



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

DESARROLLO DE UN REPOSITORIO DE ARTÍCULOS CIENTÍFICOS

TESIS PARA OPTAR AL TÍTULO DE INGENIERO CIVIL EN COMPUTACIÓN

FELIPE IGNACIO CHACÓN CANDIA

PROFESOR GUÍA:

JOSÉ PINO URTUBIA

MIEMBROS DE LA COMISIÓN:

GONZALO NAVARRO BADINO
CÉSAR GUERRERO SALDIVIA

SANTIAGO DE CHILE
ABRIL 2012

Resumen

El año 2010 visitó Chile el Dr. Tony Hey, ex decano de la Universidad de Southampton, Reino Unido. Mencionó cómo su ex universidad desarrolló un servicio (Eprints) donde se almacenan las publicaciones de académicos de la universidad y quedan disponibles a la comunidad mundial. Eprints provee búsquedas por varios criterios y así, las publicaciones propias tienen mayor visibilidad y accesibilidad. De hecho, mostró estadísticas que mostraban que estas publicaciones recibían más citas que publicaciones de otras universidades británicas incluso con mayor tradición de investigación.

Antes de este trabajo de memoria, el Departamento de Ciencias de la Computación de la Universidad de Chile – de ahora en adelante DCC - no disponía de una buena herramienta donde los artículos científicos pudiesen ser publicados de manera eficaz. Es por esto que la motivación principal era tener un espacio con el cual se pudiera dar visibilidad a la publicación de *papers* de todo el DCC. Básicamente, un espacio donde se pudiera subir estos documentos a una plataforma web y realizar búsquedas por distintos criterios.

La solución desarrollada resultó ser un sistema web llamado *U-papers*. Las características principales de este sistema lo convierten en un repositorio en donde se almacena información valiosa de publicaciones que permite realizar búsquedas. Actualmente es accesible desde la siguiente dirección: <http://upapers.dcc.uchile.cl>.

Agradecimientos

Primero quiero agradecer a mis padres María Cristina y Mauricio por haberme encaminado hacia la educación que he logrado hasta ahora y por el inmenso apoyo que me entregaron en cada etapa de mi vida.

A mi novia Claudia por haberme apoyado y ayudado en estos años de vida universitaria. Obviamente, le agradezco inmensamente el amor que me entrega cada día y ojalá que cuando lea esto sepa que su amor es completamente retribuido.

A mis hermanos y amigos, quienes son un complemento muy importante en mi vida. Siempre me apoyaron y ayudaron en todo momento.

Al Profesor José Pino le agradezco toda su buena voluntad para ayudarme durante todo el año 2011 a desarrollar un buen trabajo de memoria.

Al Profesor Gonzalo Navarro por haberme entregado un poco de su escaso tiempo en reuniones que me ayudaron a mejorar el sistema.

Índice

1. Introducción	6
2. Antecedentes.....	8
2.1 Tecnologías involucradas.....	8
2.2 Metodología de desarrollo.....	8
2.3 Arquitecturas estándar.....	8
2.4 Descripción de las soluciones existentes	10
3. Especificación del problema	18
3.1 Descripción del problema.....	18
3.2 Relevancia de la solución.....	18
3.3 Requisitos de la solución.....	18
3.3.1 Características de los usuarios.....	19
3.3.2 Requisitos funcionales	19
3.3.3 Requisitos de calidad.....	20
3.3.4 Requisitos de restricción.....	20
4. Descripción de la solución	20
4.1 Contexto de la solución	22
4.2 Arquitectura física.....	26
4.3 Arquitectura Lógica	27
4.4 Modelo de datos.....	30
4.5 Diseño de navegación del sistema.....	32
4.6 Diseño de interfaz de usuario	32
4.6.1 Página inicial.....	32
4.6.2 Vista de la búsqueda avanzada de publicaciones	34
4.6.3 Vista del listado de publicaciones por fecha.....	34
4.6.4 Vista del listado de autores	35
4.6.5 Vista del resultado de la búsqueda avanzada de publicaciones	36
4.6.6 Vista de una publicación.....	36
4.6.7 Vista de publicaciones por autor.....	37
4.6.8 Login.....	38

4.6.9 Vista del Panel de control	39
4.6.10 Vista para agregar, editar y declarar publicaciones.....	41
4.6.11 Vistas de la sección Management.....	45
4.7 Algoritmos desarrollados.....	47
4.7.1 Algoritmo de declaración de publicaciones.....	47
4.7.2 Algoritmo de búsqueda sencilla de publicaciones	48
4.7.3 Algoritmo de búsqueda avanzada de publicaciones.....	48
4.7.4 Algoritmo de parsing de archivos XML.....	49
4.7.5 Algoritmo de proceso de actualización del sistema	49
5. Seguridad del sitio	51
6. Testing	53
7. RespalDOS.....	54
8. Conclusiones.....	55
9. Referencias.....	56

1. Introducción

La publicación de artículos científicos normalmente se hace en revistas de editoriales internacionales tales como Springer Verlag, Elsevier North-Holland y Wiley. Naturalmente, para acceder a dichas revistas, es necesario estar suscrito a ellas, o alternativamente, comprar cada artículo (en línea o impreso). En el caso chileno, Conicyt paga la suscripción en línea de modo que los universitarios pueden acceder sin costo a la mayoría de las revistas científicas. Sin embargo, esto no es común en otras partes del mundo o desde la industria. Por otro lado, tanto para los autores como para sus universidades es importante que sus productos (publicaciones) sean conocidos y utilizados por otros. En consecuencia, desde el punto de vista de los científicos y sus instituciones, existe la necesidad latente de aumentar el acceso a estas publicaciones.

Una acción podría ser poner los artículos en la página web de cada autor. De hecho, muchos científicos hacen esto, y es una práctica aceptada por las principales editoriales. Sin embargo, tal dispersión no es fácil para quien busca encontrar los artículos más relevantes. Se podría pensar entonces en buenas plataformas de almacenamiento y búsqueda donde hubiera espacio para compartir estos documentos, pero a nivel universitario éstas no existen, o no están ampliamente disponibles.

Resulta interesante entonces desarrollar un proyecto que se puede resumir como un repositorio de artículos científicos. Lo anterior se entiende como un espacio donde se pueden compartir y realizar búsquedas de dichos artículos mediante una interfaz web. Lo anterior enmarca en gran medida el uso destinado a este sistema.

Inicialmente, las alternativas disponibles para desarrollar el sistema fueron Zentity¹, Eprints², algún CMS³ y desarrollo mediante un framework. Zentity es una tecnología recientemente desarrollada por Microsoft que sirve como software para repositorios. Las desventajas de esta solución eran: podía resultar un desarrollo complejo de realizar, difícil de utilizar por los usuarios y requería más esfuerzo en desarrollarlo de lo que realmente se necesitaba.

El servicio Eprints está disponible con su código fuente (desarrollado en *perl*). La principal desventaja de esta alternativa era que, muy probablemente, no se iba a lograr un control absoluto de la aplicación, lo que conllevaría orientar principalmente el esfuerzo del desarrollo en tratar de cumplir los requisitos de *U-papers* con una herramienta con código ajeno. El riesgo de que el proyecto se hubiera complicado era muy grande.

El uso de algún CMS se reduce prácticamente en instalar algún módulo que cumpla con la funcionalidad que se quiere. Lamentablemente, no existe un conjunto de módulos que haga exactamente lo que se quería de *U-papers*. Si bien los módulos pueden ser desarrollados por uno mismo, resulta mucho más sensato entonces desarrollar una aplicación completamente propia. Además, perfectamente se puede

¹ <http://research.microsoft.com/en-us/projects/zentity/>

² <http://www.eprints.org/>

³ CMS: Content Management System (Sistema de gestión de contenidos)

desarrollar en un semestre. Debido a las razones anteriores, se optó por usar un framework MVC⁴. Además, Codeigniter fue seleccionado debido a que el memorista estaba familiarizado con el framework y en mayor medida porque cumplía con todo lo solicitado.

U-papers, en términos generales, es una herramienta web que brinda los servicios de almacenamiento y búsqueda de publicaciones científicas. Para lograr lo anterior, no se debe obviar la existencia del sistema FII/EP⁵. Éste último se debe considerar como un sistema legado que servirá como fuente de información para el sistema *U-papers*. Hasta el año 2011, los académicos del DCC debían enviar un email al coordinador de investigación (Prof. Gonzalo Navarro) con un texto en XML (con formato predeterminado) que tuviera toda la información necesaria acerca de sus publicaciones. Este sistema no resultaba óptimo por la cantidad de errores presentes en dichos XML, que eran generados manualmente, además de la absoluta libertad que tenían los autores para declarar *papers*. A partir de lo anterior, se decidió en conjunto con los profesores guía y co-guía, Prof. José Pino y Prof. Gonzalo Navarro (respectivamente), que *U-papers* también proveyera, en sí mismo, un sistema de declaración de publicaciones para los académicos del DCC. Este sistema se sitúa como el punto de partida para la obtención de información relativa a lo que se debe mostrar en el sitio web.

⁴ MVC: Model View Controller (Modelo Vista Controlador)

⁵ Fondos de investigación/Estímulo a la Publicación

2. Antecedentes

2.1 Tecnologías involucradas

Las tecnologías involucradas para el desarrollo de *U-papers* fueron:

1. Sistema:
 - **Apache HTTP Server:** El Proyecto Apache HTTP Server es un esfuerzo colaborativo de desarrollo de software destinado a crear un sistema robusto, de grado comercial, con muchas características y de libre acceso al código fuente de un servidor HTTP (web).
 - **MySQL:** Es el motor de base de datos open source más popular en el mundo, debido a su gran rendimiento, alta fiabilidad, facilidad de uso y ahorro de costos.
 - **PHP:** Lenguaje de programación, del lado del servidor, que se incrusta en el HTML. Proporciona a los desarrolladores web un conjunto completo de herramientas para la construcción de sitios web dinámicos.
2. Software:
 - **CodeIgniter:** Es un potente framework PHP con un tamaño muy pequeño, construido para programadores PHP que necesitan un conjunto de herramientas sencillas y elegantes para crear una aplicación web con todas sus funcionalidades.
 - **jQuery:** Es una biblioteca de JavaScript rápida y concisa que simplifica la manipulación del documento HTML, el manejo de eventos, animación, y las interacciones Ajax para el desarrollo web rápido.

2.2 Metodología de desarrollo

La metodología utilizada en el desarrollo de *U-papers* fue el modelo incremental. Dicho modelo se enfoca en que el sistema no se entrega de una vez, sino que se divide y se entregan incrementos. Así se acordó con el profesor guía: realizar entregas semanales de lo establecido. El modelo indica que con cada incremento se entrega parte de la funcionalidad que se pretende. Los requisitos de un incremento son inamovibles. Sin embargo, estos pueden verse modificados en incrementos posteriores. Este proceso se repite hasta la obtención de un producto completo. Cabe destacar que el modelo incremental se centra en la entrega de un producto operativo en cada incremento.

2.3 Arquitecturas estándar

Para el desarrollo del proyecto se utilizó una arquitectura de tres capas: presentación, negocio y datos. El esquema estándar que tiene esta arquitectura es como la figura siguiente:

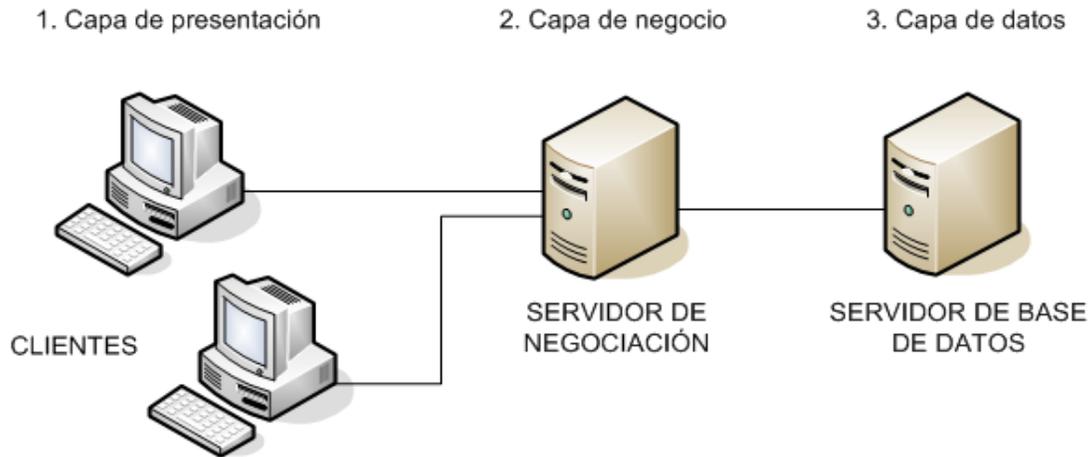


Figura 1: Arquitectura de 3 capas.

U-papers maneja estas tres capas de la siguiente manera:

1. Los clientes (usuarios de *U-papers*) podrán acceder al software mediante un navegador web, puesto que es un sistema hecho con una interfaz en HTML.
2. La capa de negocio es manejada con el código creado para efectuar, por ejemplo, las validaciones, cálculos y flujos de procesos. Este código es ejecutado en un servidor web.
3. La capa de datos se relaciona directamente con una gran base de datos en MySQL.

Si bien la anterior arquitectura es general, existe un patrón de arquitectura de software de 3 capas más específico, ampliamente utilizado en el desarrollo de aplicaciones web: Modelo-Vista-Controlador (MVC). *U-papers* hace uso directo de este patrón gracias al framework utilizado. MVC presenta un esquema muy sencillo, como se muestra a continuación:

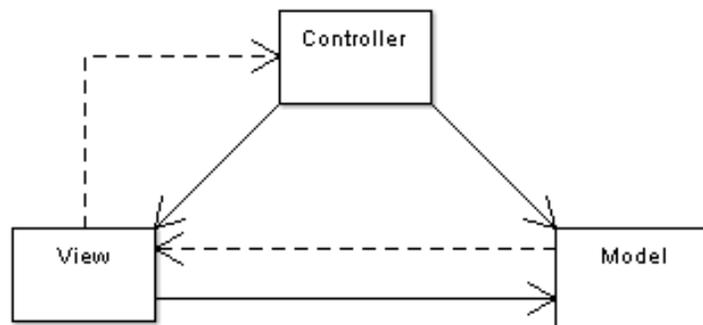


Figura 2: Modelo-Vista-Controlador

MVC sigue un flujo de control, que en general es así:

1. El usuario realiza alguna acción en la interfaz (como pulsar un botón).
2. Llega una notificación a un controlador con la acción realizada por el usuario. El controlador gestiona dichas acciones y accede al modelo para así manejar datos relativos al usuario (y al sistema).

3. El controlador envía los datos procesados a la vista para que ésta se encargue de mostrar una interfaz dinámica al usuario.
4. La interfaz espera nuevas interacciones del usuario para volver a empezar el ciclo.

2.4 Descripción de las soluciones existentes

Existen al menos 2 universidades que tienen plataformas que realizan una tarea similar a lo que se desea de *U-papers*. Estas son la Universidad de Southampton (Southampton, Inglaterra) y la Universidad Pompeu Fabra (Barcelona, España).

La Universidad de Southampton dispone de un servicio llamado Eprints. Este servicio tiene algunas de las funcionalidades que se quiere para *U-papers*. Es un sistema que permite a los científicos publicar sus documentos en una interfaz web y así hacer crecer el repositorio de artículos científicos. Esta es la primera parte del objetivo. La otra es contar con una interfaz donde se pueda realizar búsquedas especializadas de estos documentos. Eprints provee de una intranet donde sólo usuarios registrados pueden ingresar. Desde ahí pueden realizar *uploads* de los artículos científicos. Por otro lado, Eprints presenta varias formas de acceder a estos artículos:

- En la página principal existe un listado de las publicaciones añadidas más recientes, como se muestra a continuación (Figura 3):

The screenshot shows the Eprints website interface. On the left, there is a navigation menu with sections for 'Publications' (Publications Home, This weeks new additions, Search, Browse by research group, Browse by year) and 'Intranet Tools' (Add publications). Below the menu, there is a notice about the next round of REF2013 and RSS feeds for 1.0, 2.0, and Atom. The main content area displays 'Recently added publications' with a list of six items, each with a title, author, and a download icon (PDF, Video, or Zip). On the right, there are two promotional banners: 'APPLY ONLINE FOR A PhD' and 'RESEARCH PROSPECTUS PDF (7MB)'. The top of the page shows 'View: Recently added items | Download Statistics'.

Figura 3: Vista de publicaciones añadidas más recientes en Eprints.

- A la izquierda existe un menú de búsqueda donde se puede filtrar por publicaciones añadidas en la última semana (Figura 4), por grupos de investigación (Figura 5) y por año (Figura 6):

ECS Research	Latest Additions to ECS EPrints Repository
Research Groups	
Researcher Profiles	
Research Themes	
Graduate School	
Seminars	
Research Facilities	
Centres and Institutes	
Latest Technologies	
Publications	
Publications Home	
This weeks new additions	
Search	
Browse by research group	
Browse by year	
Intranet Tools	
Add publications	
	Yesterday
	Boulain, P. (2011) Open Semantic Hyperwikis . PhD thesis, University of Southampton.
	Sunday
	Tinati, R., Halford, S., Carr, L. and Pope, C. (2011) Conceptualising a Web of Linked Actors Poster. In: <i>Web Science 2011</i> , 14th to 18th June, 2011, Koblenz, Germany.
	Yip, M. (2011) An Investigation into Chinese Cybercrime and the Applicability of Social Network Analysis . In: <i>ACM WebSci '11</i> , 14-17 June 2011, Koblenz, Germany.
	Yip, M. and Webber, C. (2011) Hacktivism: a Theoretical and Empirical Exploration of China's Cyber Warriors . In: <i>ACM WebSci '11</i> , 14-17 June 2011, Koblenz, Germany.
	Saturday
	Feng, J., Zhang, R. and Hanzo, L. (2011) Auction-Style Cooperative Medium Access Control . In: <i>2011 IEEE 74th Vehicular Technology Conference</i> .
	Chang, V., Wills, G. and Walters, R. (2011) Towards Business Integration as a Service 2.0 (BaaS 2.0) . In: <i>IEEE International Conference on e-Business Engineering, The 3rd International Workshop on Cloud Services - Platform Accelerating e-Business</i> , 19-21 October, 2011, Beijing, China. (Submitted)
	Tinati, R., Halford, S., Carr, L. and Pope, C. (2011) Conceptualising a Web of Linked Actors . In: <i>ACM WebSci '11</i> , 14-17 June 2011, Koblenz, Germany.

Figura 4: Vista de publicaciones añadidas en la última semana en Eprints.

ECS Research	Browse by Research Groups
Research Groups	Please select a value to browse from the list below.
Researcher Profiles	<ul style="list-style-type: none"> • ECS Groups (15515) <ul style="list-style-type: none"> ◦ BIO@ECS Research Group (87) ◦ Communications Research Group (1806) ◦ Dependable Systems and Software Engineering Research Group (1297) ◦ Electrical Power Engineering (1381) <ul style="list-style-type: none"> ▪ High Voltage Laboratory (561) ◦ Electronic Systems and Devices Group (1099) ◦ Electronics Construction, Maintenance and Logistics (6) ◦ Grid and Pervasive Computing Research Group (9) ◦ High Performance Computing Research Group (396) ◦ Information - Signals, Images, Systems (2081) ◦ Infrared Science and Technology (203) ◦ Intelligence, Agents, Multimedia (3519) ◦ IT Innovation Centre (283) ◦ Learning Societies Lab (598) ◦ Mechanical Workshop (1) ◦ Microelectronics Industrial Unit (2) ◦ Nano Group (1899) ◦ No Specific Group (250) ◦ Optoelectronics Research Centre (869) ◦ Pervasive System Centre (96) ◦ Science and Engineering of Natural Systems (431) ◦ Systems and Networks Group (4) ◦ Systems Group (12) ◦ University of Southampton Institute of Transducer Technology (2) ◦ Web Science Doctoral Training Centre (31)
Graduate School	
Seminars	
Research Facilities	
Centres and Institutes	
Latest Technologies	
Publications	
Publications Home	
This weeks new additions	
Search	
Browse by research group	
Browse by year	
Intranet Tools	
Add publications	
<p>nb. next round of REF2013 will be using data from eprints.ecs, but the central university REF interface.</p>	

Figura 5: Vista de publicaciones agrupadas por grupos de investigación en Eprints.

ECS Research	Browse by Year	
Research Groups	Please select a value to browse from the list below.	
Researcher Profiles	<ul style="list-style-type: none"> • 2012 (3) • 2011 (470) • 2010 (899) • 2009 (974) • 2008 (1094) • 2007 (903) • 2006 (953) • 2005 (955) • 2004 (895) • 2003 (866) • 2002 (849) • 2001 (725) • 2000 (787) • 1999 (750) • 1998 (661) • 1997 (586) • 1996 (536) • 1995 (456) • 1994 (322) • 1993 (314) • 1992 (251) • 1991 (187) • 1990 (174) • 1989 (142) • 1988 (125) • 1987 (106) • 1986 (77) 	<ul style="list-style-type: none"> • 1985 (85) • 1984 (58) • 1983 (49) • 1982 (45) • 1981 (45) • 1980 (38) • 1979 (40) • 1978 (28) • 1977 (37) • 1976 (26) • 1975 (26) • 1974 (25) • 1973 (17) • 1972 (17) • 1971 (7) • 1970 (8) • 1969 (5) • 1968 (3) • 1967 (1) • 1966 (3) • 1965 (3) • 1009 (1) • 1007 (1) • 0018 (1) • 0008 (1) • Not Specified (39)
Graduate School		
Seminars		
Research Facilities		
Centres and Institutes		
Latest Technologies		
Publications		
Publications Home		
This weeks new additions		
Search		
Browse by research group		
Browse by year		
Intranet Tools		
Add publications		
nb. next round of REF2013 will be using data from eprints.ecs, but the central university REF interface.		

Figura 6: Publicaciones agrupadas por año en Eprints.

- En el mismo menú de la izquierda, existe un enlace para realizar dos tipos de búsqueda: sencilla y avanzada. En la búsqueda sencilla (Figura 7) se puede filtrar por: texto, título, resumen, palabras claves, autores, fecha, cantidad de citas y estado de publicación. En la búsqueda avanzada (Figura 8 y 9) se puede filtrar por los mismos parámetros de la búsqueda sencilla, además de algunos otros como grupo de investigación, tipo de monografía, tipo de presentación y número de descargas.

Simple Search

[Click here for an advanced search](#)

Search		Reset the form
Full Text:	Enter a term or terms to search for. all of <input type="text"/>	
Title/Abstract /Keywords:	Enter a term or terms to search for. all of <input type="text"/>	
Creator/Authors Name/Editors Name:	Enter a name or names to search for. Either the family name, or the family name followed by a comma and the first name or initial. Names with spaces may be indicated by surrounding them with double quotes, eg. "van Damme, J". all of <input type="text"/>	
Date:	Enter a date or date range. Examples: "1985-", "2001-05-17-2002-05-16", "-11-1980" <input type="text"/>	
Web of Science Citation Count:	Enter a single number (e.g. 93), or a range of numbers, e.g. `17-23`, `17-` or `-23`. <input type="text"/>	
Full Text Status:	Select one or more values from the list, and whether you want to search for records with any one or all of those values. Default is (Any). <input type="checkbox"/> Public <input type="checkbox"/> Restricted <input type="checkbox"/> None	
Retrieved records must fulfill:	all of these conditions <input type="text"/>	
Order the results:	by year (most recent first) <input type="text"/>	

Figura 7: Búsqueda sencilla en Eprints.

Advanced Search

Don't panic! Just leave the fields you don't want to search blank. [Click here for a simple search.](#)

Search		Reset the form
Full Text:	all of <input type="text"/>	?
Title:	all of <input type="text"/>	?
Creator/Authors Name:	all of <input type="text"/>	?
Abstract:	all of <input type="text"/>	?
Keywords:	all of <input type="text"/>	?
Research Projects:	all of <input type="text"/>	?
Research Group:	<div style="border: 1px solid gray; padding: 5px;"><p>BIO@ECS Research Group Communications Research Group Complexity Doctoral Training Centre Dependable Systems and Software Engineering Research Group Electrical Power Engineering ...High Voltage Laboratory Electronic Systems and Devices Group Electronics Construction, Maintenance and Logistics Grid and Pervasive Computing Research Group High Performance Computing Research Group Information - Signals, Images, Systems Infrared Science and Technology</p></div> Any of these <input type="text"/>	?
Item Type:	<input type="checkbox"/> Article <input type="checkbox"/> Conference or Workshop Item <input type="checkbox"/> Book	<input type="checkbox"/> Book Section <input type="checkbox"/> Technical Report <input type="checkbox"/> Thesis <input type="checkbox"/> Other

Figura 8: Búsqueda avanzada en Eprints (primera parte).

Monograph Type:	<input type="checkbox"/> Technical Report <input type="checkbox"/> Manual <input type="checkbox"/> Project Report <input type="checkbox"/> Working Paper <input type="checkbox"/> Documentation <input type="checkbox"/> Discussion Paper <input type="checkbox"/> Other	?	
Presentation Type:	<input type="checkbox"/> Paper <input type="checkbox"/> Oral Presentation <input type="checkbox"/> Poster <input type="checkbox"/> Other	?	
Conference or Event Title:	all of <input type="text"/>	?	
Editors Name:	all of <input type="text"/>	?	
Status:	<input type="checkbox"/> Unpublished <input type="checkbox"/> Submitted <input type="checkbox"/> In Press <input type="checkbox"/> Published	?	
Refereed:	No Preference <input type="text"/>	?	
Full Text Status:	<input type="checkbox"/> Public <input type="checkbox"/> Restricted <input type="checkbox"/> None	?	
Publication:	all of <input type="text"/>	?	
Date:	<input type="text"/>	?	
Web of Science Citation Count:	<input type="text"/>	?	
Downloads:	<input type="text"/>	?	
Issue type:	all of <input type="text"/>	?	
Additional Information:	all of <input type="text"/>	?	
Format:	<input type="checkbox"/> HTML <input type="checkbox"/> PDF <input type="checkbox"/> Postscript <input type="checkbox"/> Plain Text <input type="checkbox"/> Microsoft PowerPoint <input type="checkbox"/> Image of Cover (to show on abstract page) <input type="checkbox"/> Image collection or slideshow <input type="checkbox"/> Microsoft Word <input type="checkbox"/> Image (JPEG) <input type="checkbox"/> Image (PNG) <input type="checkbox"/> Image (GIF) <input type="checkbox"/> Image (BMP) <input type="checkbox"/> Image (TIFF)	<input type="checkbox"/> Video (MPEG) <input type="checkbox"/> Video (QuickTime) <input type="checkbox"/> Video (AVI) <input type="checkbox"/> Rich Text (RTF) <input type="checkbox"/> Microsoft Excel <input type="checkbox"/> XML <input type="checkbox"/> Archive (BZ2) <input type="checkbox"/> Archive (TGZ) <input type="checkbox"/> Archive (ZIP) <input type="checkbox"/> Audio (WAV) <input type="checkbox"/> Audio (OGG) <input type="checkbox"/> Audio (FLAC) <input type="checkbox"/> Audio (WMA) <input type="checkbox"/> Other	?
Date Deposited:	<input type="text"/>	?	
Retrieved records must fulfill:	all of these conditions <input type="text"/>	?	
Order the results:	by year (most recent first) <input type="text"/>	?	

Figura 9: Búsqueda avanzada en Eprints (segunda parte).

En la Universidad Pompeu Fabra de Barcelona disponen de un repositorio digital denominado e-Repositori. La idea general es la misma: un espacio para compartir artículos científicos y facilitar la búsqueda de los mismos. El sistema e-Repositori también presenta una sección privada para usuarios registrados en donde se pueden realizar *uploads* de las publicaciones. En cuanto a la búsqueda, también cuenta con

una de tipo sencilla y avanzada. La búsqueda sencilla consta sólo de un *input* donde se inserta el texto a buscar (ver Figura 10):

The screenshot shows the homepage of the 'REPOSITORI DIGITAL DE LA UPF' (UPF Digital Repository). The header includes the UPF logo and the text 'UNIVERSITAT POMPEU FABRA' and 'e-Repository'. Below the header, there are navigation links for 'Web BibTIC', 'Català', 'English', 'Ayuda', and 'Mi cuenta'. The main content area is divided into several sections:

- Buscar:** A search bar with a text input field and a 'Ir' button. Below it is a link for 'Búsqueda avanzada'.
- Listar por:** A menu with the following options:
 - **Todo el e-Repository**
 - Comunidades y colecciones
 - Títulos
 - Autores
 - Materias
 - Por fecha de publicación
- Mi cuenta:** A menu with the option:
 - Iniciar
- ¿Qué es el e-Repository?** A link to a page explaining the repository.
- Estadísticas:** A link to a page with statistics.
- Logos:** Logos for 'SHERPA RoMEO' and 'DULCINEA'.

The main content area features a large heading 'Bienvenidos al Repositorio Digital de la UPF' followed by a paragraph: 'El repositorio institucional recoge, difunde y preserva la producción intelectual en formato digital que resulta de la actividad académica e investigadora de la UPF, las revistas científicas y las publicaciones institucionales. Su finalidad es contribuir a aumentar el impacto de la investigación hecha en la UPF y preservar la memoria intelectual de la UPF.'

Below this is another 'Buscar' section with the text 'Introduzca el texto a buscar.' and a search bar with a 'Ir' button.

The 'Comunidades' section is titled 'Elija una comunidad para listar sus colecciones' and lists several categories with their respective counts:

- **Docència [37]**
- **Recerca: articles, congressos, llibres [33]**
- **Recerca: tesis [540]**
- **Recerca: working papers, PFC, etc. [1.380]**
- **Recursos i dades primàries [42]**
- **Revistes científiques [2.771]**
- **Continguts institucionals [84]**
- **Vida universitària [3]**

Figura 10: Búsqueda sencilla en e-Repository.

La búsqueda avanzada, al igual que la sencilla, presenta menos campos que las búsquedas en Eprints. En e-Repository, la búsqueda avanzada (ver Figura 11) utiliza sólo 4 argumentos: ámbito de la búsqueda (todo el repositorio, docencia, tesis, papers, etc.), conjunción (Y, O y NO), tipo de búsqueda (texto completo, resumen, series, autor, etc.) y el texto a buscar. Se puede notar de las figuras anteriores que en e-Repository se pueden listar las publicaciones por comunidades y colecciones, títulos, autores, materias y fecha de publicación. Estas opciones aparecen en el menú izquierdo.

UNIVERSITAT POMPEU FABRA
 REPOSITORIO DIGITAL DE LA UPF
e-Repository

[Web BibTIC](#)
[Català](#) | [English](#)
[Ayuda](#)
[Mi cuenta](#)

Inicio → [Búsqueda avanzada](#)

Buscar

[Búsqueda avanzada](#)

Listar por:

- **Todo el e-Repository**
- [Comunidades y colecciones](#)
- [Títulos](#)
- [Autores](#)
- [Materias](#)
- [Por fecha de publicación](#)

Mi cuenta

- [Iniciar](#)

[¿Qué es el e-Repository?](#)

[Estadísticas](#)

Búsqueda avanzada

Ámbito de la búsqueda:

Limite la búsqueda a una comunidad o colección.

Conjunción	Tipo de búsqueda	Buscar por
	Texto completo ▾	<input type="text"/>
Y ▾	Texto completo ▾	<input type="text"/>
Y ▾	Texto completo ▾	<input type="text"/>

Resultados/página
Ordenar documentos por pertinencia ▾
en orden descendente ▾

Figura 11: Búsqueda avanzada en e-Repository.

3. Especificación del problema

3.1 Descripción del problema

Una práctica recurrente de los investigadores es buscar y estudiar trabajos relacionados a la actual investigación que llevan a cabo. Si bien existe Google, no siempre se encuentra el trabajo deseado. El problema también, es que no todas las publicaciones tienen alguna versión disponible en la web en formato digital (como un archivo PDF). Particularmente, en la página web del DCC se tiene un listado con las publicaciones de cada año, pero sin acceso público a ninguna de ellas. Por otro lado, la Universidad de Chile tiene un repositorio académico global⁶, pero que no cumple con los requisitos impuestos para *U-papers*. Todo lo anterior genera el siguiente problema: la información en la web, relativa a las publicaciones científicas del DCC, es escasa y dispersa. Con los antecedentes anteriores, un investigador de, por ejemplo, Estados Unidos, probablemente nunca encontrará en la web las investigaciones que se desarrollan en Chile ni, en particular, en el DCC.

3.2 Relevancia de la solución

En términos ideológicos, siempre es bueno compartir el conocimiento. Por lo anterior, este repositorio representa el canal ideal para llevar a cabo esa tarea desde la Universidad de Chile hacia el mundo. En primera instancia, será el DCC quien tendrá ese espacio donde compartir las investigaciones científicas que se llevan a cabo.

En términos prácticos, una persona desde cualquier parte del mundo podrá acceder a este repositorio, aumentando la visibilidad que tiene, hasta el momento, la investigación científica en la Universidad de Chile. Podrá también, en particular, acceder a los archivos *pdf* asociados⁷.

U-papers permitirá además que los académicos del DCC puedan declarar sus publicaciones a través de una interfaz dentro del mismo sistema. Se generará un XML que será enviado al coordinador de investigación, eliminando así todos los errores de sintaxis y/o contenido que aparecían al hacerlo de forma manual.

3.3 Requisitos de la solución

Entre las funcionalidades que debe ofrecer *U-papers*, las más significativas son: la posibilidad de ingresar publicaciones científicas al repositorio, sólo por usuarios con los permisos necesarios, y la posibilidad de realizar búsquedas por parte de cualquier tipo de usuario. Otro requisito importante es que la información desplegada en *U-papers* debe ser consistente con la información existente en el sistema FII/EP. Lo anterior quiere decir que *U-papers* debe obtener prácticamente toda la información necesaria para funcionar desde aquel sistema legado. También se definió importante saber el impacto del sitio, lo que se debe mostrar a través de estadísticas de uso del sistema.

⁶ <http://captura.uchile.cl/>

⁷ Por motivos de copyright, estos archivos pdf serán una versión borrador de los archivos originales.

3.3.1 Características de los usuarios

Dentro del sistema se distinguen 4 tipos de usuarios:

1. *Super Administrador*: Encargado de gestionar la plataforma. Puede ingresar y declarar publicaciones, modificarlas, eliminarlas, gestionar usuarios, actualizar el sistema y obtener un respaldo de la base de datos. El proceso de “declarar” una publicación es distinto al de “ingresar”. Cuando se declara una publicación, se genera un XML que se envía al sistema FII/EP y se almacenan temporalmente algunos datos de la publicación (como su PDF en borrador) para luego ser “ingresada” definitivamente al sistema U-papers. Luego de 4 meses (plazo paramétrico), el PDF borrador será visible dentro del sistema.
2. *Administrador*: Encargado de gestionar las publicaciones. Puede ingresar, modificar y eliminar publicaciones, pero no puede gestionar usuarios. Puede actualizar el sistema y obtener un respaldo de la base de datos.
3. *Profesores*: Habilitados para declarar publicaciones, lo que genera un XML para apoyar el sistema actual de publicaciones que posee el DCC y maneja el Profesor Gonzalo Navarro.
4. *Usuario normal*: Habilitado para realizar búsquedas de publicaciones y descargar los archivos asociados.

3.3.2 Requisitos funcionales

A continuación se enumeran los requisitos funcionales de *U-papers*:

1. Super Administradores y Administradores pueden ingresar una nueva publicación, a través de un formulario. Inicialmente, los tipos de publicación son: artículos de journal, artículos de conferencia, libros, capítulos de libros, editores de conferencias y editores de número de journals.
2. Super Administradores y Administradores pueden editar y eliminar publicaciones existentes.
3. Super Administradores y Administradores pueden editar sus datos personales (como nombre, email y contraseña).
4. Super Administradores pueden crear nuevos usuarios Super Administradores y Administradores.
5. Super Administradores y Administradores pueden actualizar el sistema.
6. Super Administradores y Administradores pueden obtener un respaldo de la base de datos.
7. Profesores pueden declarar una publicación, con lo cual se debe generar un XML que se deriva al sistema actual FII/EP. Además, el sistema debe almacenar el PDF borrador subido por los profesores de manera de asociarlo después a la publicación final procesada por el sistema FII/EP.
8. Usuarios normales pueden realizar búsquedas sencillas y avanzadas de publicaciones.
9. El sistema debe mostrar publicaciones más recientes, publicaciones por fecha (mes y año) y por autor.
10. El sistema debe mostrar las publicaciones del año en curso de los autores, al hacer click en sus nombres.

11. El sistema debe entregar estadísticas de visitas (origen, cantidad).
12. El sistema debe tener un script de actualización del sistema que cargue en la base de datos de *U-papers* una serie de XMLs provenientes del sistema FII/EP.

3.3.3 Requisitos de calidad

A continuación se enumeran los requisitos de calidad de *U-papers*:

1. El sistema debe ser llamativo. Es importante para captar la atención de potenciales usuarios.
2. El sistema debe ser intuitivo. Se debe entender su funcionamiento de manera muy rápida.
3. El sistema debe ser seguro. Cada rol de usuario debe poder ver/hacer sólo lo que se le permite. Un usuario normal, por ejemplo, no debe poder subir publicaciones.
4. El sistema debe arrojar resultados relevantes cuando se realice una búsqueda. Esto requiere una buena implementación de los filtros.
5. El sistema debe comportarse de forma robusta en todos los casos, especialmente ante consultas típicas de SQL injection, inserción de código Javascript y en general todo tipo de datos maliciosos.

3.3.4 Requisitos de restricción

A continuación se detallan algunas restricciones que tiene la plataforma:

1. El sistema debe estar en inglés, por defecto.
2. El sistema debe ser consistente con la información existente en el sistema FII/EP.
3. El formato principal para las publicaciones es PDF, pero alternativamente se puede agregar una URL a la publicación y su DOI⁸.
4. El sistema debe tener sólo versiones PDF en borrador de las publicaciones.
5. El sistema no debe permitir dar de alta el PDF borrador de una publicación en un plazo, inicialmente, de al menos 4 meses a partir de su fecha de creación. Este plazo debe ser paramétrico y editable en el tiempo.

4. Descripción de la solución

U-papers es un medio donde se pueden compartir investigaciones. Está orientado sólo para manejar las publicaciones generadas en el DCC, pero debe permitir, eventualmente, su uso dentro de otros departamentos de la Universidad de Chile, agregando nuevo código funcional y modificando el modelo de datos.

Existen 3 intenciones que se pretenden lograr:

⁸ Document Object Identifier

1. Que un usuario cualquiera, inclusive de otro país, pueda acceder a *U-papers* y encontrar la publicación deseada. Así se da a conocer el trabajo que se realiza en el DCC.
2. Que el proceso que realizan los docentes actualmente, relativos a la declaración de nuevas publicaciones, sea un proceso automatizado que comunique el sistema FII/EP con *U-papers*.
3. Mantener la mayor cantidad de PDF's en borrador asociados a las publicaciones.

Una solución natural fue crear un sitio web como *U-papers* que permitiera realizar las tareas previamente mencionadas.

El diseño del sistema se dividió en 3 pasos. El primero fue elegir la mejor opción de desarrollo, dentro de las cuales se consideraron: CMS (como Drupal⁹ y Joomla¹⁰) y un framework (como CakePHP¹¹ y CodeIgniter¹²). Finalmente se decidió optar por la utilización de CodeIgniter, debido principalmente al dominio previo de aquel framework y a que ninguna de las otras opciones tenía una gran ventaja por sobre la otra.

CodeIgniter es un framework orientado a objetos, escrito en PHP. Utiliza el patrón de diseño Modelo-Vista-Controlador y está destinado a desarrollar proyectos de una forma ordenada y más rápida. La reutilización de código es clave en el uso de un framework, lo cual CodeIgniter logra completamente. Si bien existen distintos frameworks que realizan una labor similar, la opción por CodeIgniter tomó mayor peso debido a la excelente documentación y enorme comunidad de desarrolladores existentes.

El segundo paso fue desarrollar el modelo de datos. Se utilizó un modelo de datos relacional utilizando MySQL. La decisión del uso de MySQL en lugar de otros sistemas de gestión de bases de datos se debe tan sólo a la facilidad de uso, previo conocimiento y también a que MySQL ofrecía todo lo que se necesitaba. Existe además el argumento de software de libre acceso versus uno propietario. MySQL es un motor de base de datos *open-source* que ofrece la capacidad de permitir todas las operaciones que se necesitaban para *U-papers*. Una solución empresarial no es una opción, por el simple hecho de no contar con los recursos.

El modelo de datos se realizó gracias a la herramienta MySQL Workbench¹³. Esta herramienta provee una interfaz para el modelamiento de datos, desarrollo en SQL, herramientas de administración para la configuración del servidor, entre otras opciones. Con este software se ahorra tiempo en modelar la base de datos y se gana mucho al obtener un diagrama visual del modelo de datos del sistema.

⁹ <http://www.drupal.org>

¹⁰ <http://www.joomlaspanish.org>

¹¹ <http://www.cakephp.org>

¹² <http://www.codeigniter.com>

¹³ <http://www.mysql.com/products/workbench/>

El tercer paso fue diseñar la estructura de la interfaz, para obtener una navegación sencilla dentro del sitio. También, en paralelo, este paso se enfocó en entregar un buen atractivo visual al sitio.

Luego de realizar el diseño de *U-papers*, se comenzó a trabajar en la implementación del sitio. Este proceso se dividió en 2 partes: implementación misma de *U-papers* e implementación de un script que permitiera la extracción de información desde el sistema FII/EP. La implementación de *U-papers* constaba de entregas sucesivas que eran validadas por el profesor guía de manera semanal. Por otra parte, el script fue parte de un proceso de decisión de cómo se iba a obtener los datos necesarios para poblar la base de datos. El script debía obtener la información necesaria y consecuente con la existente en el sistema FII/EP.

4.1 Contexto de la solución

U-papers tiene como requisito utilizar la información existente en el sistema FII/EP, con el fin de tener información actualizada y correcta. El sistema FII/EP lo maneja actualmente el profesor Gonzalo Navarro, quien procesa cada declaración de las publicaciones que envían los investigadores del DCC en formato XML. Luego, una vez al mes, actualiza una página web muy sencilla que muestra un listado de las publicaciones en distintos formatos: por año, por tipo y por autor.

Con el fin de utilizar aquella información y alimentar el sistema *U-papers*, se creó un script en PHP que procesaba dicha página web del sistema FII/EP. El script leía el código fuente de dicha página e iba insertando información nueva dentro de la base de datos de *U-papers* mediante un proceso de *parsing*. Se utilizó en un principio para poder alimentar de información real a *U-papers* y así poder ejecutar búsquedas de publicaciones. Sin embargo, esta solución no era la ideal debido a que en aquella página web (con los listados de las publicaciones), no se podía obtener toda la información necesaria para que la información en la base de datos de *U-papers* fuera consistente con la información del sistema FII/EP. Por ejemplo, desde aquella página web no se podía obtener los *keys* de los journals ni de las conferencias, por lo que se debía utilizar *keys* creados arbitrariamente. Esto genera obviamente una información errónea para el sistema FII/EP al momento de generar los XMLs. Afortunadamente, este problema tiene una solución previa, para lo cual sirve ver el siguiente diagrama que muestra el contexto en que va a convivir *U-papers*:

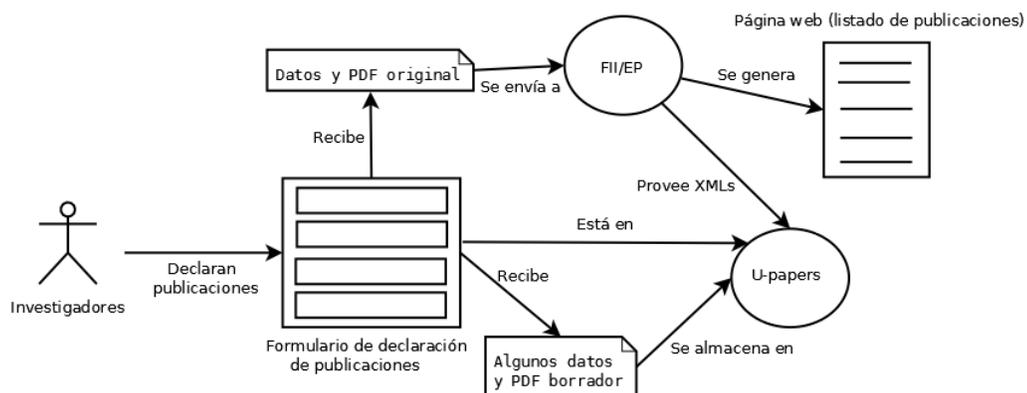


Figura 12: Diagrama del contexto de *U-papers*.

En el diagrama anterior se puede apreciar que el ciclo comienza cuando los investigadores del DCC declaran sus publicaciones dentro de un formulario provisto en *U-papers*. El sistema procesa este formulario y, si no hay errores en él, se envía un email al coordinador de investigación¹⁴ con un XML (en texto plano) junto con el PDF final (que es el PDF original de la publicación) como archivo adjunto, si fuera el caso que agregan dicho PDF en el formulario. Paralelamente, se guardan algunos datos dentro de la base de datos de *U-papers*, en una tabla de publicaciones pasivas. Estos datos son: el título de la publicación, el año y el PDF borrador (si es que se agrega).

Gracias al XML generado y enviado al correo del coordinador de investigación, el Prof. Navarro tiene por sí mismo unos scripts que procesan estos XML. Los scripts también generan los archivos HTML que dan vida a la página web anteriormente mencionada que contiene el listado de las publicaciones del DCC. Además, una vez al mes el Prof. Navarro sube a un FTP una serie de archivos XML que luego son procesados por un script de *U-papers* que actualiza toda la información necesaria para el correcto funcionamiento del sistema. Estos archivos XML son:

1. Publ<year>.xml: Tienen todas las declaraciones de las publicaciones desde el año 2009 en adelante. Un ejemplo de una declaración de una publicación en este archivo es:

```

<journal>
  <name>MTAP</name>
  <title>Improving the Space Cost of k-NN Search in Metric Spaces
    by Using Distance Estimators</title>
  <authors>
    <author>
      <dccname>BB</dccname>
      <affiliation>unica</affiliation>
    </author>
    <author>
      <dccname>GN</dccname>
      <affiliation>unica</affiliation>
    </author>
  </authors>
  <volume>41</volume>
  <number>2</number>
  <pages>215-233</pages>
  <month>1</month>
  <year>2009</year>
  <mesacreditado>1</mesacreditado>
  <key>MTAP.41.2.215</key>
  <ack>milenio</ack> <ack>coopint</ack> <ack>fondecyt</ack>
  <url>http://www.springerlink.com/content/056r2076765441p0/?p=710a362bd89745c0beb329c0bcc339de&pi=2</url>
  <abstract>
    Similarity searching in metric spaces has a vast number of applications in several fields like multimedia databases, text
    retrieval, computational biology, and pattern recognition. In this context, one of the most important similarity queries is the k$
    nearest neighbor (k$-NN) search. The standard best-first k$-NN algorithm uses the
    lower bound distance to prune objects during the search. Although optimal in several aspects, the disadvantage of this method
    is that its space requirements for the priority queue that stores unprocessed clusters can be linear in the database size. Most
    of the optimizations used in spatial access methods (e.g., pruning using MinMaxDist) cannot be applied in metric spaces, due
    to the lack of geometric properties. We propose a new k$-NN algorithm that uses \emph{distance estimators}, aiming to
    reduce the storage requirements of the search algorithm. Experimental results with synthetic and real datasets confirm the
    savings in storage space of our proposed algorithm.
  </abstract>
</journal>

```

¹⁴ El correo utilizado es: coordinador-investigacion@dcc.uchile.cl. Actualmente este email redirecciona los correos entrantes al correo del Prof. Gonzalo Navarro.

2. Académicos.xml: Tiene la información de los académicos investigadores (jornada completa) del DCC. Un ejemplo de la declaración de un académico en este archivo es:

```
<ajc>
  <key>RB</key>
  <firstname>Ricardo</firstname> <lastname>Baeza-Yates</lastname>
  <email>rbaeza@dcc.uchile.cl</email>
</ajc>
```

3. Publishers.xml: Tiene la información de las editoriales asociadas a las publicaciones. Un ejemplo de declaración de una editorial en este archivo es:

```
<publisher>
  <key>USACM</key>
  <name>U.S. Association for Computational Mechanics</name>
  <prestigio>no</prestigio>
</publisher>
```

4. Journals.xml: Tiene la información de los journals asociados a las publicaciones del tipo “artículos de journal”. Un ejemplo de declaración de un journal en este archivo es:

```
<defjournal>
  <key>IRes</key>
  <name>Information Research</name>
  <publisher>Sheffield</publisher>
  <issn>1368-1613</issn>
  <type>A</type>
  <impactos>
    <impacto>
      <ano>2007</ano>
      <if>1.027</if>
      <aif>0.649</aif>
    </impacto>
  </impactos>
</defjournal>
```

5. Confs.xml: Tiene la información de las conferencias asociadas a las publicaciones del tipo “artículos de conferencia”. Un ejemplo de declaración de una conferencia en este archivo es:

```
<defconf>
  <key>DSAL</key>
  <name>Workshop on Domain-Specific Aspect Languages</name>
  <abbrev>DSAL</abbrev>
  <editions>
    <edition>
      <year>2009</year>
      <number>4th</number>
      <venue>Charlottesville, Virginia, USA</venue>
      <editor>
        Thomas Cleenewerck, Johan Fabry, Anne-Fran\,coise Le Meur,
        Jacques Noy\e, Eric Tanter
      </editor>
      <isbn>978-1-60558-455-3</isbn>
      <from>
        <day>3</day>
        <month>3</month>
        <year>2009</year>
      </from>
      <to>
```

```

        <day>3</day>
        <month>3</month>
        <year>2009</year>
    </to>
    <acceptance>83</acceptance>
</edition>
<edition>
    <year>2011</year> (oblig)
    <number>6th</number>
    <venue>Porto do Galinhas, Brazil</venue>
    <editor></editor>
    <isbn></isbn>
    <volume></volume>
    <from>
        <day>22</day>
        <month>3</month>
        <year>2011</year>
    </from>
    <to>
        <day>22</day>
        <month>3</month>
        <year>2011</year>
    </to>
</edition>
</editions>
<publisher>ACM</publisher>
<type>L1</type>
<update> <year>2011</year> </update>
<type>L2</type>
</defconf>

```

6. Alumno.xml: Tiene la información de los tipos de alumno que participan en las investigaciones con los autores DCC. Un ejemplo de declaración de un tipo de alumno en este archivo es:

```

<tipoalumno>
    <key>phdstudent</key>
    <texto>Estudiante de doctorado de un autor DCC</texto>
</tipoalumno>

```

7. Afiliaciones.xml: Tiene la información relativa a los tipos de afiliaciones que tienen los autores DCC en cada publicación. Un ejemplo de declaración de un tipo de afiliación en este archivo es:

```

<afiliacion>
    <tipo>unica</tipo>
    <texto>Única (sólo DCC)</texto>
</afiliacion>

```

8. Acks.xml: Tiene la información de los tipos de proyectos que se asocian a las publicaciones.

```

<ack>
    <tipo>fondecyt</tipo>
    <texto>Proyecto Fondecyt</texto>
</ack>

```

Como se mencionó anteriormente, se creó un script que procesa cada uno de estos archivos. Este proceso consiste en un *parsing* de los datos existentes en los XML que son luego cargados en la base de datos de *U-papers* y así aprovechar todos los beneficios del modelo de datos relacional creado.

Para la facilidad del uso del script, se creó una sección dentro de *U-papers* (llamada *Update system*) en donde sólo apretando un botón se realiza todo el proceso de obtención de datos. La idea es que un Super Administrador (inicialmente el Prof. Gonzalo Navarro) ejecute esta acción al menos una vez al mes, al mismo tiempo en que se actualiza la información del sistema FII/EP.

Hasta ahora se ha explicado cómo es el proceso que se decidió realizar, pero no el porqué. Por un lado el DCC necesitaba una herramienta para poder dar visibilidad a sus publicaciones y por otro lado necesitaba una herramienta para apoyar el proceso de declaración de las mismas. La primera casi no existía y la segunda se desarrollaba de forma manual mediante el envío por correo de un XML muy probablemente con falta de información y en ocasiones, errónea. Como para desarrollar la primera herramienta se necesita de una fuente de información que ya existe (los XMLs del Prof. Navarro), resultaba natural entonces desarrollar una herramienta de apoyo al sistema de declaración de publicaciones para que la información enviada fuera completa y correcta. Sin embargo, no es posible utilizar directamente la información que ingresan los académicos en los formularios de *U-papers*. Esto se debe a que esa información necesita igualmente de una revisión “humana” de los datos provistos. Un ejemplo de lo anterior es cuando un académico decide definir un nuevo artículo de journal con un tipo A, cuando el coordinador de investigación piensa que no califica como tal. Es decir, existe un proceso intermedio de revisión de los datos en donde el Prof. Navarro realiza observaciones, correcciones y finalmente la generación de un nuevo XML con datos finales.

Resumiendo todo lo explicado anteriormente, el proceso es el siguiente:

1. Los académicos ingresan a *U-papers* e ingresan los datos de una nueva publicación dentro un formulario.
2. *U-papers* procesa esos datos y envía un XML al email coordinador de investigación.
3. El coordinador de investigación revisa este XML y decide si está correcto o no. Realiza modificaciones en caso de ser necesario. Si está correcto, actualiza sus archivos XML explicados anteriormente (journals.xml, confs.xml, etc.).
4. Cuando se estime conveniente, el coordinador de investigación ingresa a *U-papers* y ejecuta el proceso de actualización del sistema.
5. *U-papers* en este punto tiene toda su información actualizada y vuelve a partir el ciclo.

4.2 Arquitectura física

En el sistema se emplea una arquitectura de 3 capas. La primera capa es la interfaz de *U-papers*. Los usuarios deben utilizar un navegador y una conexión a internet para poder visualizar el sitio. El sistema no fue implementado para ser visualizado de otra forma que no sea en un navegador (como Firefox, Chrome, Opera o Internet Explorer).

La segunda capa se relaciona con el acceso a los datos y la lógica de negocios. Como el sitio web se implementa usando CodeIgniter, se requiere de un servidor que permita ejecutar código en PHP. Para *U-papers* se usó Apache 2.2.8 y PHP 5.3.1.

La tercera capa se relaciona con la base de datos, que está ubicada localmente. Para *U-papers* se usó MySQL 5.1. A continuación, se muestra un diagrama de lo explicado:

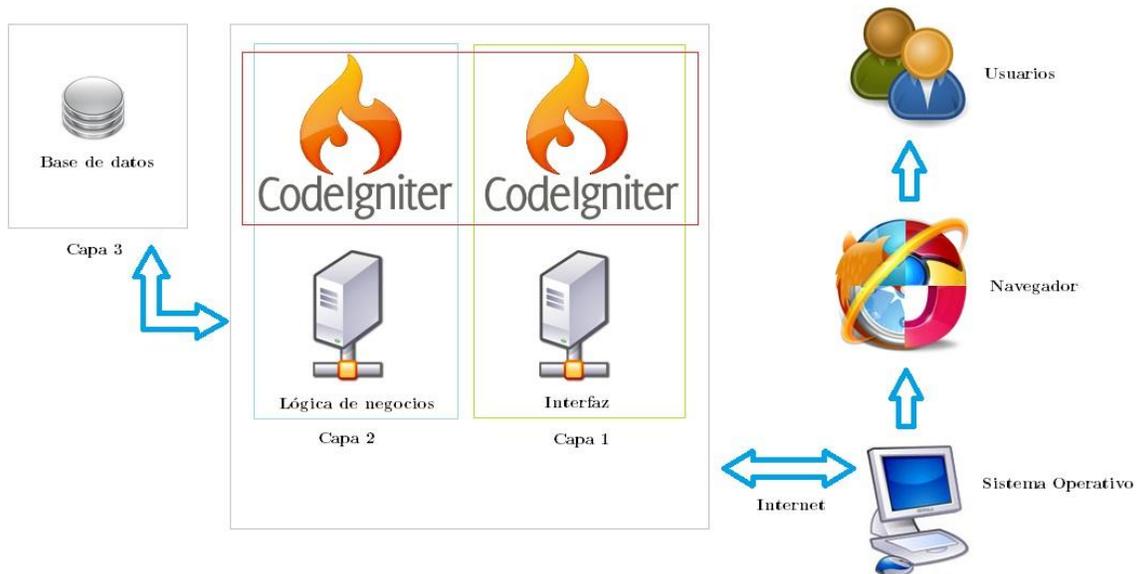


Figura 13: Arquitectura de 3 capas de *U-papers*.

4.3 Arquitectura Lógica

El sistema se creó para que un usuario cualquiera pudiera realizar una búsqueda de algún *paper* en particular. Todas las operaciones que se realizan dentro del sitio web y los cambios efectuados por los usuarios correspondientes, deben plasmarse en la base de datos de una forma consistente. A continuación se muestra el diagrama de arquitectura lógica de *U-papers*:

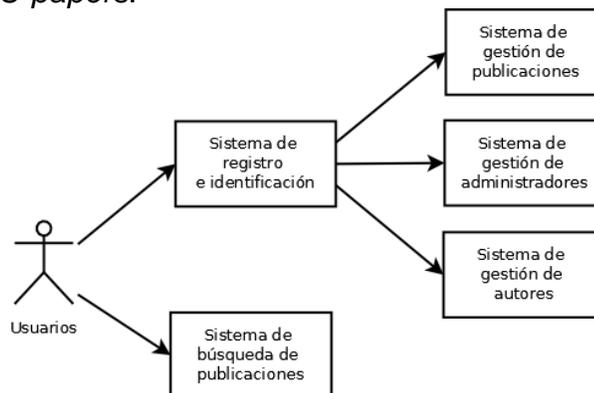


Figura 14: Diagrama de arquitectura lógica de *U-papers*.

Cada acción desencadenada en *U-papers*, es procesada por alguno de estos sistemas. Cada acción está sujeta a los permisos que tiene cada usuario. Dichos permisos son administrados por el sistema de identificación, el cual una vez que el

usuario haya ingresado su nombre de usuario y contraseña en el sistema, sabrá qué tipo de rol tiene y, por ende, qué acciones puede realizar.

Las acciones anteriores pueden ser representadas mediante los siguientes diagramas:

- Super Administrador:

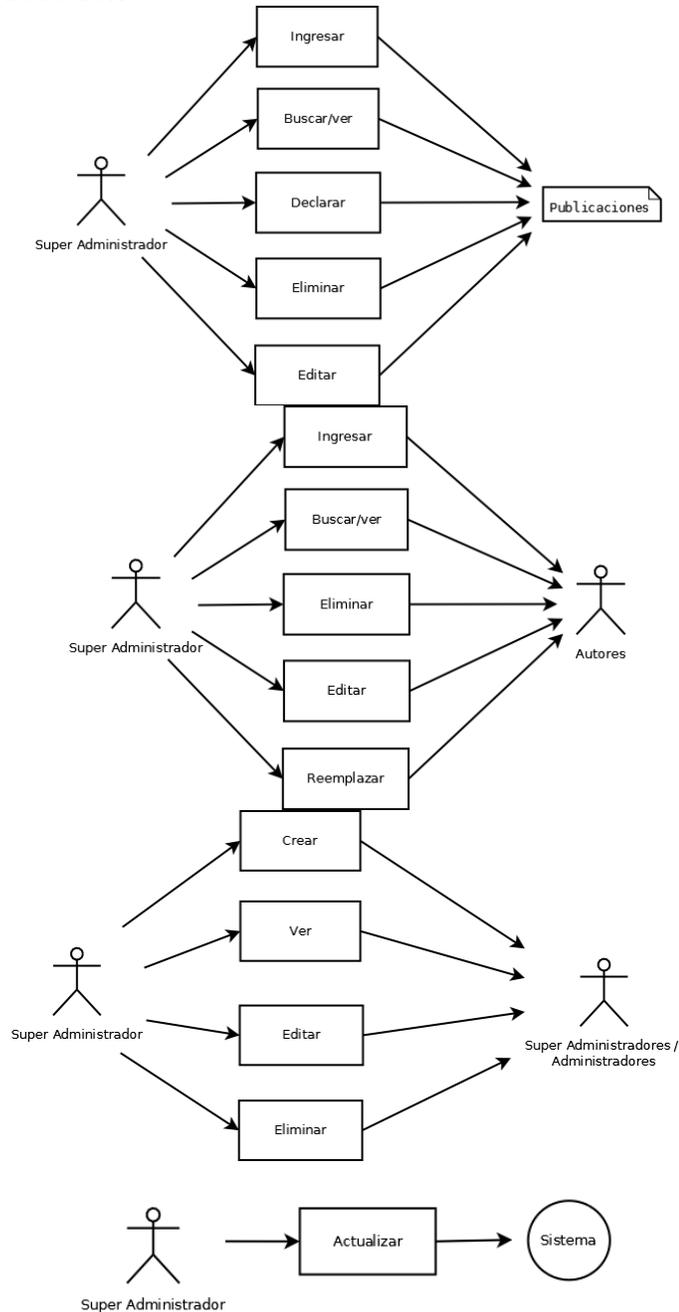


Figura 15: Acciones permitidas al usuario Super Administrador.

- Administrador:

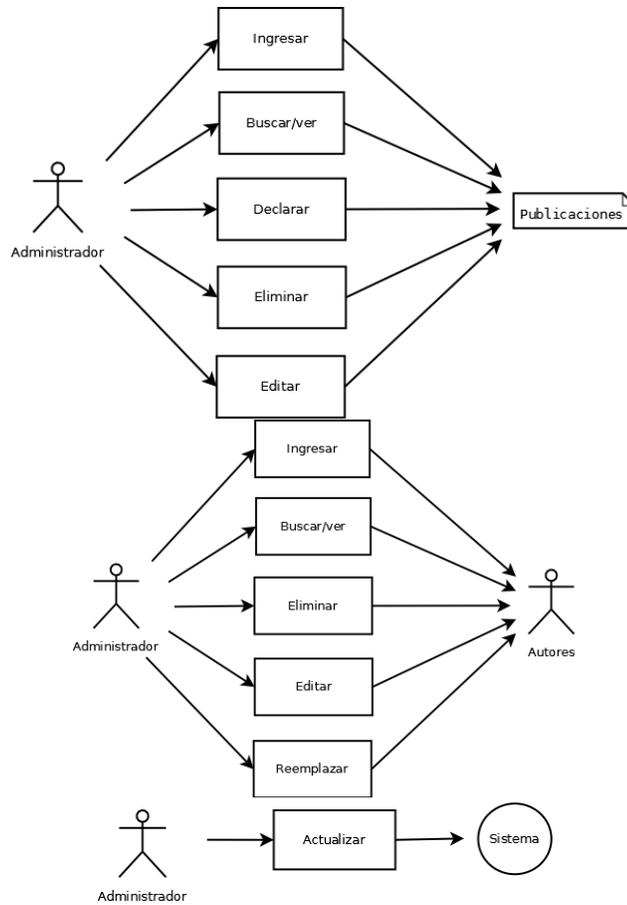


Figura 16: Acciones permitidas al usuario Administrador.

- Profesor:

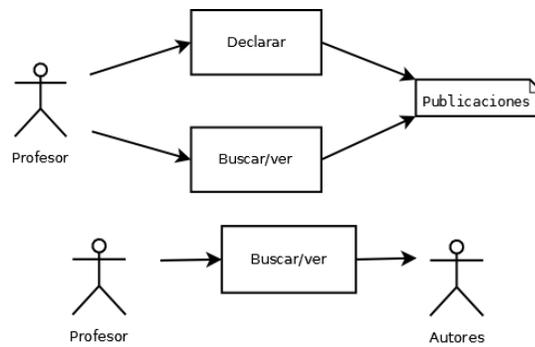


Figura 17: Acciones permitidas a los profesores

- Usuario normal:

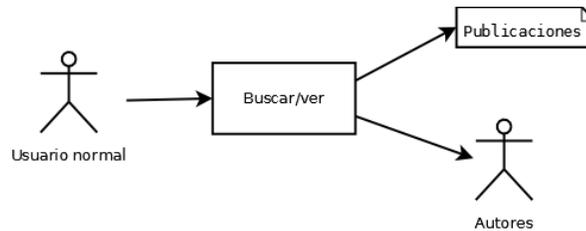


Figura 18: Acciones permitidas a un usuario normal

Claramente el sitio cuenta con otras secciones que son abiertas a cualquier tipo de usuario (FAQ, Disclaimer y Stats).

4.4 Modelo de datos

La base de datos almacena toda la información necesaria referente a, por ejemplo, publicaciones, autores y usuarios. Como se mencionó anteriormente, es una base de datos local, cuyo diseño fue realizado utilizando la herramienta MySQL Workbench. En la figura 19 se muestra el diagrama del modelo de datos relacional desarrollado con Workbench.

Básicamente, existe una tabla llamada “publicaciones” que representa a todos los *papers* que se almacenan. Esta tabla tiene la información común de las publicaciones. Lo anterior se debe a que cada publicación puede, a su vez, pertenecer a uno de los 6 tipos siguientes: artículos de journal, artículos de conferencia, libros, capítulos de libro, editores de conferencias o editores de número de journal. Cada uno de estos 6 tipos anteriores es representado por una tabla. En el caso de los artículos de journal y los artículos de conferencia, ambos se relacionan con las tablas “journals” y “conferences”, respectivamente. Esta asociación es mediante una llave foránea llamada “name”. En la realidad sucede que, por ejemplo, los artículos de journal son un extracto de un journal. Es por aquello que se requiere dicha asociación.

También se mantiene una asociación entre las publicaciones y los autores, de manera de evitar mantener tuplas repetidas con autores ya existentes. También sirve como eventual filtro de publicaciones por autores. Cada autor, a su vez, puede pertenecer a un grupo. Por ejemplo, los autores DCC pertenecen al grupo “dcc_author” referenciado en la tabla “authors_types”.

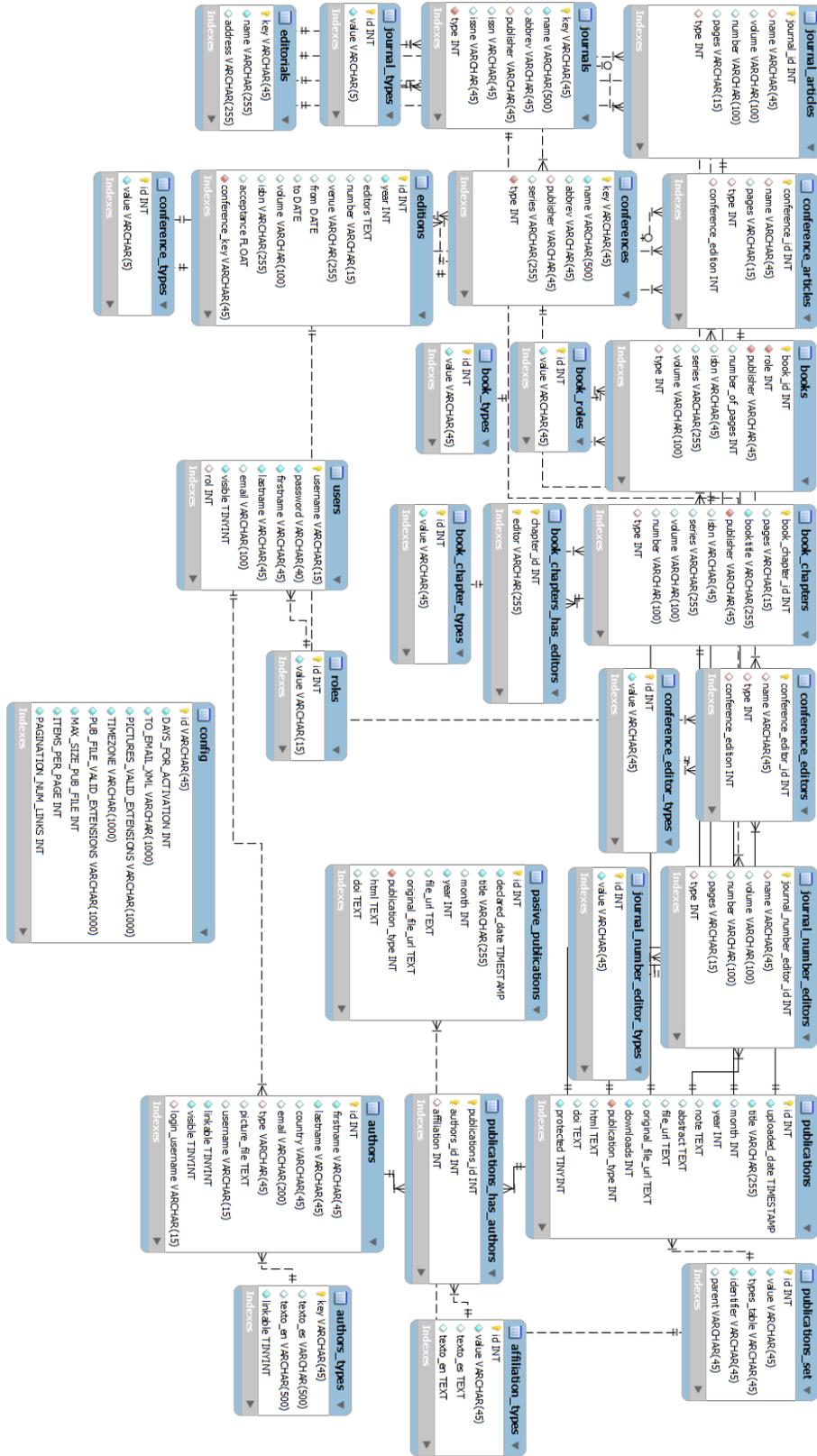


Figura 19: Modelo de datos relacional

4.5 Diseño de navegación del sistema

El sitio web contiene una serie de acciones que los usuarios pueden realizar. Para representar de mejor forma lo anterior, se muestra a continuación un diagrama de navegación de *U-papers*:

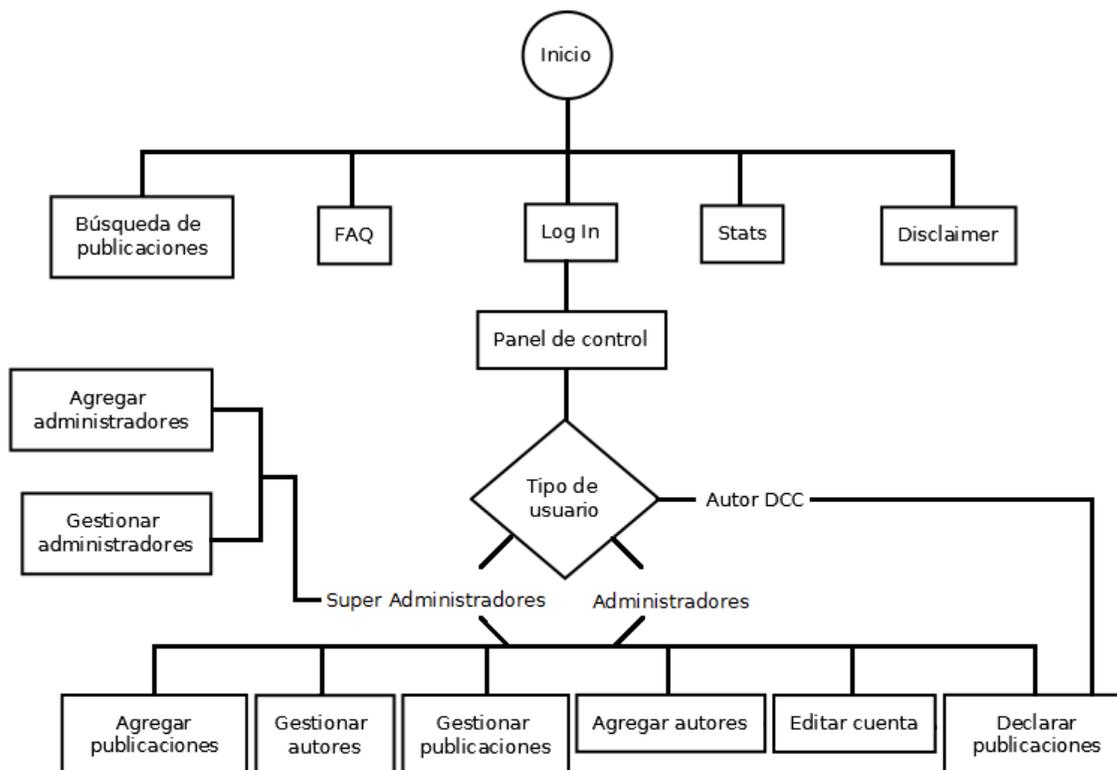


Figura 20: Diagrama de navegación.

Las secciones de Agregar y Declarar publicaciones tienen 6 posibilidades cada una:

1. Agregar/Declarar artículo de journal.
2. Agregar/Declarar artículo de conferencia.
3. Agregar/Declarar libro.
4. Agregar/Declarar capítulo de libro.
5. Agregar/Declarar editor de conferencia.
6. Agregar/Declarar editor de número de journal.

4.6 Diseño de interfaz de usuario

La interfaz gráfica se desarrolló en conjunto con el profesor guía, tratando siempre de mantener una presentación intuitiva y atractiva.

4.6.1 Página inicial

Todas las vistas de *U-papers* constan de un *header*, un menú lateral y el contenido. El *header* muestra un logo representativo de *U-papers*, un enlace a la página

web de la Universidad de Chile¹⁵, un enlace a la página web del DCC¹⁶, un enlace a Stats, un enlace a FAQ¹⁷, un enlace a Disclaimer, una opción para traducir la página y un enlace para iniciar sesión. El menú lateral presenta 2 secciones: una sección relativa a distintas formas de acceder a las publicaciones y otra privada relativa a acciones que pueden ejecutar usuarios registrados con roles definidos.

La página principal muestra, en la parte de contenidos, una explicación breve del alcance de *U-papers*, luego un buscador sencillo y finalmente muestra las publicaciones más recientes.

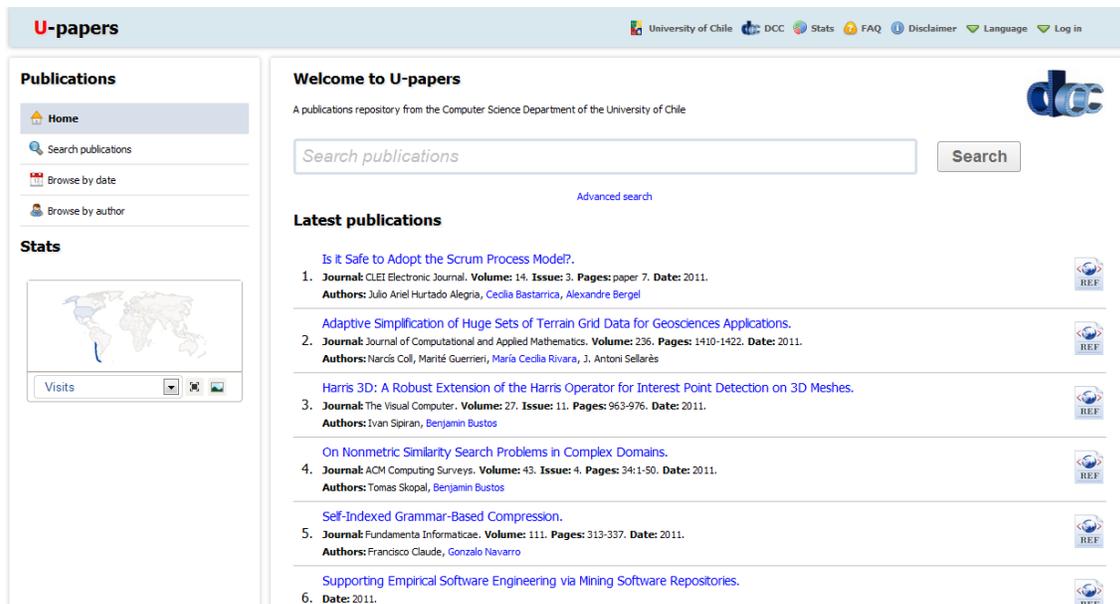


Figura 21: Página inicial

Se acordó con el Profesor Guía que la búsqueda sencilla genere resultados a partir de *matches* entre las palabras que se ingresan y los títulos, *keywords* y *abstracts* que tiene cada publicación. Los resultados son luego mostrados según relevancia, es decir, de mayor a menor cantidad de *matches*. Se acordó hacerlo también de esta forma para dedicar el mayor esfuerzo a implementar un algoritmo de búsqueda avanzada que otorgue resultados más precisos.

Las publicaciones más recientes mostradas en el inicio (*Latest publications*) son efectivamente las más recientes, pero con la restricción siguiente: deben tener un PDF borrador, una página de referencia hacia la editorial, la publicación en versión HTML o DOI asociado. Si por ejemplo la publicación más reciente subida a *U-papers* no tiene asociada ninguna de estas 4 opciones, no será mostrada en la página de inicio, para así darle prioridad y mayor visibilidad a las publicaciones que sí tienen alguna de ellas.

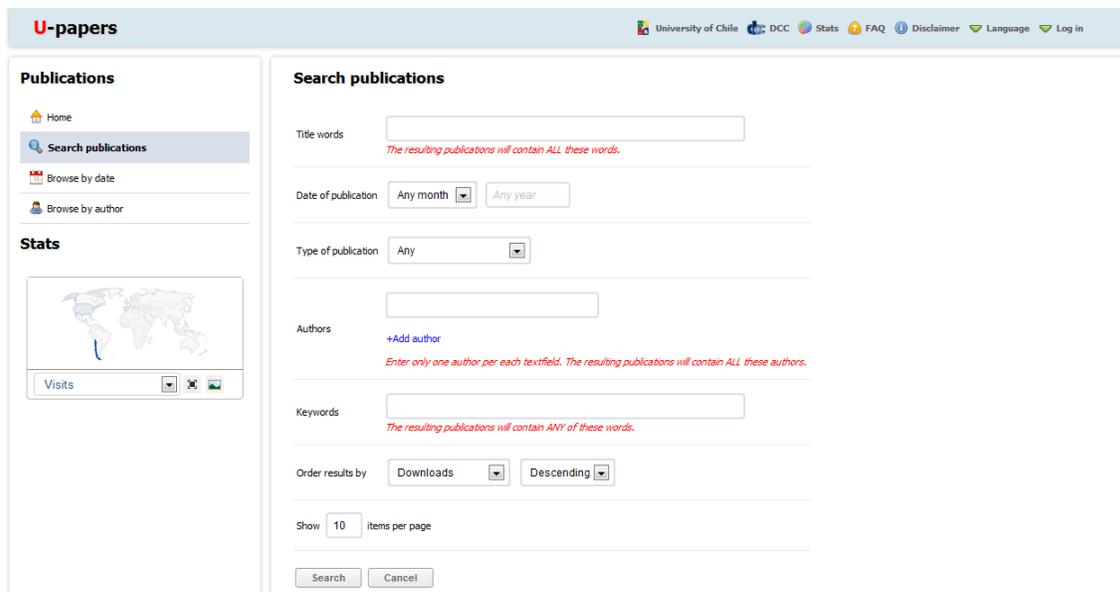
¹⁵ <http://www.uchile.cl>

¹⁶ <http://www.dcc.uchile.cl>

¹⁷ Frequently asked questions

4.6.2 Vista de la búsqueda avanzada de publicaciones

En esta vista se muestra un formulario donde se pueden ingresar criterios de búsqueda para filtrar de mejor forma los posibles resultados.



The screenshot displays the 'U-papers' website interface. At the top, there is a navigation bar with the 'U-papers' logo on the left and links for 'University of Chile', 'DCC', 'Stats', 'FAQ', 'Disclaimer', 'Language', and 'Log in' on the right. The main content area is divided into two columns. The left column contains a 'Publications' sidebar with links for 'Home', 'Search publications', 'Browse by date', and 'Browse by author', and a 'Stats' section with a world map and a 'Visits' dropdown menu. The right column is titled 'Search publications' and contains a search form with the following fields: 'Title words' (with a note: 'The resulting publications will contain ALL these words.'), 'Date of publication' (with 'Any month' and 'Any year' dropdowns), 'Type of publication' (with an 'Any' dropdown), 'Authors' (with a '+Add author' link and a note: 'Enter only one author per each textfield. The resulting publications will contain ALL these authors.'), 'Keywords' (with a note: 'The resulting publications will contain ANY of these words.'), 'Order results by' (with 'Downloads' and 'Descending' dropdowns), and 'Show 10 items per page'. At the bottom of the search form are 'Search' and 'Cancel' buttons.

Figura 22: Vista de la búsqueda avanzada de publicaciones

Este conjunto de filtros logra 2 enfoques fundamentales: mantener la sencillez del sitio y entregar los resultados más relevantes en una búsqueda. El primer enfoque tiene relación con que en la mayoría de los sitios que proveen de algún tipo de búsqueda avanzada llenan la interfaz con filtros completamente innecesarios que prácticamente nunca son utilizados. El segundo enfoque tiene relación con la precisión de los resultados entregados por la búsqueda. Se pensó que este conjunto de filtros representaba lo necesario para poder obtener los resultados adecuados.

4.6.3 Vista del listado de publicaciones por fecha

En esta vista se presenta un filtro de publicaciones por mes y año.

U-papers University of Chile DCC Stats FAQ Disclaimer Language Log in

Publications

Home
Search publications
Browse by date
Browse by author

Stats

Visits

Browse by date

All 2011 Change (if a year does not appear in the list, it is because there are no records for that particular year)

All 2011 (121)

[Mobile Audio Assistance in Bus Transportation for the Blind.](#)
Journal: International Journal on Disability and Human Development. **Volume:** 10. **Issue:** 4. **Pages:** 365-371. **Date:** November 2011.
Authors: Jaime Sánchez, Matías Espinoza

[Is it Safe to Adopt the Scrum Process Model?.](#)
Journal: CLEI Electronic Journal. **Volume:** 14. **Issue:** 3. **Pages:** paper 7. **Date:** 2011.
Authors: Julio Ariel Hurtado Alegria, Cecilia Bastarrica, Alexandre Bergel

[Adaptive Simplification of Huge Sets of Terrain Grid Data for Geosciences Applications.](#)
Journal: Journal of Computational and Applied Mathematics. **Volume:** 236. **Pages:** 1410-1422. **Date:** 2011.
Authors: Narcís Coll, Marité Guerrieri, María Cecilia Rivara, J. Antoni Sellarés

[Harris 3D: A Robust Extension of the Harris Operator for Interest Point Detection on 3D Meshes.](#)
Journal: The Visual Computer. **Volume:** 27. **Issue:** 11. **Pages:** 963-976. **Date:** 2011.
Authors: Ivan Sipiran, Benjamin Bustos

[On Nonmetric Similarity Search Problems in Complex Domains.](#)
Journal: ACM Computing Surveys. **Volume:** 43. **Issue:** 4. **Pages:** 34:1-50. **Date:** 2011.
Authors: Tomas Skopal, Benjamin Bustos

[A Communication Infrastructure to Ease the Development of Mobile Collaborative Applications.](#)
Journal: Journal of Network and Computer Applications. **Volume:** 34. **Issue:** 6. **Pages:** 1883-1893. **Date:** November 2011.
Authors: Juan Rodríguez-Covill, Sergio Ochoa, José A. Pino, Roc Messeguer, Esunly Medina, Dolores Royo

[Ejemovi!, a Web-Based Tool to Create Mobile Learning Videogames.](#)

Figura 23: Vista del listado de publicaciones por fecha

Si bien se pueden desplegar las publicaciones por mes y año, también es posible desplegarlas únicamente por año, eligiendo la opción “All” en el listado de meses.

Esta vista resulta importante desde el punto de vista estadístico. Acá se puede saber la cantidad de publicaciones que se están generando, por ejemplo, anualmente.

4.6.4 Vista del listado de autores

En esta vista se muestra un listado de todos los autores DCC existentes en el repositorio y un campo de texto para poder filtrar por nombre.

U-papers University of Chile DCC Stats FAQ Disclaimer Language Log in

Publications

Home
Search publications
Browse by date
Browse by author

Stats

Visits

Browse by author

Type part of the author's name

22 result(s)

Alejandro Hevia
Alexandre Bergel
Bárbara Poblete
Benjamin Bustos
Cecilia Bastarrica
Claudio Gutierrez

Éric Tanter
Gonzalo Navarro
Jaime Sánchez
Jérémy Barbay
Johan Fabry
Jorge Pérez

Figura 24: Vista del listado de autores

La finalidad de esta vista es poder buscar las publicaciones de los autores (jornada completa) del DCC. Si bien se podía haber hecho un filtro por todos los autores existentes en el sistema *U-papers*, se tomó la decisión de filtrar sólo por los autores jornada completa del DCC y así dar énfasis al trabajo desarrollado dentro del departamento.

4.6.5 Vista del resultado de la búsqueda avanzada de publicaciones

En esta vista se muestran los resultados de una búsqueda. Se aplica para ambos tipos de búsqueda: sencilla y avanzada.

The screenshot shows the 'U-papers' website interface. At the top, there is a navigation bar with the 'U-papers' logo and links for University of Chile, DCC, Stats, FAQ, Disclaimer, Language, and Log in. The main content area is divided into two columns. The left column contains a 'Publications' sidebar with options for Home, Search publications, Browse by date, and Browse by author, along with a 'Stats' section featuring a world map and a 'Visits' counter. The right column is titled 'Search publications' and shows 'Your search returned 34 results'. Below this, a list of 10 search results is displayed, each with a title, journal information (Journal, Volume, Issue, Pages, Date), and author names. Some results include a small icon representing a journal or a reference symbol.

Figura 25: Vista del resultado de la búsqueda avanzada de publicaciones

A diferencia de la página de inicio, en esta vista se muestran todas las publicaciones que presenten algún *match* con el criterio de búsqueda, pero sin la restricción que tienen las publicaciones presentes en la página de inicio, es decir, que contengan de alguna forma un enlace a la publicación.

4.6.6 Vista de una publicación

En esta vista se muestran los datos asociados a una publicación. Esta vista se despliega cuando se hace click en el nombre de alguna publicación.

U-papers University of Chile DCC Stats FAQ Disclaimer Language Log in

Publications

Home

Search publications

Browse by date

Browse by author

Stats



Visits

View publication

Title	Evaluating Collaborative Learning Processes using System-based Measurement
Authors	Cesar A. Collazos, Crescencio Bravo, Jane Kobas, José A. Pino, Luis A. Guerrero, Manuel Ortega, Miguel A. Redondo, Stefano Renzi
Publication date	July 2007
Abstract	Much of the research on collaborative work focuses on the quality of the group outcome as a measure of success. There is less research on the collaboration process itself, but an understanding of the process should help to improve both the process and the outcomes of collaboration. Understanding and analyzing collaborative learning processes requires a fine-grained analysis of group interaction in the context of learning goals. Taking into account the relationships among tasks, products and collaboration this paper presents a set of measures designed to evaluate the collaborative learning process. We emphasise: direct system-based measures based on data produced by a collaborative learning system during the collaboration process, and suggest that these measures can be enhanced by also considering participants' perceptions of the process.
Downloaded	0 times
Pages	257-274
Volume	10
Journal name	Journal of Educational Technology and Society (ETS)
Publisher	International Forum of Educational Technology & Society (New York, NY, USA)
PDF	

Figura 26: Vista de una publicación

La idea detrás de esta vista es mostrar la información más relevante de las publicaciones. Se acordó con el profesor guía cuáles iban a ser los campos mostrados para cada tipo. Estos campos son los siguientes:

- Para los artículos de journal: Autores, Título, Nombre del journal, Volumen, Issue, Páginas y Fecha de la publicación.
- Para los artículos de conferencia: Autores, Título, Nombre de la conferencia, Lugar de la conferencia, Fecha de la conferencia, Nombre de la editorial y Páginas.
- Para los libros: Autores, Título, Nombre de la editorial, Ciudad de la editorial y Año de la publicación.
- Para los capítulos de libro: Autores, Título del capítulo, Editores, Título del libro, Nombre de la editorial, Ciudad de la editorial, Año de la publicación y páginas del capítulo.
- Para los editores de conferencia: Igual que los artículos de conferencia.
- Para los editores de número de journal: Igual que los artículos de journal.

4.6.7 Vista de publicaciones por autor

En esta vista se muestran las publicaciones asociadas a los autores. También existe un filtro para poder mostrar las publicaciones del autor por año. Esta vista se despliega cuando se hace click en el nombre de algún autor.

U-papers University of Chile DCC Stats FAQ Disclaimer Language Log in

Publications of Alexandre Bergel

2011 (6) Change *(If a year does not appear in the list, it is because there are no records for that particular year)*

Year 2011

[Toward Lean Development in Formally Specified Software Processes.](#)
Editors: Richard Messnarz, Damjan Ekert, Mads Christiansen, Jørn Johansen, Sonja König. **Conference:** European System & Software Process Improvement and Innovation (Roskilde, Denmark). **Pages:** 5-11-5.21. **Date:** 2011.
Authors: Cecilia Bastarrica, Julio Ariel Hurtado, Alexandre Bergel

[Counting Messages as a Proxy for Average Execution Time in Pharo.](#) REF
Conference: European Conference on Object-Oriented Programming. **Publisher:** Springer-Verlag. **Pages:** 533-557. **Date:** 2011.
Authors: Alexandre Bergel

[Analyzing Software Process Models with AVISPA.](#)
Editors: David Raffo, Dietmar Pfahl, Li Zhang. **Conference:** International Conference on Software and System Process (Honolulu, Hawaii, USA). **Publisher:** ACM Press. **Pages:** 23-32. **Date:** 2011.
Authors: Julio Ariel Hurtado, Cecilia Bastarrica, Alexandre Bergel

[Is it Safe to Adopt the Scrum Process Model?.](#) REF
Journal: CLEI Electronic Journal. **Volume:** 14. **Issue:** 3. **Pages:** paper 7. **Date:** 2011.
Authors: Julio Ariel Hurtado Alegria, Cecilia Bastarrica, Alexandre Bergel

[FlowTalk: Language Support for Long-Latency Operations in Embedded Devices.](#) REF
Journal: IEEE Transactions on Software Engineering. **Volume:** 37. **Issue:** 4. **Pages:** 526-543. **Date:** July 2011.
Authors: Alexandre Bergel, William Harrison, Vinny Cahill, Siobhan Clarke

[Reconciling Method Overloading and Dynamically Typed Scripting Languages.](#) REF
Journal: Computer Languages, Systems & Structures. **Volume:** 37. **Issue:** 3. **Pages:** 132-150. **Date:** 2011.
Authors: Alexandre Bergel

Figura 27: Vista de publicaciones por autor

Esta vista está disponible sólo con los autores que tengan activada esta opción. Lo anterior está representado por un campo binario dentro de la base de datos, en la tabla 'authors' llamado 'linkable'. Por defecto, este campo no está activo, pero sí lo está para los autores DCC. La idea es poder fácilmente activar/desactivarlo en todos los usuarios. El motivo de esta decisión es que eventualmente se desee mostrar las publicaciones de algún autor que no sea DCC.

4.6.8 Login

El *login* es un recuadro emergente que aparece en la esquina superior derecha del sitio y sirve para que los usuarios inicien sesión.

U-papers University of Chile DCC Stats FAQ Disclaimer Language Log in

Welcome to U-papers
 A publications repository from the Computer Science Department of the University of Chile

Search publications
 Advanced search

Latest publications

[Is it Safe to Adopt the Scrum Process Model?.](#) REF
Journal: CLEI Electronic Journal. **Volume:** 14. **Issue:** 3. **Pages:** paper 7. **Date:** 2011.
Authors: Julio Ariel Hurtado Alegria, Cecilia Bastarrica, Alexandre Bergel

[Adaptive Simplification of Huge Sets of Terrain Grid Data for Geosciences Applications.](#) REF
Journal: Journal of Computational and Applied Mathematics. **Volume:** 236. **Pages:** 1410-1422. **Date:** 2011.
Authors: Narcís Coll, Marité Guerrieri, María Cecilia Rvava, J. Antoni Sellarès

[Harris 3D: A Robust Extension of the Harris Operator for Interest Point Detection on 3D Meshes.](#) REF
Journal: The Visual Computer. **Volume:** 27. **Issue:** 11. **Pages:** 963-976. **Date:** 2011.
Authors: Ivan Spiran, Benjamin Bustos

[On Nonmetric Similarity Search Problems in Complex Domains.](#) REF
Journal: ACM Computing Surveys. **Volume:** 43. **Issue:** 4. **Pages:** 341-50. **Date:** 2011.
Authors: Tomas Skopal, Benjamin Bustos

[Self-Indexed Grammar-Based Compression.](#) REF
Journal: Fundamenta Informaticae. **Volume:** 111. **Pages:** 313-337. **Date:** 2011.
Authors: Francisco Claude, Gonzalo Navarro

[Supporting Empirical Software Engineering via Mining Software Repositories.](#) REF
Date: 2011.

Figura 28: Login

El inicio de sesión permite autenticar un usuario y de esa forma identificar los permisos que tiene. Dependiendo de los permisos se presentan las acciones que pueden realizar. De ahora en adelante se mostrarán las vistas que requieren de la autenticación de los usuarios.

4.6.9 Vista del Panel de control

Esta vista es dinámica (dependiendo de los permisos de usuario) y muestra las acciones que pueden ejecutar los distintos roles. Algunas de las opciones administrativas disponibles desaparecen de esta vista dependiendo del rol que tenga el usuario dentro del sistema. Por ejemplo, los Profesores no tienen acceso a las secciones *Adding* ni *Management* y los administradores no pueden agregar usuarios ni gestionarlos. La siguiente figura muestra el Panel de control para un usuario Super Administrador:

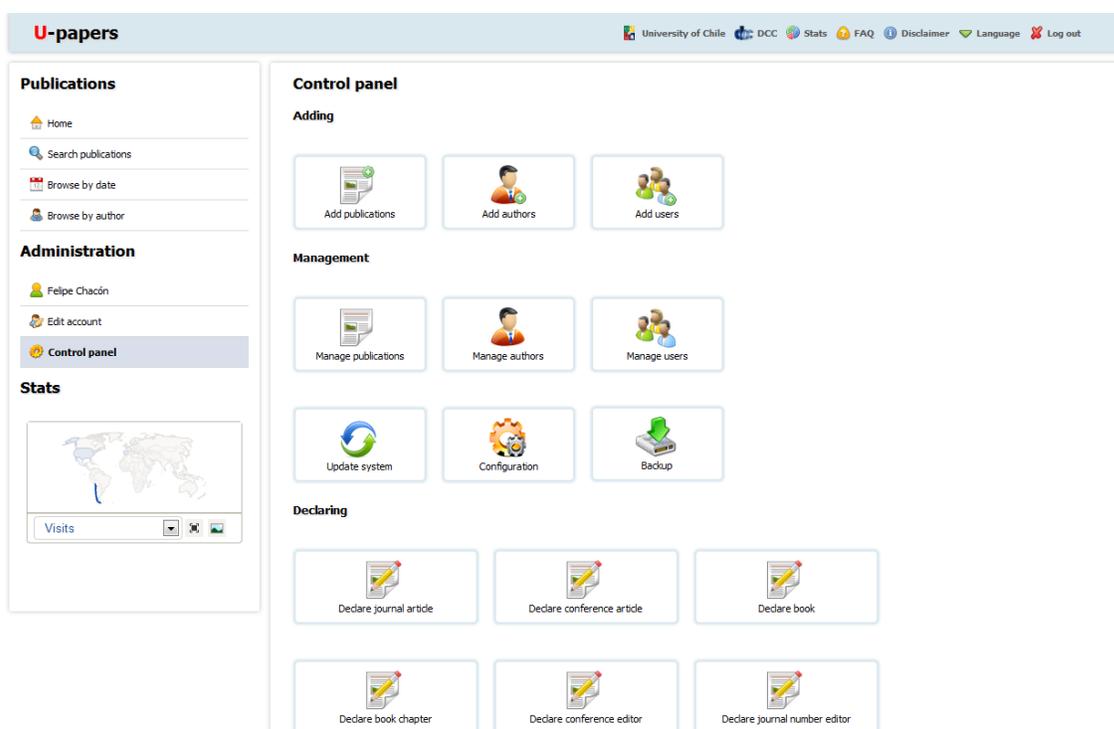


Figura 29: Vista del Panel de control para un usuario Super Administrador.

Cabe recordar que un usuario Super Administrador tiene los privilegios para realizar cualquier tipo de acción, por lo cual la vista en el Panel de Control muestra todas las opciones disponibles.

A continuación se muestra el Panel de Control para un usuario Administrador:

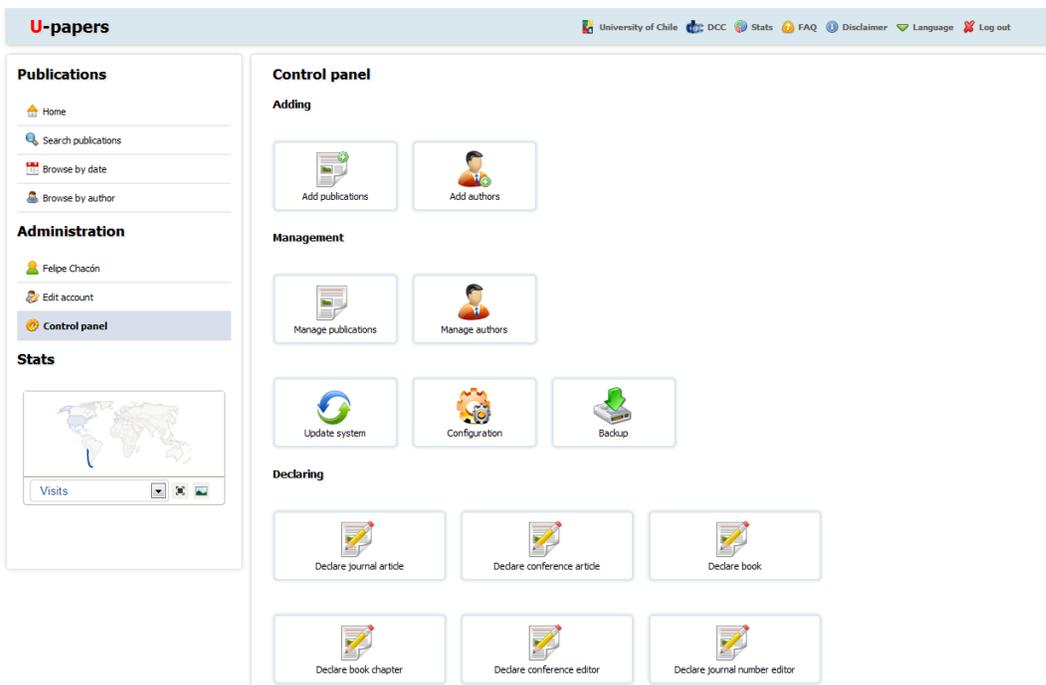


Figura 30: Vista del Panel de control para un usuario Administrador.

En la imagen anterior se puede notar que desaparecen las opciones para agregar y gestionar usuarios. A continuación se muestra la vista del Panel de Control para un usuario tipo Profesor:

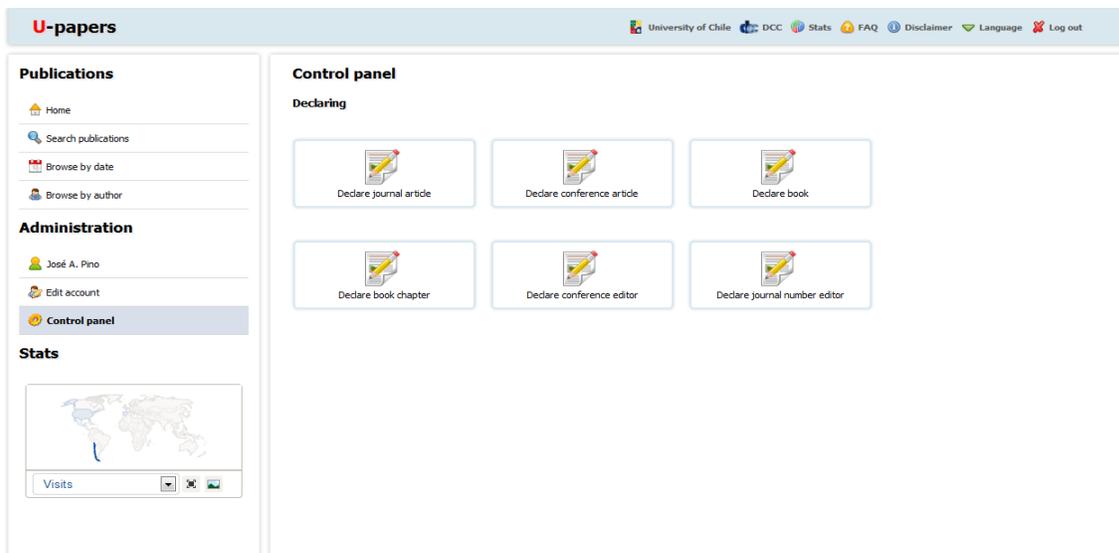


Figura 31: Vista del Panel de control para un usuario Profesor.

En la vista anterior se puede ver que un usuario Profesor sólo tiene la posibilidad de declarar publicaciones dentro del Panel de Control.

4.6.10 Vista para agregar, editar y declarar publicaciones.

Existen 3 acciones dentro de *U-papers* que utilizan la misma vista: agregar, editar y declarar publicaciones. Cabe destacar que es la misma vista para un mismo objeto, es decir, es la misma vista para agregar, editar y declarar un artículo de journal, pero es otra vista para agregar, editar y declarar un artículo de conferencia. Lo mismo ocurre para los demás objetos.

La acción de agregar publicaciones se utiliza principalmente para ingresar al sistema (es decir, a la base de datos) aquellas que son antes del 2009. El motivo de lo anterior es porque las publicaciones que fueron creadas desde el año 2009 en adelante se van a agregar desde los XML que posee el sistema FII/EP.

La acción de editar publicaciones aplica también principalmente para aquellas que fueron creadas antes del 2009, puesto que la edición no es persistente en aquellas que fueron creadas desde el 2009 en adelante. No es persistente debido a que el proceso de actualización del sistema utiliza completamente la información existente en los XML del sistema FII/EP, por ende, las actualizaciones deben ser efectuadas directamente en esos archivos para que *U-papers* muestre correctamente la información. Cabe mencionar que es posible editar, igualmente, publicaciones que hayan sido creadas desde el 2009 en adelante, pero los cambios se verán reflejados sólo hasta la próxima actualización del sistema (que se piensa hacer mínimo una vez al mes).

La acción de declarar publicaciones no refleja cambios directos en *U-papers*, puesto que se utiliza para generar un XML que llega al coordinador de investigación. Sin embargo, esta acción almacena algunos datos para usarse a futuro, cuando se actualiza el sistema. A continuación se mostrará la vista común que se utiliza para agregar, editar y declarar publicaciones:

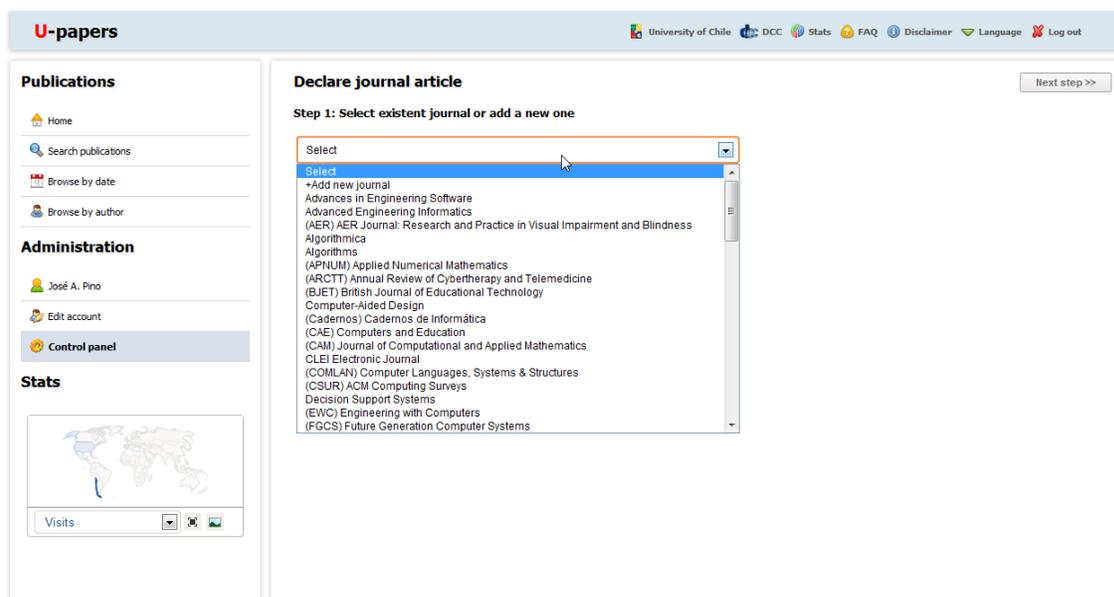
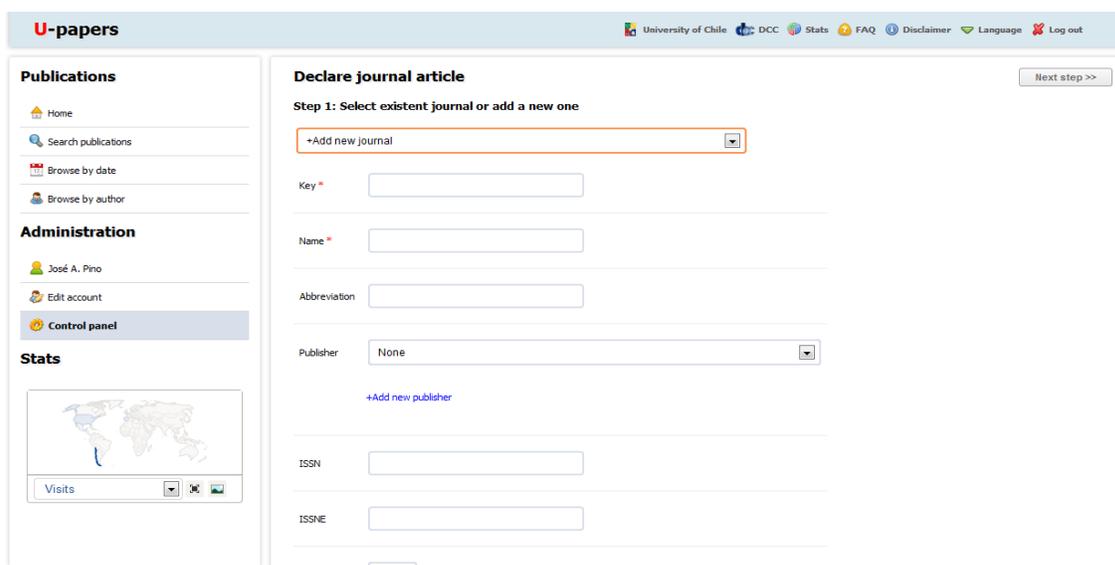


Figura 32: Vista para agregar, editar y declarar un artículo de journal (Paso 1).

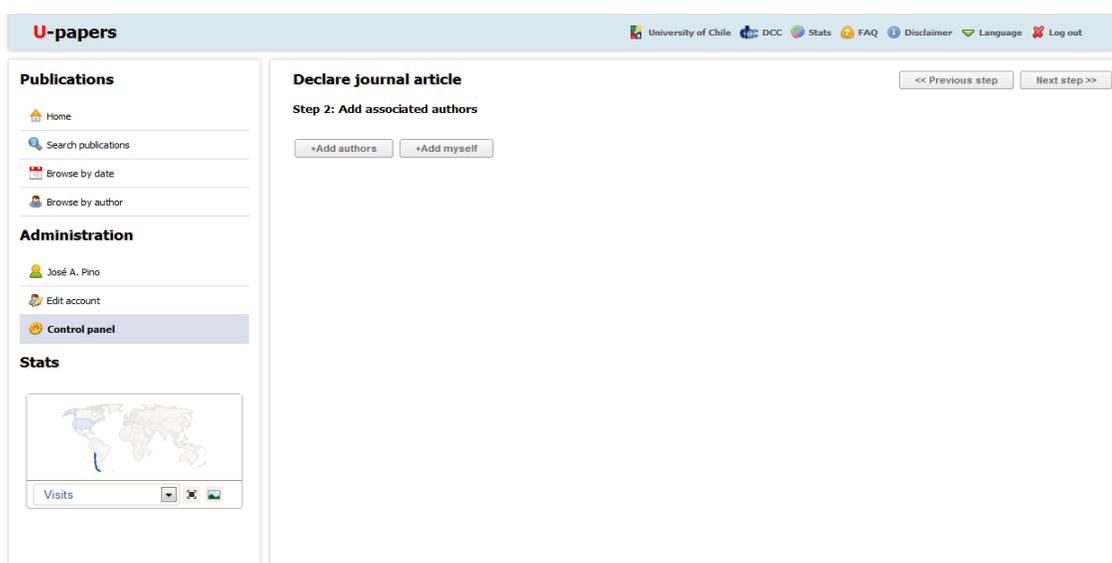
La figura anterior muestra el primer paso (de un total de cuatro) para declarar un artículo de journal. Básicamente, es un *dropdown* con todos los journals existentes en la base de datos. La primera opción es *Add new journal*, lo cual indica que si se presiona dicha opción, se despliega un formulario para agregar un nuevo journal. En general, este tipo de situaciones se gestiona de esa forma en *U-papers*: se puede agregar *on the fly* elementos nuevos al sistema. A continuación se muestra la situación anterior:



The screenshot shows the 'U-papers' interface. On the left is a sidebar with 'Publications' (Home, Search publications, Browse by date, Browse by author), 'Administration' (José A. Pino, Edit account, Control panel), and 'Stats' (Visits). The main content area is titled 'Declare journal article' and shows 'Step 1: Select existent journal or add a new one'. It features a '+Add new journal' dropdown menu, followed by input fields for 'Key', 'Name', and 'Abbreviation'. The 'Publisher' field is a dropdown menu currently set to 'None', with a '+Add new publisher' link below it. Below these are input fields for 'ISSN' and 'ISSNE'. A 'Next step >>' button is located in the top right corner of the form area.

Figura 33: Agregar *on the fly* un journal nuevo.

Se puede notar que arriba, a la derecha, existe un botón llamado *Next step*. Dicho botón permite ir avanzando por los distintos pasos. Se hizo de esta forma debido a que la cantidad de información a ingresar es numerosa, por lo que desplegar un gran formulario en una sola página iba a resultar invasivo y confuso. Al presionar el botón antes mencionado, se muestra el segundo paso que sirve para asociar autores a la publicación:



The screenshot shows the 'U-papers' interface. The sidebar is identical to the previous figure. The main content area is titled 'Declare journal article' and shows 'Step 2: Add associated authors'. It features two buttons: '+Add authors' and '+Add myself'. Navigation buttons '<< Previous step' and 'Next step >>' are located in the top right corner of the form area.

Figura 34: Asociar autores (Paso 2).

En este segundo paso se observan dos botones: *Add authors* y *Add myself*. El primero permite agregar autores desde un listado existente o para agregar *on the fly* uno nuevo. A continuación se muestra lo anterior:

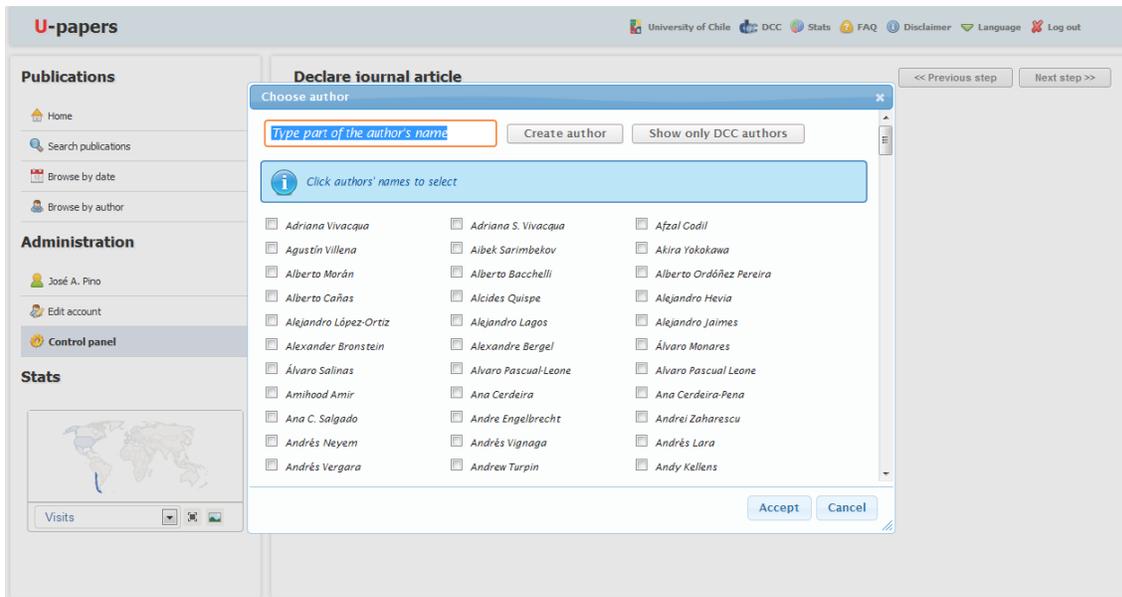


Figura 35: Ventana emergente al hacer click en *Add authors*.

Desde la ventana emergente se puede ir seleccionando autores. Se respeta el orden seleccionado. También se puede filtrar autores mediante el *textfield* de arriba. Arriba de esta ventana hay dos botones: *Create autor* y *Show only DCC authors*. El primero permite agregar *on the fly* un autor que no exista aún en *U-papers* y el segundo permite reducir la lista de autores a sólo los que son DCC. Una vez que se seleccionan los autores y se presiona *Accept*, se muestran en un listado los autores:

El botón *Add myself* utiliza los datos de sesión para agregar directamente al autor que está declarando la publicación. Se acordó con el Prof. Navarro que el sistema debía explotar al máximo la información que tiene para ser lo más intuitivo posible. La idea es siempre facilitar el proceso.

Una vez elegidos los autores, se muestra el listado:

U-papers University of Chile DCC Stats FAQ Disclaimer Language Log out

Publications

- Home
- Search publications
- Browse by date
- Browse by author

Administration

- José A. Pino
- Edit account
- Control panel

Stats

Visits

Declare journal article << Previous step Next step >>

Step 2: Add associated authors

+Add authors +Add myself

Alejandro Hevia DCC Author Affiliation Unique (only DCC) Delete

José A. Pino DCC Author Affiliation Shared (DCC not first) Delete

Álvaro Salinas Author type PhD student of a DCC Author Chile Delete

Figura 36: Listado de autores seleccionados.

Cuando un autor no es DCC, tiene 2 datos asociados que se muestran en el listado anterior: el tipo de autor y el país. Estos datos son pre-llenados con la información existente en la base de datos, pero se guarda la información si se modifica.

El tercer paso muestra un formulario que sirve para ingresar la información misma de la publicación, como se muestra a continuación:

U-papers University of Chile DCC Stats FAQ Disclaimer Language Log out

Publications

- Home
- Search publications
- Browse by date
- Browse by author

Administration

- José A. Pino
- Edit account
- Control panel

Stats

Visits

Declare journal article << Previous step Next step >>

Step 3: Add the journal article info

Title *

Volume

Number

Pages

Month Select

Year *

Abstract

Figura 37: Datos de la publicación (Paso 3).

Claramente el paso anterior varía para los distintos tipos de publicación que se manejan en *U-papers*.

El último paso es para agregar los ítems que sirven para decidir si la publicación puede ir dentro de la página de inicio o no, es decir, aquellos que contienen el PDF borrador, enlace a una página de referencia de la editorial, una página con una versión en HTML de la publicación o el DOI. A continuación se muestra el formulario del último paso:

The screenshot shows the 'U-papers' web interface. The top navigation bar includes the University of Chile logo and links for DCC, Stats, FAQ, Disclaimer, Language, and Log out. The left sidebar contains navigation options: Home, Search publications, Browse by date, Browse by author, Administration (with a sub-menu for José A. Pino, Edit account, and Control panel), and Stats (with a world map and a Visits dropdown). The main content area is titled 'Declare journal article' and is at 'Step 4: Add the associated file'. It features a '<< Previous step' button. The form contains the following fields:

- PDF original:** A file selection button labeled 'Seleccionar archivo' followed by the text 'No se ha...archivo' and a red note: 'Only PDF is allowed. 40 MB maximum.'
- PDF borrador:** A file selection button labeled 'Seleccionar archivo' followed by the text 'No se ha...archivo' and a red note: 'Only PDF is allowed. 40 MB maximum.'
- Publication URL:** A text input field containing 'http://', with a red note below it: 'This URL points to a page where the publication is contained like a webpage instead of a PDF.'
- Reference URL:** A text input field containing 'http://', with a red note below it: 'This URL points to the publisher's page where the publication can be downloaded/purchased.'
- DOI:** An empty text input field.

At the bottom of the form are 'Save' and 'Cancel' buttons.

Figura 38: Datos de la publicación (Paso 4).

Luego de completar los 4 pasos y darle click al botón Save, se muestra un mensaje de éxito en caso de haber completado correctamente el formulario.

4.6.11 Vistas de la sección Management.

La sección anterior era muy relevante para el funcionamiento de *U-papers*, ya que se trata del punto fundamental del sistema, es decir, las publicaciones. En la sección *management* se enfoca toda la mantención del sistema. Aquí se pueden editar y eliminar publicaciones, autores o usuarios del sistema. También se puede actualizar el sistema y configurar algunos parámetros para el correcto funcionamiento de *U-papers*. Las vistas para gestionar publicaciones, autores y usuarios es muy similar: se muestra un listado con los ítems correspondientes y luego una serie de opciones. En este punto se mostrará sólo la gestión de autores, en la figura siguiente:

U-papers University of Chile DCC Stats FAQ Disclaimer Language Log out

Publications

- Home
- Search publications
- Browse by date
- Browse by author

Administration

- Felipe Chacón
- Edit account
- Control panel**

Stats

Visits

Manage authors

Type part of the author's name Show 10 Number of items to display on paginations Search

Name	Country	Type	Linkable?	Picture	Options
Adriana Vivacqua	Brazil	None	No		Edit Delete Replace
Adriana S. Vivacqua	Brazil	None	No		Edit Delete Replace
Afzal Godil	United States	None	No		Edit Delete Replace
Agustín Villena	Chile	None	No		Edit Delete Replace
Albek Sarimbekov	Switzerland	None	No		Edit Delete Replace
Akira Yokokawa	Switzerland	None	No		Edit Delete Replace

Figura 39: Gestión de autores.

En este caso, se muestran tres opciones: editar, eliminar y reemplazar. Editar un autor significa modificar sus datos, manteniendo su presencia. Es distinto el caso de eliminar y reemplazar. Cuando se quiere eliminar un autor, existen tres caminos posibles a seguir, en caso de que el autor tenga publicaciones asociadas. Las tres opciones se muestran en una pregunta de decisión, como se muestra a continuación:

U-papers University of Chile DCC Stats FAQ Disclaimer Language Log out

Publications

- Home
- Search publications
- Browse by date
- Browse by author

Administration

- Felipe Chacón
- Edit account
- Control panel**

Stats

Visits

What do you want to do?

The author Adriana Vivacqua has some associated publications. Please decide what to do next.

- Delete author. This includes deleting his/her name of the corresponding publications.
- Keep only his/her name in the corresponding publications, but hide him/her in all other instances.
- Replace him/her with another author. You have to choose one after selecting this option.

Save Cancel

Figura 40: Pregunta de decisión sobre la eliminación de un autor.

La primera opción es eliminar completamente al autor y no dejar rastro de él dentro de *U-papers*. La segunda opción es ocultarlo en todas las instancias, excepto cuando aparece su nombre en una publicación. Lo anterior quiere decir que el autor no podrá ser sugerido en una búsqueda ni cuando se desee asociar a una nueva publicación. Finalmente, la tercera opción es reemplazarlo por otro autor. La situación

anterior ocurre, generalmente, cuando se ingresa un autor dos o más veces dentro del sistema.

4.7 Algoritmos desarrollados

Existen algunos procesos dentro del funcionamiento de *U-papers* que son muy relevantes. Los más destacables son:

1. Declaración de publicaciones.
2. Búsqueda sencilla de publicaciones.
3. Búsqueda avanzada de publicaciones.
4. *Parsing* de archivos XML.
5. Proceso de actualización del sistema.

4.7.1 Algoritmo de declaración de publicaciones

La declaración de publicaciones se realiza en los siguientes pasos:

1. Obtener datos desde el formulario.
2. Generar un *string* con el XML resultante.
3. Guardar datos en la tabla 'pasive publications'.
4. Enviar un email a coordinador-investigacion@dcc.uchile.cl y a los autores DCC asociados a la publicación.
5. Mostrar un mensaje de éxito.

El primer paso consiste en validar los datos entregados en el formulario y, en caso de éxito, ir guardando los datos en variables con nombres apropiados para luego formar un gran *string* que representa el XML resultante. Por ejemplo, el título se guarda de la siguiente forma:

```
$title = $space."<title>".$this->input->post('title')."</title>". "\r\n";
```

La variable `$space` es una indentación predefinida en el código que se concatena con un texto XML (que en este caso es `<title>`), luego el dato ingresado en el formulario (`$this->input->post('title')`), luego se cierra el XML (con `</title>`) y finalmente un salto de línea (`\r\n`).

Finalmente, el *string* XML resultante (para el caso de un artículo de journal) es de la forma siguiente:

```
$xml .= $name.$title.$volume.$authors['xml'].$number.$pages.$month.$year  
      .$note.$ack.$doi.$type_article.$url.$persurl.$abstract;  
$xml .= "</journal>". "\r\n";  
$xml .= $def_journal;  
$xml .= $publisher;
```

Inicialmente, la variable \$xml es igual al *string* '<journal>.\r\n'. Luego se concatenan los datos ingresados en el formulario y finalmente se concatenan dos XML que representan los casos donde se define un nuevo journal (no confundir con artículo de journal!) y una nueva editorial.

Los mismos datos que se usan para completar el XML son usados (algunos) para guardarse en la tabla 'passive publications'. Ya fue explicado anteriormente que estos datos son usados durante el proceso de actualización del sistema para poder obtener algunos datos que no vienen dentro del XML, como por ejemplo, la fecha de la declaración de la publicación. La fecha anterior sirve para poder decidir si ha pasado el plazo predefinido para poder activar el PDF borrador dentro de *U-papers*.

Finalmente, cuando se han realizado los subprocessos anteriores, se envía un email con el XML resultante al coordinador de investigación, a los autores asociados a la publicación y luego un mensaje de éxito en la página de *U-papers*.

4.7.2 Algoritmo de búsqueda sencilla de publicaciones

La búsqueda sencilla recibe como parámetro una frase. Esta frase se utiliza para realizar un match con el *abstract* de la publicación. Se acordó con el Prof. Navarro que se les iba a pedir a los autores que agregaran los *keywords* de las publicaciones que declararan dentro del *abstract*. El *match* se realiza mediante SQL de la siguiente forma:

```
MATCH(P.abstract) AGAINST (".$this->db->escape($phrase).")
```

Para poder utilizar el *statement* "MATCH ... AGAINST" de SQL, la tabla en cuestión debe utilizar el tipo de almacenamiento MyISAM. El tipo de almacenamiento por defecto de MySQL (InnoDB) no permite el uso de lo anterior.

Además de realizar un *match* con el *abstract* de la publicación, se realiza un *match* de cada palabra presente en la frase con el título de la publicación. Finalmente se retornan los resultados encontrados, los cuales son menos precisos que los encontrados en el siguiente tipo de búsqueda.

4.7.3 Algoritmo de búsqueda avanzada de publicaciones

El algoritmo de búsqueda avanzada recibe una serie de datos como input. Estos datos son: *title words*, *date of publication*, *type of publication*, *authors*, *keywords*, *order results by* y *show 'x' items per page*.

Las palabras ingresadas en *title words* deben aparecer todas en el *title* de la publicación (se usa el operador *AND* de SQL). Las palabras ingresadas en *keywords* son utilizadas para hacer un *match* con el *abstract* de la publicación, de la misma forma que se hace en la búsqueda sencilla. Los demás datos son utilizados para una comparación usando el operador '=' de SQL. Todos los datos anteriores son agregados en una gran consulta usando el operador *AND* y finalmente los resultados se ordenan

mediante el dato ingresado en *order results by*, mostrando una cantidad 'x' de resultados por página.

4.7.4 Algoritmo de parsing de archivos XML

Existen algunas librerías en PHP que sirven para realizar un proceso de *parsing* de archivos XML y así obtener datos. Se probó la librería SimpleXML pero no con muy buenos resultados. La librería no funcionaba muy bien debido a dos cosas: los archivos XML del sistema FII/EP no eran archivos XML bien formados: en general no tenían un *tag* de inicio y además tenían texto insertado fuera de los *tags*. Con lo anterior, el *array* retornado por la librería era incorrecto.

Se decidió crear una serie de librerías propias que preveían esta serie de situaciones y procesaban correctamente estos archivos XML. Cada librería se representa por un archivo PHP con dos funciones: *parse* y *load*. La primera lee un archivo XML y retorna un *array* con los datos deseados. La segunda función sirve para el proceso de actualización del sistema y realiza una carga dentro de la base de datos.

Cada función *parse* de las librerías utiliza principalmente la función *stripos* de PHP. Según el manual de PHP¹⁸, la función *stripos* encuentra la posición de la primera aparición de un *substring* insensible a mayúsculas y minúsculas en un *string*. En general el proceso es el siguiente, para cada dato:

```
/*Key*/
$inicio_key = stripos($current_source,"<key>",$current_inicio) + 5;
$fin_key = stripos($current_source,"</key>",$inicio_key) - 1;
$data_aux['KEY'] = trim($CI->my_string_lib->replace_strange_xml_chars(
utf8_encode(substr($current_source, $inicio_key, $fin_key - $inicio_key + 1))));
```

El código anterior representa la obtención del dato *key* de la declaración de un *journal* dentro del archivo *journals.xml*. Se puede observar que los datos son guardados dentro de un *array*, el cual es insertado dentro de otro gran *array* que guarda los datos de todas las declaraciones de *journals* (el *array* anterior guardaba los datos para sólo una declaración de *journal*).

4.7.5 Algoritmo de proceso de actualización del sistema

El proceso de actualización es de hecho bastante sencillo de desarrollar una vez hechas las librerías de *parsing* y *load* de los XML. El código de este proceso es el siguiente:

```
function fp_update() {
    if(!$this->_is_logged_in()) return;
    if($this->session->userdata('rol') != 'superadmin' && $this->session->userdata('rol') != 'admin') return;
    $libs = array(
        'my_acks_xml_reader','my_affiliations_xml_reader','my_alumno_xml_reader',
        'my_authors_xml_reader','my_conf_xml_reader','my_journals_xml_reader',
        'my_countries_xml_reader','my_publishers_xml_reader','my_publications_xml_reader',
        'my_publications_types_xml_reader'
    );
}
```

¹⁸ <http://php.net/manual/es/function.stripos.php>

```

$j = 0;
$global_result = true;
$this->data['ERRORS'] = array();
for($i = 0; $i < count($libs); $i++) {
    $this->load->library($libs[$i]);
    if($libs[$i] == 'my_publications_xml_reader')
        continue;
    $result = $this->$libs[$i]->parse();

    if(isset($result['result']) && $result['result'] === FALSE) {
        $global_result = false;
        $this->data['ERRORS'][$j++] = $result['error_msg'];
    }
}

if($global_result) {
    for($i = 0; $i < count($libs); $i++)
        $this->$libs[$i]->load();

    $this->data['EXITO'] = lang('site_update_successfully');
}
$this->data['MAIN_CONTENT'] = 'update';
$this->load->view('includes/template',$this->data);
}

```

Primero se verifican los permisos necesarios para poder ejecutar esta función. Luego se crea un *array* con los nombres de las librerías. A continuación, se entra a un ciclo para ejecutar la función *parse* de cada librería (salvo el caso especial de la librería 'my_publications_xml_reader'). Si algún proceso de *parsing* de las distintas librerías entrega un resultado incorrecto, se detiene el proceso y se guarda la variable *\$global_result* como *false*. Si todo el proceso de *parsing* es correcto, entonces se ejecuta la carga de los datos y se guarda un mensaje de éxito que se muestra en la vista *update*.

5. Seguridad del sitio

El sistema fue desarrollado con el requisito de al menos abordar problemas típicos de seguridad de forma correcta. Los problemas de seguridad más frecuentes son:

- *SQL injections*.
- Interceptación de datos importantes.
- Acceso a áreas restringidas del sitio.
- Suplantación de identidad.

Para el caso de las *SQL injections*, fue un problema abordado de forma muy sencilla. Como se mencionó en el capítulo 2, sección 2.1 (Tecnologías involucradas), *U-papers* fue desarrollado usando el *framework* llamado Codelgniter. Esta herramienta tiene una librería llamada *Database*. Dentro de esta librería, además, existe una clase llamada *Active Record*. Esta clase, además de algunas funciones extras presentes en la librería *Database*, permiten la ejecución de *SQL queries* de forma segura. Es el mismo *framework* el que se encarga de sanitizar todo tipo de sentencia SQL. De esta forma, el problema de *SQL injections* se soluciona gracias al uso mismo de las funciones presentes en la librería *Database*. Por ejemplo, para insertar un registro en la base de datos, basta invocar la función *insert* de la librería *Database*, pasando como argumentos el nombre de la tabla donde se quiere ingresar el registro y un *array* con los datos a insertar:

```
$datos = array('title' => 'Collaborative work', 'year' => '2011');  
$this->db->insert('publications', $datos);
```

Para el caso de la interceptación de datos importantes, la solución viene de la mano con utilizar una conexión que encripte los datos transferidos. Para eso se debió configurar el servidor *Apache* y luego utilizar un certificado de seguridad. En resumen, lo anterior es una conexión segura utilizando SSL. El certificado se obtuvo de forma gratuita gracias a RapidSSL¹⁹. Según el Prof. Guía, la idea es probar con este certificado por al menos un mes y luego comprar uno, para poder tener todos los beneficios que ofrecen dentro de la página de RapidSSL.

Para el caso del acceso a áreas restringidas del sitio, se utilizó un sistema de *login*. Éste permite autenticar un usuario y saber cuál es su rol dentro del sistema. Teniendo el rol del usuario, se restringe mediante código el acceso a las distintas áreas del sistema.

Para el último caso de suplantación de identidad, sólo se aborda el caso en que a un usuario con un rol definido deja su sesión abierta sin realmente desearlo y alguna otra persona podría cambiar los datos personales de dicha persona (por ejemplo, la contraseña). La forma de combatir este problema es pedir nuevamente la contraseña original, la cual sólo debería saber la persona quien dejó la sesión abierta dentro del sistema.

¹⁹ <https://www.rapidsslonline.com/freessl-certificates.aspx>

Estos cuatro puntos referentes a seguridad dentro del sitio, fueron los únicos abordados en conjunto con el Prof. Guía y Co- guía.

6. Testing

El *testing* realizado a *U-papers* fue básico. El sistema se desarrolló funcionalidad por funcionalidad y se realizaban pequeñas pruebas cada vez que se terminaba de desarrollar cada una de las funcionalidades. Una vez que se iba teniendo un producto funcional, se insertó mediante un *script* datos reales que iban a ir dentro del sistema. La primera vez que se insertaron datos, fueron obtenidos desde la página del sistema FII/EP. Finalmente, se decidió que los datos fueran obtenidos desde archivos XML.

La inserción de datos desde los archivos XML fue ejecutada en numerosas ocasiones, para probar el funcionamiento del *script*. Luego de los ajustes necesarios, se obtuvo un código que ejecuta de forma correcta la inserción de datos desde todos los XML necesarios para alimentar *U-papers*.

Durante la última semana de desarrollo, el sistema fue puesto en marcha blanca y presentado a todos los docentes del DCC. Este proceso sirvió como un *testing* del sitio y para que ellos mismos verificaran el funcionamiento del sistema. Gracias a esto, se obtuvo un valioso *feedback* que sirvió para mejorar el sistema.

7. Respaldos

La política de respaldos no fue discutida junto al Prof. Guía puesto que *U-papers* no es un sistema que crece rápidamente. De hecho las actualizaciones (inserciones de nuevas publicaciones) están pensadas para realizarse una vez al mes. A su vez, el número de publicaciones es alrededor de 120 cada año. Por otro lado, los usuarios asociados al sistema son pocos, alrededor de 25. La decisión fue propia del alumno memorista y se optó por entregar un enlace llamado *Backup* dentro del Panel de Control. Este enlace, al ser ejecutado, permite guardar un archivo de texto, con las instrucciones *SQL* que permiten reproducir la base de datos en ese momento.

8. Conclusiones

Las expectativas que existían sobre *U-papers* fueron cubiertas en este trabajo de memoria. Resultó un trabajo bien elaborado, intuitivo, amigable y funcional. Gracias a este trabajo, existe un antes y un después en las publicaciones del DCC. El “antes” era una época en donde las publicaciones tenían muy poca visibilidad y sin acceso libre a los archivos *PDF*. El “después” es una época en donde la visibilidad está dada por el sitio web de *U-papers*, en donde se provee a los investigadores un espacio para compartir los *PDF* de sus publicaciones.

Además de cumplir el objetivo principal, *U-papers* resultó ser también una herramienta muy útil para agilizar y corregir el proceso de declaración de publicaciones. Cabe recordar que el proceso de declaración de publicaciones era una tarea de los investigadores hacia el sistema FII/EP en donde se enviaba un XML (con muchas fallas) por email. De ahora en adelante, la declaración de publicaciones se desarrolla en *U-papers*, el cual provee una interfaz muy sencilla. La interfaz guía a los docentes mediante un formulario con información precargada. Terminado el formulario, automáticamente se genera un XML (ahora sin errores) y se envía el email al coordinador de investigación.

Desde el término de la memoria en adelante, sólo falta que el sistema crezca en su uso, inicialmente por el DCC, y así la cantidad de archivos *PDF* asociados a las publicaciones empiece a aumentar. Con esto, el incentivo a que las personas ajenas al DCC utilicen el sistema será aún mayor.

9. Referencias

- [1] Repositorio de la Universidad de Southampton.
<http://eprints.ecs.soton.ac.uk>
- [2] Repositorio de la Universidad de Pompeu Fabra. *<http://repositori.upf.edu/>*
- [3] ISI web of knowledge.
http://apps.isiknowledge.com/WOS_GeneralSearch_input.do?product=WOS&search_mode=GeneralSearch&SID=4BgOcklnkPHLHKmf6jL&preferencesSaved=&highlighted_tab=WOS
- [4] Índices H.
http://images.isiknowledge.com/WOK45/help/WOS/h_citationrpt.html
- [5] Documentación de Code Igniter.
http://codeigniter.com/user_guide/index.html