

Generación automática de resúmenes de congresos e integración con el Repositorio Institucional

De Giusti, Marisa Raquel*; Terruzzi, Franco Agustín; Villarreal, Gonzalo Luján**

* PrEBi UNLP y Comisión de Investigaciones Científicas (CIC) de la Provincia de Buenos Aires.
La Plata, Argentina

† PrEBi UNLP. La Plata, Argentina.

** PrEBi UNLP y Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). La Plata, Argentina

Resumen

La Universidad Nacional de La Plata (UNLP) organiza anualmente distintos tipos de reuniones académicas y científicas a través de sus docentes, investigadores, departamentos y laboratorios. Estas reuniones abarcan congresos, simposios, conferencias, seminarios, presentaciones, etcétera (por simplicidad, se utilizará el término congreso para referirse a todos los tipos de reuniones). Además de apoyar y fomentar la organización de congresos por parte de sus docentes e investigadores, la UNLP busca también preservar y difundir toda la creación intelectual generada a partir de los mismos, generalmente en forma de artículos, presentaciones, posters e informes. Para facilitar las tareas relativas a la gestión y organización de congresos, en el año 2009 se creó el Portal de Congresos de la UNLP. El mismo ha sido utilizado por más de 20 congresos regionales, nacionales e internacionales hasta la fecha. El portal web cuenta con un equipo técnico que realiza diferentes actividades para brindar capacitación y soporte a los organizadores de congresos, como reuniones, talleres y comunicación por correo electrónico. Otras, en cambio, abarcan tareas transparentes para los usuarios de mantenimiento y actualización del software OCS que soporta al portal y del servidor donde el mismo se aloja. También se genera documentación en una wiki, que permite a los gestores de congresos comprender cómo dar los primeros pasos en el uso del portal así como también realizar algunas configuraciones avanzadas comunes a la mayoría de los congresos.

Durante el tiempo que se ha brindado soporte y atención a usuarios, el equipo del Portal de Congresos ha detectado algunas necesidades comunes a la mayoría de los usuarios antes, durante y después de la realización de los congresos. Entre las necesidades posteriores a la realización de los congresos, se observó que la elaboración de los libros de resúmenes se realizaba por lo general de manera artesanal, es decir, que los organizadores de cada congreso debían gestionar la conformación de las actas manualmente, recopilando todas las versiones finales de los trabajos, de a una a la vez. Esto les insume una importante dedicación de tiempo y a más de una persona en varios casos. Además, este mecanismo manual requiere una duplicación de esfuerzos por parte del equipo del SEDICI (Servicio de Difusión de la Creación Intelectual), el repositorio institucional de la UNLP. Las tareas de recopilación de la producción intelectual de los congresos, su incorporación al repositorio institucional, su correcta catalogación y difusión requieren desglosar manualmente cada uno de los trabajos de las actas ya generadas. Evidentemente, se hizo necesario en establecimiento de un flujo de trabajo que evitase la

duplicación de esfuerzos y que simplificara la integración entre el Portal de Congresos y el repositorio institucional de la UNLP.

Durante el año 2012, el equipo del Portal de Congresos de la UNLP desarrolló un plugin para el software OCS que permite automatizar la generación de actas de congresos, y que a la vez simplifica de manera considerable la integración de dichas actas con el SEDICI. Este plugin puede, en líneas generales, exportar información de un mismo congreso en diferentes formatos, según se requiera en cada caso. Se han implementado distintos formatos de exportación, como por ejemplo CSV (valores separados por coma), HTML e incluso documento de texto (Word). Estas tres exportaciones toman los datos de todos los trabajos (título, autores, resumen, instituciones, etc.) y los combinan en un único archivo, que el gestor de cada congreso puede descargar. También se implementó la exportación de todos los artículos de cada congreso en un único archivo .zip, y se modificaron los nombres de los archivos incluidos en el archivo .zip para que adopten el título del trabajo al que corresponden. De esta forma, los gestores pueden generar los libros de resúmenes de manera muy simple, y el repositorio institucional puede cargar todos los trabajos de cada congreso de una forma mucho más rápida.

Si bien se han implementado hasta ahora algunos formatos de exportación, el plugin fue pensado desde el principio para ser extendido hacia nuevos formatos, mediante un diseño basado en objetos que implementa el patrón de diseño conocido como Template Method.

Esta herramienta fue puesta a disposibilidad para usuarios en varios congresos del Portal de Congresos de la UNLP a modo de pruebas y los resultados fueron satisfactorios. En la actualidad, la herramienta ya está siendo utilizada para incorporar las producciones de los congresos al repositorio institucional SEDICI.

Aún cuando este desarrollo permitió simplificar la generación de actas de resúmenes de congresos, puede optimizarse más la incorporación de los trabajos al repositorio institucional, a fin de evitar la carga manual de los mismos. Por este motivo, el equipo del Portal de Congresos se encuentra evaluando diversas tecnologías de interoperabilidad entre sistemas. Entre ellas se destaca el protocolo sword en su versión 2 como una opción muy prometedora y relativamente simple. Este protocolo permite a los gestores de los congresos realizar el depósito automático o semiautomático de los trabajos desde el Portal de Congresos en el repositorio institucional, soportado por Dspace. Dicho depósito puede realizarse sobre una colección privada, lo que permitirá que el equipo de SEDICI revise los trabajos antes de confirmar su incorporación definitiva al repositorio.

Introducción

Una gran parte de la producción académica de la Universidad Nacional de la Plata [1] resulta de distintas reuniones de índole académico que sus miembros organizan, ya sean conferencias, congresos, simposios, etcétera. Con el objetivo de facilitar la gestión de este tipo de eventos y de albergar, administrar y resguardar la producción académica surgida como consecuencia de la organización de los mismos, se estableció el Portal de Congresos de la UNLP [2] como un servicio y una herramienta en línea tanto para la gestión como para la difusión de los distintos eventos.

El Portal de Congresos de la UNLP se encuentra disponible para todos los miembros de esta universidad –docentes, investigadores, laboratorios, cátedras, departamentos, etcétera– que organicen congresos y deseen realizar la gestión en línea de los mismos. El portal permite a cada congreso administrarse de manera autónoma y utilizar una gran variedad de funciones diseñadas para facilitar la gestión de los mismos: registro de autores, sistemas de revisión de trabajos, formato de publicaciones, organización del calendario del congreso, registro de participantes, etcétera. El portal está soportado por la plataforma Open Conference Systems (OCS) [3] del Public Knowledge Project (PKP), que agrupa a universidades e instituciones de renombre internacional como British Columbia, Simon Fraser y Stanford University. Esta plataforma, personalizada por PREBI-SEDICI [4,5] para la Universidad Nacional de La Plata, ofrece numerosas funciones y una amplia flexibilidad que le permite adaptarse a las diferentes necesidades de cada congreso (congreso, simposio, conferencia, workshop, mesa de debate, etcétera). Entre estas funciones y herramientas se incluyen:

- Personalización estética del portal: encabezados, colores y fuentes, fondos, logos.
- Soporte para múltiples idiomas
- Carga de textos de las distintas secciones del congreso con información para autores, revisores, participantes, llamado a trabajos, alojamiento, entre otras.
- Difusión organizada por secciones de la información del congreso para los visitantes del portal: programa, equipo organizador, fechas importantes, formas de pago, etcétera.
- Configuración de la metodología de publicación de presentaciones, envío de trabajos, circuito de revisión y suscripción al congreso.
- Gestión del congreso en etapas por medio del calendario.
- Configuración de mecanismos de acceso en línea a los trabajos publicados: abierto y gratuito, solo usuarios registrados, mediante suscripción.
- Publicación de mecanismos de inscripción y políticas del registro: tipos de inscripción, información sobre la inscripción, opciones de registro.
- Generación de estadísticas, sistemas de importación y exportación de listados, y reportes.

Desde su creación en el año 2009, se ha brindado asistencia personalizada a muchos organizadores de congresos, y se han establecido mecanismos y herramientas de ayuda para facilitar el trabajo de aquellos usuarios no familiarizados con este sistema. Para brindar este tipo de asistencia se reciben consultas mediante correo electrónico, se organizan reuniones para tratar aspectos específicos a cada conferencia, se redactan documentos de ayuda (FAQ, manuales de usuario, documento de interfaz, wiki [6]), se elaboran traducciones, se asiste a los usuarios con el procedimiento de carga de la información y se brindan recomendaciones acerca de los formatos adecuados para la misma. Además se realizan esfuerzos más específicos como el desarrollo o la incorporación de herramientas de carácter técnico para asistir a usuarios más experimentados con necesidades concretas. Esto abarca desde el desarrollo y/o instalación de nuevos plugins para el portal, hasta la implementación de cambios en el flujo de trabajo preestablecido por el sistema. Todas las modificaciones de la herramienta desarrollada por PKP son puestas al alcance de la comunidad.

El portal de congresos ha sido utilizado por un muchos usuarios que conforman los distintos equipos organizadores de los congresos, de quienes se ha obtenido retroalimentación sobre los usos más comunes, sus preferencias y sus necesidades. A partir de allí, se ha observado que muchos de ellos han solicitado asistencia durante la fase de generación de informes posteriores a la finalización del congreso, incluyendo listados de trabajos publicados y de sus autores, libros de resúmenes, y actas del congreso. Si bien el software OCS posee algunas funciones de generación de reportes, la elaboración de un libro de resúmenes requiere un trabajo fundamentalmente artesanal: recopilación desde las distintas secciones de la administración del portal del congreso de toda esta información (título de cada trabajo, autores, instituciones y resúmenes), y posterior compaginación de un único archivo de texto. Para facilitar esta tarea a los organizadores de congresos, se desarrolló una herramienta de generación de múltiples reportes.

Desde un enfoque global a todas las conferencias, ha surgido la necesidad de establecer un flujo de trabajo que defina los criterios y mecanismos mínimos para asegurar la transferencia de información entre los distintos sistemas de gestión. Este requerimiento no fue relevado a partir de los usuarios del portal, sino que proviene desde otros espacios de la universidad que utilizan la información generada en cada congreso con distintos propósitos. El envío y recepción simple de información desde las distintas herramientas de administración de la producción académica resultaría muy beneficioso para las distintas dependencias, que ahorrarán en tiempo y se reducirá considerablemente la tasa de errores por copiar información entre sistemas. Un caso concreto aquí está representado por la interacción con el repositorio institucional de la UNLP, conocido como Servicio de Difusión de la Creación Intelectual (SEDICI). Para posibilitar la misma, se estableció como primera medida un método de exportación que facilita la recopilación de la producción intelectual que se encuentra presente en cada congreso gestionado con el portal. Se buscaron aprovechar las construcciones ya mencionadas (actas, informes, listados) constituidas en cada congreso y con estas generar una exportación que pueda ser manipulada desde el repositorio.

Como una posible solución a estas dos problemáticas, se construyó la herramienta generadora de reportes mencionada previamente. La misma se integra con el software OCS como un *plugin* de reportes, y se encarga de recopilar esta información y generar un documento según las preferencias del organizador. A continuación se detallan los principales aspectos de este desarrollo y los resultados alcanzados..

Detalles del desarrollo

La solución desarrollada combina algunas herramientas ya disponibles en OCS, como una categoría de “plugins de reportes” [7] bajo la que se encuentran todas las extensiones dedicadas a la exportación de datos. Se destacan en particular el plugin generador de informes de revisión, el de reportar artículos y el dedicado a mostrar información sobre los usuarios registrados en los distintos congresos. Mediante la utilización de los informes generados con estos plugins, cada organizador de un congreso tiene un mecanismo para obtener datos crudos (artículos, metadatos de los artículos, revisiones, participantes del congreso, usuarios registrados, etc) y estadísticas a partir de estos datos (cantidad de artículos publicados, de usuarios registrados, de participantes al congreso, porcentaje de artículos rechazados, etc).

Los usuarios gestores del portal de congresos utilizan constantemente varias de estas herramientas. Si bien la información obtenida resulta útil bajo ciertos contextos, se encuentra demasiado dispersa y el formato de presentación es en muchos casos difícil de interpretar. Sumado a esto, no siempre toda la información obtenida es útil para todos los usuarios, pues cada organizador requiere distinta información y estadísticas a partir de los informes generados por estas herramientas: listados de participantes, trabajos completos y sus autores, cantidad de visitas al portal, cantidad de descargas de cada trabajo, etcétera. Por último, la construcción de las actas de congresos requiere la recopilación de mucha de la información que contiene el portal, como ser títulos y resúmenes de los trabajos, datos de los autores, y finalmente los trabajos completos. Toda esta información se combina y presenta bajo distintos formatos como listados, tablas, documentos de texto plano, entre otros.

Con el objetivo de brindar una solución que facilite las tareas arriba mencionadas, se desarrolló un nuevo plugin, llamado Plugin Multi-Generador de Reportes [8]. Este plugin permite a los organizadores de cada congreso generar reportes completos sobre los artículos publicados en los mismos. Los reportes pueden mostrar información sobre los trabajos, las secciones en las que están categorizados, los datos de sus autores (nombre, apellido, filiación, documento, etc.) y los datos del congreso en cuestión (título del congreso, lugar, fecha de apertura y fecha de cierre,etc.). Los reportes generados pueden ser exportados en varios formatos, que van desde tablas (.csv), hasta documentos en formato HTML.

El plugin, desarrollado por el equipo del PrEBI-SEDICI de la UNLP, puede adaptarse a diferentes requerimientos, cambiando por ejemplo el estilo de la información o seleccionando qué información se quiere mostrar. Su utilización aporta principalmente productividad para los organizadores de los congresos, que pueden generar un libro de resúmenes con pocos clics. El diseño de esta herramienta fue pensado para ser adaptado a cambios en los formatos de exportación existentes, y para ser expandido con nuevos *exportadores*¹, a fin de generar nuevos formatos o incluir etapas de procesamiento de la información antes de realizar la exportación.

Uno de los requerimientos comunes de muchos usuarios era la obtención de datos similares para distintos usos, para lo cual requerían diferentes presentaciones: listado plano, planilla de cálculo, documento de texto, HTML, entre otros. Esto requiere, como primera medida, la obtención de los datos de la base de datos, más allá del formato con el cual los mismos serán exportados. A continuación, se requiere iterar sobre el conjunto de datos, y aplicar sobre cada uno las transformaciones y operaciones necesarias para conseguir el formato deseado. Si bien estas operaciones y transformaciones son particulares a cada formato de salida, algunos formatos comparten ciertas características comunes que implican realizar las mismas operaciones y transformaciones. Por ejemplo, los diferentes mecanismos de exportación requieren, como

¹Formatos distintos para generar las exportaciones de reportes

mínimo, iterar sobre la lista de autores de un trabajo, y para cada autor, sobre sus datos personales y sus instituciones.

Mediante la aplicación del patrón de diseño *Template Method*[9], es posible reproducir este comportamiento común y especificar el comportamiento particular de cada formato de exportación. En líneas generales, este patrón permite, a través de una jerarquía de clases, definir ciertas etapas genéricas de un algoritmo en una superclase y redefinir en las subclases solamente las operaciones específicas del algoritmo. En este contexto, la implementación de este patrón soluciona el problema de la diversidad de los formatos de presentación, mientras que también permite generalizar la manera de procesar y obtener los datos a exportar. Como resultado de esta implementación se obtiene un diseño fácilmente escalable, que permite construir diversos tipos de reportes variando no solo la presentación sino también los datos que son mostrados.

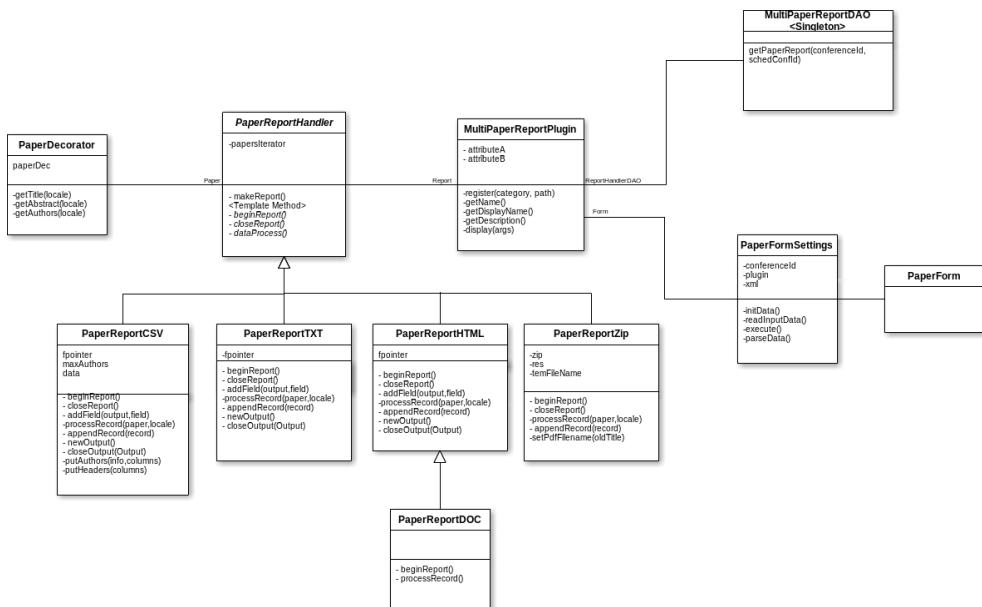


Figura 1. Diagrama UML del plugin Multi-Generador de Reportes.

En el diseño propuesto (figura 1) se organizan los distintos tipos de reportes en una jerarquía de clases, en la cual la clase abstracta *PaperReportHandler* contiene, dentro del método *MakeReport*, el esqueleto básico para la generación de cualquier reporte. Este esqueleto consiste en tres pasos genéricos: *beginReport()*, *dataProcess()* y *closeReport()*, los que deben ser implementados por las subclases de dicha jerarquía. El método *dataProcess()*, en particular, se comporta también como el patrón *TemplateMethod*, ya que exige que las subclases implementen los métodos *processRecord()* y *appendRecord()*; cabe destacar que las subclases pueden también definir estos dos métodos, y no definir el método *dataProcess()*, ya que heredarán el comportamiento genérico de la superclase *PaperReportHandler*. Finalmente, cada una de las subclases resultantes son las encargadas de generar los diferentes reportes. Algunos ejemplos ya implementados son *PaperReportHTML* que genera documentos HTML y define una estructura jerárquica para cada artículo, *PaperReportCSV* que genera archivos en formato CSV (valores separados por comas) útiles para trabajar como cualquier planilla de cálculo, y *PaperReportDoc* que genera documentos de texto con formato (.doc). Además, se cuenta con la clase *PaperReportZip* que da la posibilidad de generar un paquete con todas las versiones galeradas de los trabajos que conforman un congreso determinado, ordenado por los títulos de los mismos. Con este paquete se pueden cargar los trabajos en el repositorio institucional, evitando al administrador de dicho repositorio la tarea de buscar y descargar a mano cada uno de los trabajos desde el portal de congresos..

El diseño arriba descrito permite generar nuevos formatos de presentación de reportes creando simplemente una subclase de *PaperReportHandler* que implemente los tres métodos mencionados. Adicionalmente, esta solución permite la generación de nuevos formatos derivados de otros formatos, ya que sólo es necesario generar una nueva clase que tenga como superclase

a aquella que genera el formato a derivar. En esta nueva subclase, deberá definirse solamente el o los métodos que modifiquen el resultado final. Este concepto fue aplicado para la generación de documentos de texto, cuya clase *PaperReportDoc* es subclase de *PaperReportHtml*, y solo redefine los métodos *beginReport()* (para generar los encabezados del documento) y *processRecord()* (para generar una la estructura apropiada de cada registro antes de ser almacenado en el archivo de salida).

La obtención de datos desde la base de datos fue implementada dentro de la clase *MultiPaperReportDAO*, a fin de encapsular este comportamiento particular y reusarlo de en todas las clases que generar reportes, sin importar el formato de exportación de los mismos.[Figura 2]

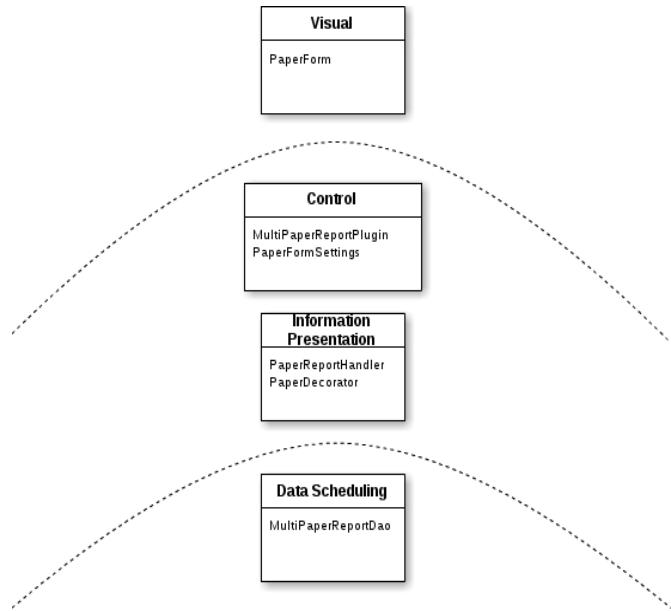


Figura 2 .Diagrama utilizado en el diseño del plugin.

Conclusiones

Durante la organización de un congreso, es evidente la necesidad de contar con información sobre presentaciones, autores, sesiones, etc..., así como también de automatizar tanto como sea posible la generación de reportes, actas, y documentos en general. Asimismo, es importante brindar mecanismos que faciliten la interoperabilidad entre el sistema de gestión del congreso y otros sistemas de la misma organización, como por ejemplo su propio repositorio institucional. En este trabajo, se presenta una extensión (plugin) del software OCS que permite generar información bajo múltiples formatos, y que fue pensado principalmente para ser ampliado con nuevos formatos o derivaciones de los formatos ya existentes. El código del plugin ha sido liberado bajo licencia GPL v2 [10] y subido al repositorio de plugins de PKP en un paquete que contiene el desarrollo y la documentación requerida. A la fecha, esta herramienta fue puesta a disponibilidad por los organizadores de algunos congresos del Portal de Congresos de la UNLP a modo de pruebas, y de acuerdo los comentarios enviados la herramienta ha resultado útil para los usuarios. Dentro de los mensajes enviados, los usuarios han solicitado nuevos formatos o algunos cambios en los formatos generados, a fin de automatizar aún más la generación de reportes, documentos y actas de resúmenes.

Con respecto a la interoperabilidad del portal con el repositorio, la exportación de un paquete comprimido de todos los trabajos del congreso ha simplificado considerablemente el trabajo de los administradores del repositorio, que pueden valerse del reporte en formato textual para extraer los metadatos de cada trabajo, y del paquete .zip para recuperar el documento completo. Sin

embargo, se sigue trabajando de manera manual, cargando uno a uno los trabajos del congreso en el repositorio. Con el fin de automatizar esta tarea, al menos parcialmente, se está considerando una posible ampliación de este plugin o la construcción de uno diferente que permita comunicar de manera transparente las distintas herramientas. Esto permitirá enviar de forma automática o semi-automática los documentos enviados al congreso hacia el repositorio institucional, donde se recibirán y procesarán de manera mucho más ágil. Para realizar esto, se están estudiando protocolos de interoperabilidad tales como OAI PMH [11], Swordv1 y Swordv2 [12] y Rss/Atom [13], que brindan lo necesario para generar una solución en relativamente poco tiempo.

Referencias

- [1] Universidad Nacional de La Plata. Web. 11 de Junio de 2013. www.unlp.edu.ar
- [2] Portal de Congresos de la UNLP. Web. 11 de Junio de 2013. www.congresos.unlp.edu.ar
- [3] OCS in an Hour, version 2.1. PKP, John Willinsky, Charles Eckman
- [4] PrEBI | Proyecto de Enlace de Bibliotecas. Web. 11 de Junio de 2013. <http://prebi.unlp.edu.ar/>
- [5] Servicio de Difusión de la Creación Intelectual de la UNLP. Web. 11 de Junio de 2013. <http://sedici.unlp.edu.ar/>
- [6] Wiki-PREBI. Web. 11 de Junio de 2013. <http://wiki.prebi.unlp.edu.ar/wiki/>
- [7] van Heesch, Dimitri, Doxygen. *Plugins, OCS API Reference*. Sep 20 2012. En línea
- [8] "Multiple report plugin". Web. 11 de Junio de 2013. <http://pkp.sfu.ca/support/forum/viewtopic.php?f=28&t=8810>
- [9] Gamma Erich, Helm Richard, Johnson Ralph, Vlissides John. *Design Patterns: Elements of Reusable Object-Oriented*. Addison-Wesley Professional, 31 de Octubre de 1994. Impreso.
- [10] GNU General Public License v2.0 - GNU Project - Free Software Foundation. Web. 11 de Junio de 2013. www.gnu.org/licenses/gpl-2.0.html
- [11] Rusch-Feja, D. "The Open Archives Initiative and the OAI Protocol for Metadata
- [12] R, Jones, Stuart Lewis. *Sword 2.0 Profile*. swordapp.org, 20 de Marzo de 2012. En línea
- [13] M, Nottingham, R. Sayre. The Atom Syndication Format. RFC 4287, Diciembre 2005. En línea