

Informe Final Beca de Entrenamiento 2014

BECA DE Entrenamiento PERIODO 2014-2015

1. DATOS PERSONALES

APELLIDO: Terruzzi

NOMBRES: Franco Agustín

Dirección electrónica: agustinterruzzi@sedici.unlp.edu.ar

2. TEMA DE INVESTIGACIÓN (Debe adjuntarse copia del plan de actividades presentado con la solicitud de Beca)

Estrategias para la interoperabilidad entre sistemas de gestión de producción académica y científica

3. INSTITUCIÓN DONDE DESARROLLA LOS TRABAJOS

Universidad y/o Centro: Universidad Nacional de La Plata

Facultad: Facultad De Ingeniería

Departamento: Ciencias Básicas

Cátedra: -

Otros: PREBI-SEDICI Universidad Nacional de La Plata

Dirección: Calle: 49 y 115 *Nº:* s/n

Localidad: La Plata *CP:* 1900 *Tel:* (0221) 423-6696/6677 int. 141

4. DIRECTOR DE BECA

Apellido y Nombres: De Giusti, Marisa Raquel

Dirección electrónica: marisa.degiusti@sedici.unlp.edu.ar

5. EXPOSICIÓN SINTÉTICA DE LAS TAREAS DESARROLLADAS EN EL PERIODO.

Durante el período de la beca, se realizaron distintas tareas relacionadas al tema de investigación planteado en el plan de actividades de la misma, a continuación se listan y describen sintéticamente cada una:

- Estudio de alternativas de interoperabilidad máquina-máquina: mediante ciertas metodologías que serán descritas en el inciso 6, se investigaron distintos protocolos de interoperabilidad conocidos y se establecieron conclusiones sobre la proporción costo-beneficio de aplicarlos en el repositorio institucional, la gestión de portales, la preservación de contenidos, la recuperación de datos, etc. Algunos protocolos

estudiados fueron: XML/XSL, RSS/Atom, SWORD (v1 y v2), OAI PMH, OAI ORE, OpenSearch, etc.

- Actualización de servicios para interoperabilidad: con el objetivo de adecuar ciertos servicios del repositorio institucional, para su comunicación a través de alguno de los protocolos mencionados en el párrafo anterior, se realizaron diversas actualizaciones (explicadas con más detalle en el inciso 6), para incorporar módulos de interoperabilidad a estos sistemas. Algunos de los sistemas que presentaron cambios fueron: el Portal de Revistas, el Portal de Congresos, el Portal de Revistas Científicas y el SEDICI (al menos en una versión para pruebas de éste).
- Estudio de interoperabilidad y comunicación humano-máquina: en el marco de esta beca, también se estudiaron técnicas para mejorar la interacción entre el humano y la máquina a través de diversas interfaces orientadas a extracción de información en sistemas de repositorios digitales. Esto significó el abordaje de una nueva línea de estudio: el desarrollo de lenguajes específicos de dominio para tareas de gestión y administración de repositorios. Esta tarea se vio concretada en la tesina de grado, descrita en el inciso 7.

6. RESÚMENES DE LOS MÉTODOS Y TÉCNICAS EMPLEADAS (así como del instrumental y métodos de cálculos utilizados).

A continuación se describe la metodología empleada para cada una de las tareas descritas en el inciso anterior (inciso 5).

- Estudio de alternativas de interoperabilidad: se investigaron distintos protocolos y su aplicabilidad a la problemática diaria del repositorio institucional. Para realizar esta tarea se buscó en la documentación de los distintos protocolos y se estudiaron formas de emplearlos en los servicios que brinda el repositorio institucional de la UNLP. Para cada protocolo, se propusieron distintas formas de aplicación a los servicios del repositorio. Se evaluaron los costos de implementación de estas alternativas y se aplicaron aquellas cuyos costos son proporcionales a los beneficios que brindan. Los protocolos sobre los que se hicieron avances en el período de esta beca fueron:
 - **RSS/ATOM:** se estudiaron estos protocolos de interoperabilidad para brindar la posibilidad de actualizar la información que mostraban ciertos módulos existentes en servicios del repositorio como el Portal de Revistas Científicas. Mediante la utilización del protocolo RSS/ATOM se implementó y configuró un conector para el portal, que permite mostrar tanto los últimos artículos de una revista en particular (cargados en SEDICI), como todos los números correspondientes a esa revista. El conector se utiliza para todas las revistas del portal y ha sido muy útil para mantener la información actualizada. En el enlace : <http://revistas.unlp.edu.ar/cientificas/?p=309>, se puede ver un ejemplo de su uso.
 - **SWORD:** se analizaron las dos versiones del protocolo SWORD. En primer lugar se mejoró la configuración del módulo Swordv1 que se utiliza en el Portal de Revistas de la UNLP (mediante el mismo se realiza autoarchivo remoto de los nuevos números de las revistas del portal, en el repositorio institucional), con el objetivo de conseguir que se enviaran más datos desde el portal de revistas, para que los gestores del SEDICI, que deben revisar los autoarchivos como parte del flujo de trabajo, tengan menos datos que rellenar para completar la información. Esto se logró mediante la modificación del conector del Portal de Revistas y la realización de cambios en el *crosswalk* (el archivo encargado de los mapeos de metadatos) en la versión de DSpace utilizada para gestionar el SEDICI. En la actualidad, se continúan

realizando depósitos remotos de los nuevos números de la revistas del Portal.

En segundo lugar, se estudió la adaptación de módulo que utiliza el Portal de Revistas para que emplee la segunda versión del protocolo Sword (Swordv2), con el objetivo de aprovechar las ventajas que este ofrece: actualización, eliminación y sincronización de los objetos remotos, para asegurar que los depósitos se realicen correctamente. Como parte de este análisis, se tuvo en cuenta que ya disponía de una herramienta en el software del repositorio, que brindaba soporte para Swordv2. No obstante, a la hora de integrar los dos servicios, se encontraron distintos errores en las versiones actuales de los software utilizados, tanto en el lado del Portal de Revistas (OJS), como del lado del SEDICI (DSpace), por lo que se decidió que la incorporación de Swordv2 es demasiado costosa en implementación por el momento, y se la propuso como trabajo futuro, cuando los servicios se hayan migrado a su última versión.

- **OAI-PMH:** el análisis de este protocolo se produjo como consecuencia del desarrollo de otra tarea que se especifica más adelante: el estudio de la comunicación y la interoperabilidad humano-máquina. Como parte de esta tarea, fue necesario analizar el módulo de OAI del software de SEDICI (DSpace), y comprender cómo funciona mediante su utilización en un entorno real, para mejorar la configuración del módulo, de modo que sea más sencillo y eficiente de gestionar para un usuario administrador.
- **OpenSearch:** se investigó la documentación y se analizó la implementación de este protocolo en distintos módulos de los portales, con el objetivo de favorecer a la recuperación automática de información, desde el servicio de OpenSearch que ya ofrece SEDICI. Como parte de esta investigación, se colaboró en el testeo y la configuración de una herramienta que busca recuperar ciertos datos específicos del repositorio institucional, en el marco de una iniciativa más amplia: el proyecto de visibilidad para sitios institucionales de la UNLP (que se describe mejor en el inciso “actualización de servicios para interoperabilidad”). Esta herramienta quedó configurada y se puede ver un ejemplo de su utilización en el siguiente enlace: <http://lisea.uids.testing.sedici.unlp.edu.ar/investigacion/publicaciones/>
- **Actualización de servicios para interoperabilidad:** de manera conjunta al estudio de los diferentes mecanismos y protocolos de interoperabilidad, fue necesario actualizar y adaptar los distintos servicios y las herramientas que utilizaban estos servicios primero en el repositorio institucional (SEDICI) y, de manera más amplia, en la Universidad Nacional de La Plata. La metodología para esta tarea principalmente consistió en la actualización de los distintos softwares utilizados para gestionar el Portal de Congresos (OCS), el Portal de Revistas (OJS) y el Portal de Revistas Científicas (Wordpress) a sus últimas versiones, compatibles con muchos de los protocolos de interoperabilidad ya mencionados.

No obstante, también se trabajó en un proyecto más amplio: la visibilidad de la UNLP. Dentro de este marco de trabajo se realizaron actividades para la interoperabilidad entre los distintos centros de investigación pertenecientes a la universidad, incluso con una participación activa en el desarrollo de un Portal de Sitios Institucionales, con el objetivo de aumentar la visibilidad de aquellos centros de investigación pertenecientes a la UNLP que no poseían un espacio para compartir el trabajo científico que realizaban. Todo esto significó muchos avances en el marco de la interoperabilidad, puesto que gran parte de la información a mostrar,

se encuentra disponible en el SEDICI, y el hecho de poder interoperar con el repositorio, facilitó el desarrollo del portal y de los sitios particulares.

- Estudio de interoperabilidad y comunicación humano-máquina: además de la interoperabilidad entre sistemas automáticos, durante el período de la beca se encontraron nuevas líneas de investigación e incluso se realizaron trabajos (ver inciso 7) sobre estos nuevos marcos teóricos. Con el objetivo de mejorar la experiencia del usuario y de facilitar ciertas tareas que son recurrentes en el repositorio y que tienen que ver con los mecanismos existentes para el control de la calidad de los datos, se realizaron estudios sobre distintos módulos del software, tales como el módulo de curación: que brinda la posibilidad de programar a través de una interfaz, tareas que realizan controles personalizados sobre elementos del repositorio.

De manera conjunta, se investigaron formas de facilitar las interfaces sobre las cuales se ejecutaban los controles de calidad. En este contexto, se trabajó sobre una nueva línea de investigación: el desarrollo de lenguajes, específicamente de lenguajes específicos de dominio. El objetivo de esto fue el desarrollo de una herramienta que permite mejorar la comunicación humano-máquina para definir los mecanismos de control de calidad, de manera tal que ciertos usuarios con conocimientos en el contexto de repositorios, pero no demasiado técnicos, puedan configurar y ejecutar los mecanismos, sin necesidad de recurrir a una solución programática. Esto incluso fue motivo de un trabajo de investigación, publicado y concretado en una tesina de grado (se define en el inciso 7).

7. TRABAJOS DE INVESTIGACIÓN REALIZADOS O PUBLICADOS EN EL PERIODO.

Tesina De Grado

Título: Herramienta de validación aplicada a las tareas de gestión de calidad en un repositorio digital

Autores: Franco Agustín Terruzzi

Director: Dra. Marisa R. De Giusti

Codirector: -

Asesor profesional: Dr. Gonzalo L. Villarreal

Asesor profesional: Lic. Ariel J. Lira

Carrera: Licenciatura en Sistemas

Resumen:

Esta tesina describe el trabajo realizado para la implementación de un aporte a los mecanismos de control de calidad existentes en un Repositorio Institucional. Se parte con el objetivo de desarrollar una herramienta dinámica de validación, que permita a los administradores del repositorio, la posibilidad de implementar ciertas tareas enmarcadas en el ámbito de la preservación de los objetos digitales. El desarrollo de esta herramienta se enmarca en el software DSpace, el sistema de gestión que utiliza el repositorio institucional de la UNLP, el SEDICI.

Para realizar el desarrollo se aborda la temática de los lenguajes específicos de dominio, y se realiza un aporte en el diseño y el análisis de un lenguaje específico de dominio (DSL) que reúna los conceptos pertenecientes al dominio de los repositorios gestionados con DSpace. Para esto, se aplican distintos conceptos pertenecientes a esta rama de investigación: metamodelos, árboles AST, sintaxis abstracta y concreta, etc.

Después, se describe la implementación desarrollada a partir del aporte mencionado en el párrafo anterior, y informa acerca de la herramienta de desarrollo escogida para la implementación: el lenguaje de expresiones especificado en la JSR-341.

Finalmente, se documenta un caso de uso puntual para el desarrollo realizado: las tareas de curación.

Publicaciones

Revista: Scholarly and Research Communication

ISSN: 1923-0702

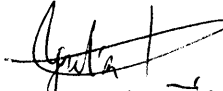
URL: <http://src-online.ca/index.php/src/about>

Título del artículo: Fostering the institutional repository through policies and interoperability with online services: the case of Universidad Nacional de La Plata"

Autores: Villarreal Gonzalo Luján, Marisa Raquel De Giusti, Franco Agustín Terruzzi

Estado: (A publicar)

URL: <http://src-online.ca/index.php/src/issue/view/19>


FRANCO AGUSTÍN TERRUZZI
DNI: 26.625.848

Informe correcto

SeGi

DE GIUSTI MARISAR.