

Explorador de recursos de información.

2DO. WORKSHOP NACIONAL_

28_SEPTIEMBRE 2016
LANÚS | BUENOS AIRES

Tecnologías e infraestructuras de la información para
Bibliotecas y Centros de Documentación e Información



Vufind

Busca, descubre y comparte.

Aspectos técnicos para la elección de VUFIND
como explorador de recursos de información
para la Universidad.

Codificado casi enteramente en PHP

- Tecnologías OPENSOURCE
- Desarrollo continuo.
- Amplia comunidad de usuarios / desarrolladores.
- Lectura sencilla de código, código base muy estable.
- Respeta estándares de programación.
- Gran respeto por los novatos en las listas de discusión.

Solo PHP ? No es tan cierto.

- Java en el motor SOLR.
 - Java en los scripts de importación.
 - Java en los scripts de conversión de datos.
 - Java en los scripts de control y updates.
 - Javascript en el frontend.
 - XML y XSLT .
- Ventaja: poco hay que tocar o comprender, los script realmente funcionan.

Otras tecnologías asociadas

- PHP
 - Symfony
- SOLR
 - Jetty webserver o Tomcat
- Javascript
 - JQuery.
- XML / XSLT
 - Usado para los transformaciones.
- SQL
 - Mysql o Postgresql.



Componentes de VUFIND

- Un motor de búsqueda: APACHE SOLR
- Una base de datos para guardar preferencias de usuarios y configuraciones.
- Scripts de importación.
- Un front-end para el público
 - Acceso al motor de indexación
 - Conexión a otros sistemas mediante módulos genéricos.
 - Conexión a con servicios webs (portadas, wikipedia, etc.)



Componente Apache SOLR

- Motor de búsqueda de código abierto
- Principales características en la búsqueda
 - Búsquedas a texto completo.
 - Soporta búsqueda por facetas, resultados agrupados.
 - Ordenamiento por relevancia.
 - Búsqueda por similitud.
 - Cache.
 - Destacado de texto (coloreado).



Componente Apache SOLR

- Características técnicas
 - Escrito en JAVA
 - Definición del índice mediante archivos xml.
 - Interfase de administración integrada.
 - Alto rendimiento.
 - Escalable.
 - Mantenido desde la fundación APACHE.
 - Indexación en tiempo real.
 - Librerías en distintos lenguajes (PHP,Perl, Python, etc.)

Base de datos

- Mysql o Postgresql
 - Usada para guardar preferencias de usuarios, búsquedas anteriores, etc.
 - No opcional, requerido por el sistema.
 - Estructura simple de datos.
 - No representa carga para el sistema.

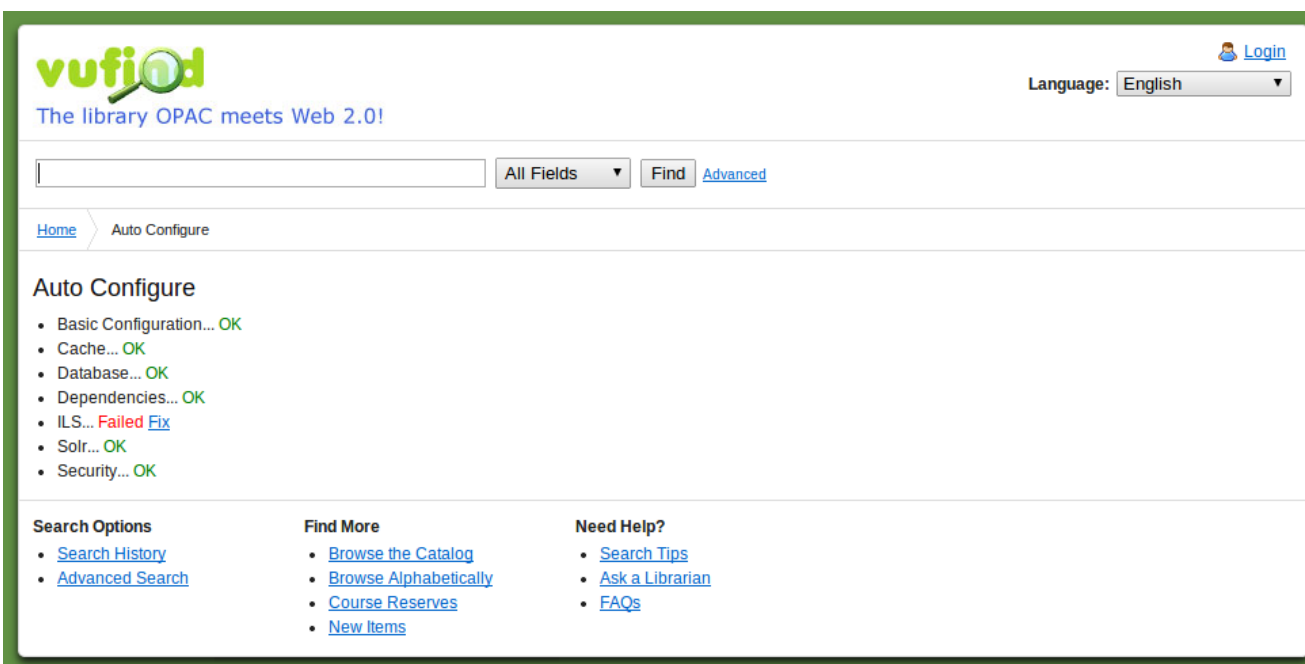
Componente: scripts de importación

- Para la ingesta de los datos se utilizan:
 - Scripts realizados en bash que contienen llamadas a
 - Scripts en java que utilizan
 - Archivos texto .properties que sirven para obtener los datos utilizando
 - Scripts java beanshell que hacen modificaciones en los datos obtenidos y uso de listas de traducción.
 - Expresadas en Archivos planos de texto con formatos diversos.
- El punto más complejo para las personalizaciones de VUFIND. Sobre todo para bases de datos no normalizadas que requieran cambios.



Componente: Front-end

- Escrito en PHP, XSLTs, HTML, Javascript.
- Requiere instalación de extensiones de php específicas para su funcionamiento.
- Instalador simplificado en las últimas versiones.



The screenshot displays the vufind library OPAC front-end. At the top left is the vufind logo with the tagline "The library OPAC meets Web 2.0!". On the top right, there is a "Language: English" dropdown menu and a "Login" button. Below the header is a search bar with a dropdown menu set to "All Fields" and a "Find" button with a link to "Advanced". A breadcrumb trail shows "Home" and "Auto Configure". The main content area is titled "Auto Configure" and lists several system status items: "Basic Configuration... OK", "Cache... OK", "Database... OK", "Dependencies... OK", "ILS... Failed Fix", "Solr... OK", and "Security... OK". At the bottom, there are three columns of links: "Search Options" (Search History, Advanced Search), "Find More" (Browse the Catalog, Browse Alphabetically, Course Reserves, New Items), and "Need Help?" (Search Tips, Ask a Librarian, FAQs).

Esquema del frontend

- Implementa todas las funcionalidades para obtener los datos desde el índice SOLR o desde los módulos activados.
- Uso de archivos de configuración completamente documentados para modificar funcionalidades, activar o desactivar opciones y cambiar comportamientos.
- El uso de plantillas modificables permite adaptar diseño, configurar funcionalidades, desactivar ciertas cosas.

Ejemplo de archivo configuración

FACETS.ini

```
; The name of the index field is on the left
; The display name of the field is on the right
[Results]
institution      = Institution
building        = Library
format          = Format

; Use callnumber-first for LC call numbers, dewey-hundreds for Dewey Decimal:
callnumber-first = "Call Number"
;dewey-hundreds  = "Call Number"
;hierarchy_top_title = Collections
authorStr       = Author
language        = Language
genre_facet     = Genre
era_facet       = Era
geographic_facet = Region
publishDate     = "adv_search_year" ; share year string w/advanced search page

; Facets that will appear at the top of search results when the TopFacets
; recommendations module is used. See the [TopRecommendations] section of
; searches.ini for more details.
[ResultsTop]
topic_facet     = "Suggested Topics"

; This section is used to identify facets for special treatment by the SideFacets
; recommendations module.
[SpecialFacets]
; Any fields listed below will be treated as year-based date ranges rather than plain
; facets:
dateRange[] = publishDate
; Any fields listed below will be treated as year/month/day-based date ranges rather
; than plain facets:
fullDateRange[] = example_field_date
; Any fields listed below will be treated as numeric ranges rather than plain facets:
```

.ini ▾ Ancho de la tabulación: 8 ▾ Ln 1, Col 1 INS



Plantilla de visualización

- Desarrolladas en PHP, dependen de un controlador para cada tipo de registro (SOLRdefault, SOLRweb, Summon, PRIMO, etc). Se requieren conocimientos de PHP y de los objetos SOLR.

```
$this->escapeHtml($journalTitle)?></a>
  <? $ref = $this->driver->getContainerReference(); if (!empty($ref)) { echo $this->escapeHtml($ref); } ?>
</td>
</tr>
<? endif; ?>

<? $nextTitles = $this->driver->getNewerTitles(); $prevTitles = $this->driver->getPreviousTitles(); ?>
<? if (!empty($nextTitles)): ?>
<tr>
  <th><?=$this->transEsc('New Title')?>: </th>
  <td>
    <? foreach($nextTitles as $field): ?>
      <a href="<?=$this->record($this->driver)->getLink('title', $field)?>"><?=$this->escapeHtml($field)?></a><br/>
    <? endforeach; ?>
  </td>
</tr>
<? endif; ?>

<? if (!empty($prevTitles)): ?>
<tr>
  <th><?=$this->transEsc('Previous Title')?>: </th>
  <td>
    <? foreach($prevTitles as $field): ?>
      <a href="<?=$this->record($this->driver)->getLink('title', $field)?>"><?=$this->escapeHtml($field)?></a><br/>
    <? endforeach; ?>
  </td>
</tr>
<? endif; ?>
```

Funcionalidades principales

- Resultados Agrupados (FACETAS)
 - Permite definir que facetas aparecen y crear nuevas.
 - Las facetas permiten realizar filtros AND y OR (selección simple y múltiple).
 - Gran detector de errores de carga y captura.
 - Definidos en facets.ini


Colecciones	▲
Libros-FFyL	63,864
Libros B. Central	46,923
Libros FCPyS	41,249
Libros FCA	19,581
Biblioteca Inst. Balseiro	17,051
Libros de la FAD	16,264
más ...	

Formato	▲
<input type="checkbox"/> Impreso	16,058
<input type="checkbox"/> Artículo en línea	2,624
<input type="checkbox"/> CD-Rom	144
<input type="checkbox"/> Libro en línea	85
<input type="checkbox"/> Online	36
<input type="checkbox"/> Video disco digital	11
más ...	

Resultados similares

- Implementación algoritmo MoreLikeThis
- Permite detectar las distintas ediciones de un documento, resultados similares, gran aporte para bases de datos con múltiples orígenes.

Ejemplares Similares

 [Orígenes sociales y culturales del nacionalismo marroquí\(1830-1912\).](#)
por: Laroui, Abdallah,
Publicado: (1997)

 [Sistemas políticos del Magreb actual.](#)
por: Saaf, Abdallah
Publicado: (1996)

 [Communication progressive du Français : avec365 activités.](#)
por: Miquel, Claire
Publicado: (2003)

Orden por relevancia

- Ordenamiento por relevancia, permite realizar un ranking de los resultados.
- El orden responde a las distintos tipos de búsqueda.
- Se define en searchspecs.yaml (otro lenguaje)
- Permite definir que campos están involucrados en el orden, qué pesos, qué ordenamiento. Requiere lectura del manual de SOLR.
- Ejemplo:
 - Author:
 - DismaxFields:
 - - author^100
 - - author_fuller^50
 - - author2
 - - author_additional
 - **Aplicar el principio básico: si anda no tocar.**

Coloreado de resultados

- Al realizar una búsqueda se colorea las ocurrencias encontradas.
- Permite detectar que es lo que el sistema detecta como resultado, simplifica la lectura de los resultados.
- Implementación del motor SOLR.
- Se implementa en searches.ini

1

Economía : elementos de micro y macroeconomía.

por Beker, Víctor Alberto

Publicado 1995

Materias: "; "... Economía ..."

2

Economía .

Publicado 1988

Materias: "; "... Economía ..."



BIBLIOTECAS
.uncuyo.edu.ar

Módulo de exportación y citas

- Permite ofrecer los registros en formatos como marcxml, marc, refworks, endnote.
- Las citas en formato apa, chicago style, MLA.
- Se habilitan por .ini, se programan en las plantillas.



The screenshot displays a user interface for document management. On the left, a button labeled 'Exportar Registro' is active, opening a dropdown menu with three options: 'Exportar a RefWorks', 'Exportar a EndNoteWeb', and 'Exportar a EndNote'. To the right, a 'Citar' (Cite) window is open, showing citation templates for three styles: APA, Chicago Style, and MLA. Each style includes the author names (Abdallah-Pretceille, M., & Porcher, L.) and the title of the work (*Education et communication interculturelle*).

Exportar Registro * Citar

Exportar a RefWorks
Exportar a EndNoteWeb
Exportar a EndNote

Citar

Cita APA
Abdallah-Pretceille, M. , & Porcher, L. (1996). *Education et communication interculturelle* (1a. ed.). Paris: Presses Universitaires de France.

Citación Chicago Style
Abdallah-Pretceille, Martine. , and Louis Porcher. *Education Et Communication Interculturelle*. 1a. ed. Paris: Presses Universitaires de France, 1996.

Cita MLA
Abdallah-Pretceille, Martine. , and Louis Porcher. *Education Et Communication Interculturelle*. 1a. ed. Paris: Presses Universitaires de France, 1996.

Precaución: Estas citas no son 100% exactas.

Facetas en la home

- La primera vista puede surgir de un filtro.
- Se pueden colocar facetas en la home desde el archivo de configuración facets.ini.
- Genera una vista inicial del contenido del índice

Revisar por Biblioteca

Biblioteca Central
Biblioteca Instituto Balseiro
Departamento de Aplicación
Docente
Fac. de Artes y Diseño
Fac. de Ciencias Agrarias
Fac. de Ciencias Médicas
Fac. de Derecho
Fac. de Filosofía y Letras
Fac. de Ingeniería
Fac. de Odontología
Más opciones...

Revisar por Colecciones

Artículos de revistas UNCuyo
Biblioteca Inst. Balseiro
Centro de Documentación
Audiovisual
Itu Central
Libros - FCM
Libros-FFyL
Libros B. Central
Libros DAD
Libros de la FAD
Libros Digitales UNCuyo
Más opciones...

Revisar por Formato

Artículo en línea
CD-Rom
CD de música
COD
Diapositiva
Disquete
IMPRES
Impreso
Libro en línea
Mapas y planos tridimensionales
Más opciones...

Autocompletar en la caja de búsqueda

- Permite una selección rápida de registros basado en el texto ingresado.
- Funciona mediante el selector de campo de búsqueda.
- Muestra títulos, materias o autores dependiendo del campo seleccionado.
- Se configura mediante Searches.ini
-



The screenshot shows a search interface with a search box containing the text "borge". Below the search box, a dropdown menu displays several search results. The first result is "Evolución de la poesía **borg**eana." The second result is "Vaghe Eufone : temi **borg**esiani nella cultura contemporanea." The third result is "Padres nuestros que están en los cielos : **borg**esperón / Rodolfo Braceli." The fourth result is "Ficciones barrocas : una lectura de **Borges**, Bioy Casares, Silvina Ocampo, Cortázar, Onetti y Felisberto Hernández / Carlos Gamerro." To the right of the search box, there is a dropdown menu labeled "Todos los Campos" and a blue button labeled "Buscar". Below the search box, there is a section titled "Revisar por Colecciones" with a list of collections: "Artículos de revistas UNCuyo", "Biblioteca Inst. Balseiro", "Centro de Documentación", "Audiovisual", "Ru Central", "Libros - FCM", "Libros-FFyL", "Libros B. Central", "Libros DAD", "Libros de la FAD", "Libros Digitales UNCuyo", and "Más opciones...". At the bottom right, there is a logo for "CAS" and the text ".uncuyo.edu.ar".

Información externa al índice

- Wikipedia para información de autores.
 - Gran módulo con grandes errores en los aciertos.
- Portadas de libros: varios proveedores, alto consumo de cpu del server. Redimensión, cache, búsqueda remota, espacio, etc. Utiliza el ISBN para identificar las portadas.
- Conexión con sistema de circulación (ILS), sistema de drivers para los más populares. No es complejo escribir uno propio.
- Previsualización mediante GoogleBooks.

5

Borges : la pasión de una cita sin fin.

por Block de Behar, Lisa

Publicado 1999



3LIOTECAS
cuyo.edu.ar

Procesos de ingesta

- Vufind provee scripts de ingesta de datos para los servicios más populares como archivos marc21, marcxml, cosecha OAI, importación de CSV.
- El motor SOLR mediante la utilización de librerías de programación permite incorporar registros en forma directa.
- Se requiere un conocimiento del SOLR para realizar ingesta directa.

Proceso adoptado en el sistema de la UNCuyo

- Registros originales en MARCXML
- Proceso de armado del registro para enviar mediante un lenguaje de formateo
 - Se normalizan los datos a enviar al motor.
 - Se agregan valores por defecto.
 - Se corrigen errores comunes de carga.
- Se envían los datos al motor desde la aplicación.
 - Se registran los rechazos.

Opciones desde el sistema de gestión

- Borrado de colecciones
 - Se pueden borrar las colecciones en el índice.
- Actualización de cada registro o por colección
 - Actualización en línea o por rangos de registros.
- Estado de las colecciones
 - Reporte del estado de las colecciones.
- Vista general del estado del índice solr



Estado de colecciones SOLR

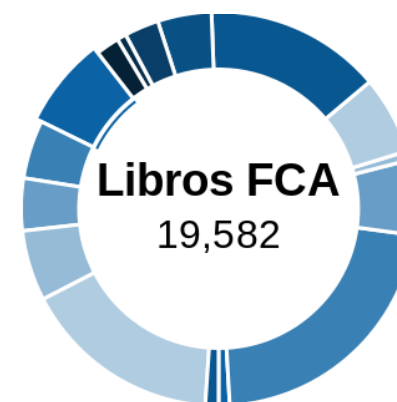
Tablas Gráficos

Registros en el explorador por colección

Libros Odontología	526
Libros-FFyL	63867
Libros de la FAD	16265
Libros DAD	55
Biblioteca Inst. Balseiro	17051
Libros FCPyS	41251
Libros - FCM	11091
Centro de Documentación Audiovisual	6285

Estado del OPAC

Bibliotecas	Registros en el opac
Libros-FFyL	63864 (de 63867) 
Libros B. Central	46923 (de 46958) 
Libros FCPyS	41249 (de 41251) 
Libros FCA	19581 (de 19582) 
Biblioteca Inst. Balseiro	17051 (de 17051) 



Proceso detallado de ingesta de registros MarcXml

- Registro original en formato marcxml
- Lectura con PHP conversión a objeto.
- Conversión a objeto SOLR mediante lenguaje de formateo. Transformaciones.
- Envío al SOLR y registro.

XML format=952y>Soporte
author=100ab
title=245abc|primero

Format:Libro
author:Borges, Jorge
Luis
Title:

Otros tipos de registros

- Cosecha de OAI/PMH:
 - Registros de la biblioteca digital de la universidad.
 - Realizamos la cosecha por SETs
 - No se realizan grandes modificaciones ni transformaciones mediante scripts.
- Cosecha de registros OJS:
 - Mediante protocolo OAI/PMH
 - Se realizan por título de revista.
 - Se incorporan como artículos de revista en línea.



Siguientes pasos en la implementación

- Habilitar los perfiles de usuario.
- Habilitar las opciones de “sociales”.
 - Compartir
 - Agregar etiquetas
 - Comentar
- Descartar los procesos no automáticos de ingesta
 - OJS → VUFIND.
 - Bdigital → VUFIND.
- Seleccionar e incorporar otros recursos de información:
 - Base de datos de enlaces
 - Archivo histórico
 - Base de datos de investigaciones (referencial) de la Universidad.

{ Contacto }

- bdigital@uncu.edu.ar
- hdegorgi@uncu.edu.ar

2DO. WORKSHOP NACIONAL_

28_SEPTIEMBRE 2016
LANÚS | BUENOS AIRES

Tecnologías e infraestructuras de la información para
Bibliotecas y Centros de Documentación e Información

