

Interoperabilidad Semántica de la Producción Científica En Repositorios Académicos

Patricia Zachman¹, Mariela Gola^{1,2}, Elena Durán^{1,2}, Margarita Alvarez^{1,2},
Jonatan Aguilera¹

(1) Departamento de Ciencias Básicas y Aplicadas.- Universidad Nacional del Chaco Austral.-

Cdte. Fernández N° 755.- Pcia. Roque Sáenz Peña.- Chaco – TE 0364-420137 Int 155

(2) Departamento de Informática, Facultad de Ciencias Exactas y Tecnologías, Universidad Nacional de Santiago del Estero.

Avda. Belgrano N° 1912.- Santiago del Estero – TE 0385-4509560 Int 1838

ppz@uncaus.edu.ar, marielaag@hotmail.com, eduran@unse.edu.ar, alvarez@unse.edu.ar,
jonemmanu@gmail.com.ar

Resumen

Las necesidades actuales de interoperabilidad, usabilidad, interpretabilidad y uniformidad de vocabularios han llevado a repensar los modelos clásicos de organización del conocimiento en repositorios académicos. En este contexto es donde comienza a adquirir funcionalidad, las búsquedas semánticas. En la Universidad Nacional del Chaco Austral (UNCAus) existen numerosas producciones académicas resultado de los proyectos de investigación, las tesis de grado y de posgrado, así como publicaciones científicas, entre otras. Estas producciones, aunque se almacenan y organizan en bibliotecas digitales personales o la biblioteca institucional, están aisladas, responden a normalizaciones heterogéneas, y presentan una gran disparidad en cuanto a niveles de descripción en la terminología utilizada. La desvinculación existente entre las mismas dificulta su visibilidad para los nuevos procesos de investigación, así como la interoperabilidad y la usabilidad de los recursos de conocimiento generados.

Dando continuidad a la línea de investigación culminada a principios de 2014, sobre objetos de aprendizaje y repositorios académicos para investigadores noveles, se ha iniciado una segunda fase de investigación que

propone el diseño y desarrollo de aplicaciones de software que permitan la creación y gestión de un repositorio recomendador ontológico, del conocimiento científico generado en la UNCAus: trabajos finales de grado y posgrado, proyectos de investigación y transferencia tecnológica y publicaciones, para mejorar la accesibilidad a este tipo de conocimiento y la usabilidad e interoperabilidad de la información almacenada. Este artículo presenta el estado de avance alcanzado en este proyecto, los resultados logrados y la formación de recursos humanos concretada en el marco del mismo.

Palabras clave: Metadatos, Ontologías, Repositorio académico

Contexto

Esta línea de I+D, se encuadra en el proyecto denominado “Sistemas de Búsquedas Semánticas en Repositorios Académicos”, convocatoria 2015-2018 de la Secretaría de Investigación, Ciencia y Tecnología de la UNCAus. Ha sido, evaluado externamente y financiado por dicha institución, y se orienta a ampliar la línea de investigación institucional en el área de Semánticas Web (PI27/2012).

Introducción

Aunque la proliferación de repositorios institucionales es un gran paso para la difusión y preservación de la ciencia y la cultura, este proceso viene acompañado de retos pues requiere un esfuerzo por parte de las instituciones para normalizar un conjunto de propiedades útiles para la descripción de recursos de información. Y es que los contenidos informacionales científicos se presentan en múltiples formatos (multimedias, fotos, libros, documentos de archivos, películas, *sites*, artículos) y se multiplican cada día, en volumen y características, ocasionado como contrapartida dificultades para acceder a esta información cuando se la necesita.

Esto ha motivado diferentes iniciativas en las universidades y centros de investigación del exterior y también algunas iniciativas en el ámbito nacional, orientadas a estructurar y organizar este conocimiento para lograr interoperabilidad consistente.

La UNCAus no es ajena a esta problemática. La desarticulación de las producciones científicas torna dificultoso definir algún parentesco programático entre dos o más de los trabajos de investigación existentes, y no contribuye a facilitar – acrecentar – vincular, la actividad científica. (Durán, 2014) (Zachman, 2013)

Considerando la organización de la información como elemento fundamental para la garantía de calidad en la recuperación, se presenta como desafío institucional la necesidad de singularización contextual en la reconstrucción del conocimiento, buscando proporcionar la recuperación y el uso de la información a partir de la identificación, requisitos de pertinencia y relevancia en contextos específicos.

Desde el punto de vista tecnológico, los metadatos constituyen una parte fundamental en los repositorios, ya que hacen posible que aumenten los niveles de confianza en la utilización de herramientas de búsqueda y consulta, favoreciendo mejoras en los procesos de representación, organización y recuperación de información en bibliotecas digitales. En

este caso, uno de los estándares más utilizados y reconocidos a nivel de metadatos es Dublin Core. (Dublin Core, 2013)

De acuerdo con Koper (2004) el surgimiento de la web semántica ha permitido obtener soluciones a problemas habituales en la búsqueda de información gracias a la utilización de una infraestructura común, mediante la cual, es posible compartir, procesar y transferir información de forma sencilla. En este contexto es donde comienzan a tomar protagonismo la web semántica, como solución a dichas necesidades.

Una de las herramientas que ofrece la web semántica para describir un dominio y facilitar las búsquedas son las ontologías. Estas poseen un enorme potencial para proporcionar homogeneidad entre recursos heterogéneos de acceso común (Castells, 2003).

En el modelo Web Semántica la información está dotada de un significado bien definido que permite una mejor colaboración entre humanos y máquinas. De esta forma se busca el marcado semántico de los recursos, lo que implica una separación formal entre el contenido y la estructura de los documentos y el desarrollo de agentes de software capaces de procesar y operar con estos recursos a nivel semántico utilizando ontologías, entendidas como conceptualizaciones formalizadas de una parcela de realidad que dan sentido pleno a la información, situándola en un contexto. (Gruber, 1995)

A partir de ello, desde el año 2015 se ha iniciado una nueva línea de investigación sobre ontologías aplicadas a la producción científica almacenada en repositorios institucionales, con el fin de modelar, desarrollar e implementar un repositorio basado en metadatos y ontologías, sobre la web de la Universidad.

En este trabajo se sintetizan los avances logrados en el proyecto desde su inicio (2012).

Línea de Investigación, Desarrollo e Innovación

La línea de trabajo 2012-2014 concluyó con el prototipo del repositorio institucional que incluyó interfaz para la gestión de usuarios y la

gestión de proyectos de investigación. Sólo incluyó el trabajo con proyectos de investigación realizándose las validaciones correspondientes, al mismo tiempo que se encuentra en proceso de carga de datos pertenecientes a proyectos de anteriores.

En la actual línea de investigación se está trabajando, tomando como ejes centrales los repositorios institucionales, las ontologías y los metadatos. La finalidad del proyecto es analizar, diseñar y/o reusar, construir, y evaluar ontologías para el repositorio digital de producciones científicas, destinada tanto a docentes, investigadores como a estudiantes. Así lograr que el sistema de gestión implementado y la estructura dada al repositorio permita inferir información que proporcione una comprensión común de un dominio determinado eliminando confusiones conceptuales y terminológicas.

Para ello, se han planteado los siguientes objetivos prioritarios:

1. Modelar estructuras semánticas como forma de describir los recursos informacionales ingenieriles, de tal manera que se encuentren interrelacionados e interoperables.
2. Desarrollar e implementar un repositorio ontológico, en la UNCAus.
3. Favorecer el desarrollo de conocimiento científico-tecnológico en el campo de las ontologías y repositorios web.

La metodología propuesta comprende; planificación, diseño conceptual y lógico, implantación, mantenimiento y evaluación del repositorio ontológico.

Resultados Alcanzados

Para lograr estos objetivos, y dando continuidad al proyecto iniciado en 2012, el equipo se encuentra validación y testeando el prototipo de repositorio centralizado intranet existente. Esta actividad incluye un análisis acerca de la normalización y clasificación de los recursos con el objeto de resolver los errores y gaps de objetivos y discrepancias de bases de datos encontradas con el objeto de un

acercamiento a la identificación y construcción de ontologías.

Dado que en la primera etapa piloto de desarrollo se incluyó en el repositorio únicamente a los Proyectos de Investigación, se está realizando una actualización del material recopilado que incluye tesis y proyectos finales de carrera de grado y posgrado para la futura expansión de los recursos ofrecidos por el repositorio.

Asimismo, se está trabajando en un modelo más avanzado de los metadatos y en un modelo preliminar de las ontologías que conforman el repositorio. donde se definan conceptos, se asocien significados y se construyan lógicas de inferencia basadas en la automatización y acceso a colecciones estructuradas de información, razonamientos y procesos cognitivos para investigadores en el área de Ingeniería.

Cabe acotar que el desarrollo se realiza como un módulo externo a los servicios web que ofrece la Universidad, sin embargo, a futuro está prevista la integración entre ambos módulos de servicios.

Como etapa final, se prevé el diseño de políticas institucionales para el uso del repositorio: acceso, estructuración jerárquica de contenidos, manejo y transferencia de contenidos, metabúsquedas, preservación, interoperabilidad.

El modelo de repositorio ontológico institucional, se considera innovador dentro de la región y de la provincia, pues no se conocen publicaciones referidas a repositorios ontológicos con dominio en el campo de las investigaciones en ingeniería.

Los primeros beneficios se evidencian en una mejora del flujo y gestión de recursos científicos para la iniciación y sustentabilidad de proyectos de investigación en Ingeniería, y en el suministro de esquemas estándares de metadatos que contribuyen a una articulación entre las líneas de investigación existentes.

Formación de Recursos Humanos

El equipo de trabajo está integrado por el Director, un Codirector y un docente

investigador y dos alumnos de los últimos años de la carrera Ingeniería en Sistemas de Información, más dos asesores en el tema de ontologías.

El proyecto ha permitido que un docente de la carrera Ingeniería en Sistemas de Información se inicie en actividades de investigación bajo la dirección de un investigador formado y categorizado en el sistema de incentivos. Por otra parte, uno de los estudiantes ha finalizado su pregrado con el diseño web del repositorio digital de la institución y se encuentra en proceso de iniciar su Proyecto Final de carrera.

Referencias

- Castells, P., *La Web Semántica*. Escuela Politécnica Superior, Universidad Autónoma de Madrid. Conferencia impartida en el Curso de Verano sobre Interacción en la Web, Universidad de Castilla - La Mancha, Puertollano, Septiembre 2003.
- Dublin Core Metadata Initiative. Página Web, . Recuperado Septiembre 19, 2013, a partir de <http://www.dublincore.org>.
- Durán E., Zachman P., Alvarez M., Aguilera J, Nuñez J., Umaño M. *Repositorio Ontológico Institucional*. Anales de XV Workshop de Investigadores en Ciencias de la Computación (WICC 2013), ISBN: 9789872817961. 2013 Disponible en: <http://sedici.unlp.edu.ar/handle/10915/27073?show=full>, 2013
- Durán E Zachman P., Alvarez M. Umaño M., Nuñez J., Aguilera, J., *Metadatos y Ontologías aplicados al diseño de Repositorios Institucionales*. XIV Workshop de Investigadores en Ciencias de la Computación WICC 2014, Disponible en [//sedici.unlp.edu.ar](http://sedici.unlp.edu.ar), 2014
- Gruber, T. R. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal of Human and Computer Studies*, 43(5/6): 907-928, 1995
- Koper, R., *Use of the Semantic Web to Solve Some Basic Problems in Education: Increase Flexible, Distributed Lifelong Learning. Decrease Teachers' Workload*. *Journal of Interactive Media in Education. Special Issue on the Educational Semantic Web*, 2004
- Pedraza-Jiménez, R., Codina, L. & Rovira, C. *Web semántica y ontologías en el procesamiento de la información documental*. *El profesional de la Información* 16, 10, 2007.
- Zachman P., Umaño M., Nuñez J., Aguilera, J., Durán E. *Arquitectura de un Sistema de Gestión de un Repositorio Ontológico Institucional de Apoyo a Investigadores Noveles*. Actas del 3er Congreso Iberoamericano de Ingeniería de Proyectos, Noviembre 2012, Mar del Plata, Argentina. Disponible en: <http://sites.riipro.org/ciip/home>., 2012
- Zachman P., Durán E., Aguilera J. *Una aproximación ontológica al modelado de un repositorio académico-científico para investigadores noveles*. Anales IX Jornadas de Ciencia y Tecnología de Facultades de Ingeniería del NOA, 2013, Libro de Resúmenes pag. 100, Edición completa en CD. ISSN 1853-7871, 2013