelador para mejorar las ontologías diseñadas, aplicando herramientas como DL-learner [18, 19], continuando con el desarrollo del ciclo de vida de aplicaciones Linked Data [20] comenzado en [15].

Finalmente, se propone extender *crowd* para visualizar los datos en RDF y en algún formato más amigable utilizando alguna biblioteca PHP para consultar las tripletas producidas.

La estructura del presente trabajo es la siguiente. En la sección 2, presentamos los objetivos de los proyectos de investigación en los que se enmarca este trabajo y describimos la línea de investigación, el problema que se

estudia y los objetivos. En la sección 3, indicamos algunos resultados obtenidos y trabajos futuros. Finalmente, comentamos aspectos referentes a la formación de recursos humanos en esta temática.

2. Línea de Investigación y Desarrollo

Dentro de los objetivos del nuevo proyecto de investigación Agentes inteligentes y Web Semántica, UNCo, se planteó el de generar conocimiento especializado en el área de agentes inteligentes y en lo referente a la representación y el uso del conocimiento en sistemas computacionales basados en la web, es decir lo que se ha llamado la Web Semántica. Para ello, es necesario profundizar el estudio de 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. Se pretende aplicar estos conceptos como soporte para comunidades de desarrollo de ontologías, entre otros.

El proyecto de investigación Integración de Información y Servicios en la Web, UNS, se propone la integración y búsqueda, en forma semántica, de la información y de servicios en la Web utilizando lenguajes y tecnologías que están innovando en la actualidad.

Las investigaciones de ambos proyectos pueden relacionarse y ser aplicadas en el estudio de tecnologías y formalismos para solventar las necesidades de compartir, intercambiar y asegurar la correctitud de la información, que se encuentra disponible en sistemas existentes, de una forma precisa y consistente.

Particularmente, en esta línea de investigación se experimentará sobre la integración de tecnologías y lenguajes ya existentes para la generación de datos abiertos, enlazados y correctos y la publicación de los mismos en la Web. Del mismo modo, se espera, identificar falencias, proponer mejoras y evaluar el impacto en los usuarios, de esta nueva infraestructura.

Como primer paso se trabajará en el análisis de la arquitectura y del funcionamiento in-terno

de la herramienta *crowd*, para determinar las formas de inserción de datos y el tipo de los mismos que admite. Luego, se adaptará la arquitectura de la herramienta, de modo de crear una conexión entre diferentes fuentes de datos RDF y *crowd*. Esto mejorará la asistencia al modelador ya que permitirá visualizar gráficamente, además de las relaciones entre los datos, las ontologías subyacentes a dichas fuentes enlazadas.

En una segunda instancia, se pretende extraer nuevo conocimiento a partir de los datos, usando herramientas basadas en machine learning y generando axiomas y conceptos que se correspondan con los datos subyacentes. De esta manera, soportaremos el proceso de integración de los datos [21] con las ontologías de alto nivel.

Nuestra intención con esto, es establecer relaciones entre datos pertenecientes a bases de datos linkeadas en la Web, como así también exponer toda esta información a través de front-ends visuales para habilitar búsquedas semánticas a usuarios convencionales.

3. Resultados Obtenidos y Trabajo Futuro

De anteriores trabajos surgió LinkedGILIA [15], que es una aplicación Web específica para el dominio académico, que enlaza información científica desde diferentes fuentes como, por ejemplo, DBLP, entre otras. Dicha aplicación permite visualizar la información actualizada asociada a los datos disponibles y realizar consultas en el lenguaje SPARQL [17]. Como caso de estudio se utilizó, para la implementación del prototipo, al grupo de investigación GILIA del Departamento de Teoría de la Computación, de la Facultad de Informática, de la Universidad Nacional del Comahue.

Actualmente, nos encontramos trabajando en el análisis de la arquitectura y funcionalidades de *crowd* [12, 13] y en la evaluación de las diferentes formas de interacción que esta posee con elementos externos. A partir de los resultados de estos estudios, se extenderá la arquitectura *crowd* y se desarrollará una interfaz, que nos permita navegar por los datos

y la ontología modelada, accediendo a través de sus enlaces y visualizando información relacionada con otras ontologías existentes en la Web.

Se espera implementar un prototipo de la arquitectura extendida y evaluarla sobre distintos casos de estudio.

4. Formación de Recursos Humanos

Uno de los autores de este trabajo está inscripto en el Doctorado en Ciencias de la Computación en la Universidad Nacional del Sur (beca interna doctoral CONICET).

Otro de los autores, alumno avanzado de la carrera Licenciatura en Ciencias de la Computación, ha obtenido una Beca de Iniciación en la Investigación para Alumnos Universitarios de la Universidad Nacional del Comahue. Dicho becario realizará su tesis de grado en la temática de la línea de investigación presentada en el marco del GILIA.

5. Referencias

- [1] Bou-Wen Lin. *Information technology capability and value creation: Evidence from the us banking industry*. Technology in Society, 2007.
- [2] Tim Berners-Lee. Linked Data, 2006. <https://www.w3.org/DesignIssues/Linkedata.html>, accedida en septiembre de 2016.
- [3] Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked Data - The Story So Far. *Int. J. Semantic Web Inf. Syst.*, 2009.
- [4] Tom Heath and Christian Bizer. *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool, 1st edition, 2011.
- [5] T. Berners-Lee, J. Hendler, and O. Lassila. *The semantic web*. Scientific American, May 2001.

- [6] Liyang Yu. *A Developer's Guide to the Semantic Web*, Second Edition. Springer, 2014.
- [7] V. Haarslev and R. Moller. Racer system description. In R. Gore, A. Leitsch, and T. Nipkow, editors, *International Joint Conference on Automated Reasoning*, IJ-CAR'2001, June 18-23, Siena, Italy, pages 701–705. Springer-Verlag, 2001.
- [8] Evren Sirin, Bijan Parsia, Bernardo Cuenca Grau, Aditya Kalyanpur, and Yarden Katz. Pellet: A practical owl-dl reasoner. *Web Semant.*, 5(2):51–53, June 2007.
- [9] Dmitry Tsarkov and Ian Horrocks. Fact++ description logic reasoner: System description. In *In Proc. of the Int. Joint Conf. on Automated Reasoning (IJCAR 2006)*, pages 292–297. Springer, 2006.
- [10] Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, New York, NY, USA, 2003.
- [11] Christian Gimenez, Germán Braun, Laura Cecchi, and Pablo Fillottrani. Una arquitectura cliente-servidor para modelado conceptual asistido por razonamiento automático. In *Proceedings of XVIII Workshop de Investigadores en Ciencias de la Computación*, Argentina, 2016. Universidad Nacional de Entre Ríos.
- [12] Christian Gimenez, Germán Braun, Laura Cecchi, and Pablo Fillottrani. crowd: A tool for conceptual modelling assisted by automated reasoning - preliminary report. In *Proceedings of Simposio Argentino de Ontologías y sus Aplicaciones*. 45 JAIIO, Buenos Aires, 2016. Universidad de Tres de Febrero.
- [13] Germán Braun, Christian Gimenez, Laura Cecchi, and Pablo Fillottrani. Towards a visualisation process for ontology-based conceptual modelling. In *Proceedings of VIII Brazilian Seminar Ontology*. ONTOBRAS, Curitiba-PR, Brasil.
- [14] Gaston Michelan, Germán Braun, Laura Cecchi, and Pablo Fillottrani. Linked open data para la integración de información científica. In *Proceedings of XVIII Workshop de Investigadores en Ciencias de la Computación*, Argentina, 2016. Universidad Nacional de Entre Ríos.
- [15] Gaston Michelan, Germán Braun, Laura Cecchi, and Pablo Fillottrani. Integration of scientific information through linked data - preliminary report. In *Proceedings of Simposio Argentino de Ontologías y sus Aplicaciones*. 45 JAIIO, Buenos Aires, 2016. Universidad de Tres de Febrero.
- [16] Marcelo Arenas, Claudio Gutierrez, and Jorge Pérez. Foundations of rdf databases. In *Reasoning Web*, pages 158–204, 2009.
- [17] World Wide Web Consortium. SPARQL 1.1 Query Language. Available at <http://www.w3.org/TR/sparql11-query/>, last accessed May 2016
- [18] DL-Learner Homepage. <http://dl-learner.org/>.
- [19] Lorenz Buhmann, Jens Lehmann, and Patrick Westphal. DL-learner—a framework for inductive learning on the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 39:15–24, 2016.

- [20] Anja Jentzsch, Max Schmachtenberg, Christian Bizer and Richard Cyganiak. Linking Open Data cloud diagram, 2014. <http://lod-cloud.net/>, accedida en septiembre de 2016.

- [21] AnHai Doan, Alon Halevy, and Zachary Ives. *Principles of Data Integration*. Elsevier, 2012.