

Soporte de vocabularios controlados y autoridades en repositorios digitales

Autor: de Albuquerque, Pablo César
Directora: Dra. De Giusti, Marisa Raquel
Asesor técnico: Lic. Lira, Ariel Jorge



Objetivos

- Diseñar e implementar un sistema de gestión de autoridades y vocabularios controlados para repositorios digitales.

Objetivos secundarios

- Permitir la reutilización de vocabularios controlados
- Implementar un sistema genérico para gestión de vocabularios controlados en el ámbito del repositorio SEDICI
- Migrar sistema de autoridades en SEDICI
- Implementar un conector para DSpace que se comuniquen con el nuevo sistema.



Repositorio institucional

Infraestructura web cuya finalidad es:

- Almacenar, organizar, preservar y publicar la producción digital generada en una institución
- Describir a través de **metadatos** los recursos almacenados
- Interoperar con otros repositorios similares.

El origen de la información que describa al recurso depende del metadato



Vocabularios controlados en repositorios

Conjunto limitado de términos y reglas que forman parte de una estructura

- Mejora la búsqueda y descubrimiento de contenidos.
- Asisten a la usuarios:
 - búsqueda específicas de contenido
 - navegación de un sitio
- Asocian diversos sinónimos en un término controlado



SEDICI

Repositorio institucional central de la Universidad Nacional de La Plata

- Aloja más de 64000 recursos
- Desarrollado sobre DSpace
- Gestiona autoridades externas en CelsiusDL





Autoridades

Vocabulario controlado constituido por:

- Un registro
 - información asociada con una entidad
- Una clave
 - identificador unívoco
- Variantes
 - variaciones sintácticas de un registro

René Favaloro, se identifica con la URI

http://catalogo.bn.gov.ar/F/?func=direct&doc_number=000036235&local_base=BN A10

- Favaloro, René, 1923-2000
- Favaloro, René Gerónimo, 1923-2000
- Favaloro, René G. (René Gerónimo), 1923-2000



Control de autoridades

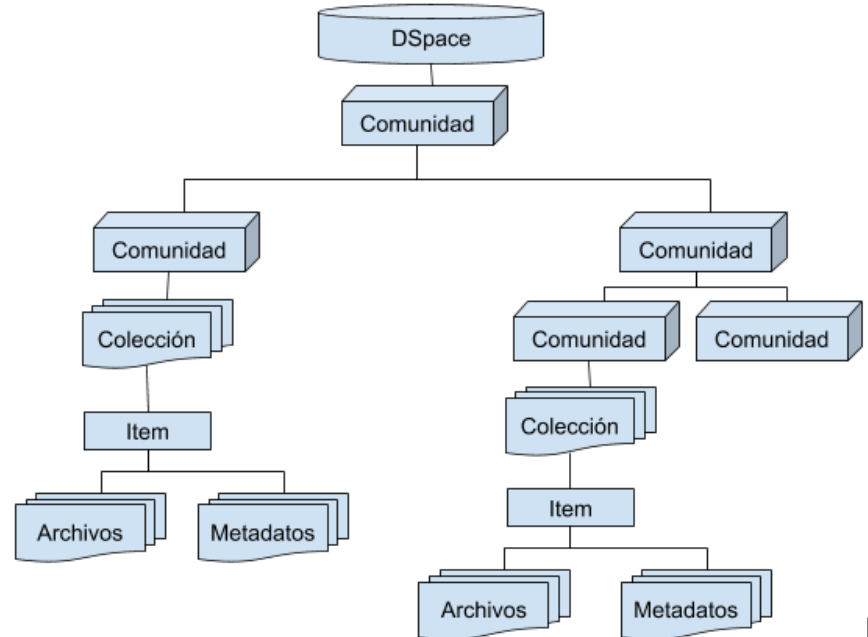
La práctica de mantener un catálogo autoridades

Este proceso tiene como objetivo:

- Probar que dos valores son idénticos
- Ayudar a completar metadatos con valores correctos.
- Mejorar la calidad de los metadatos
- Mejorar la interoperabilidad
- Reducir los tiempos de carga de metadatos

Modelo de datos de DSpace

Los repositorios basados en DSpace poseen una o más comunidades de nivel base que se organizan jerárquicamente en subcomunidades.





Uso de autoridades en DSpace

3 formas de controlar los valores

- value-pairs
- vocabularios controlados
- módulo de autoridades



DSpace - Módulo de autoridades

Utiliza una una arquitectura de plug-in

- Afecta todas las fases de la gestión de metadatos
- Aplicarse a cualquier metadato.

CelsiusDL módulo de autoridades que usaba SEDICI

El módulo de autoridades fue diseñado para controlar dos aspectos básicos llamados **Choice Management** y **Authority Control**.

Módulo de autoridades - Choice Management

Choice Management

Mecanismo que permite seleccionar los posibles valores de un metadato provenientes de una base de autoridades, a partir de un valor de consulta.

Autor Institucional:
Institución creadora del contenido intelectual de la obra

uni

Universidad de Salamanca (USAL)

AGH University of Science and Technology (AGH)

Benemérita Universidad Autónoma de Puebla (BUAP)

Cadi Ayyad University (UCA)

Centro de Investigación en Economía Política y Comunicación (CIEPYC)

Editorial de la Universidad Nacional de La Plata (EduLP)

Facultad de Periodismo y Comunicación Social

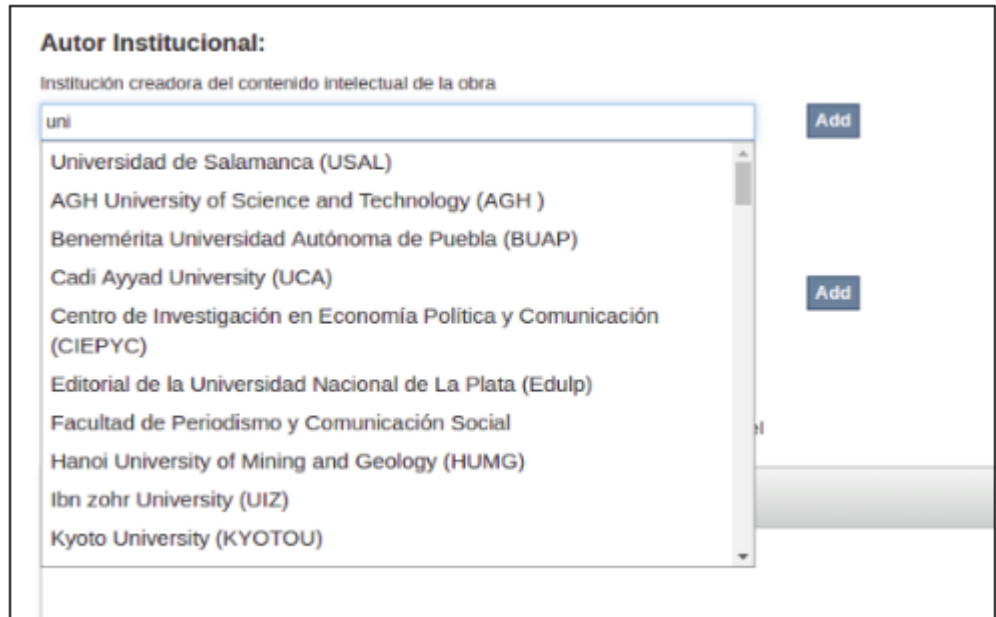
Hanoi University of Mining and Geology (HUMG)

Ibn zohr University (UIZ)

Kyoto University (KYOTOU)

Add

Add





Módulo de autoridades - Authority Control

Authority Control

Este mecanismo trabaja junto con el módulo Choice Management identificando una autoridad que contiene un valor elegido.

- confidence
- authority

<u>text_value</u>	authority	Confidence
Universidad Nacional de La Plata (UNLP)	http://ejemplo/auth/node/86555	600
Cañueto, Matías F.	http://ejemplo/auth/node/204702	600
Attribution 4.0 International (BY 4.0)	http://creativecommons.org/licenses/by/4.0/	600
UNLP	http://ejemplo/auth/node/86555	500
Cañueto, Matías G.		-1



Etapas de la solución

1. Relevar funcionamiento de CelsiusDL
2. Analizar y elegir herramientas de gestión de VC
3. Desarrollar nuevo módulo de gestión de autoridades
4. Normalizar entidades en CelsiusDL y exportarlas.
5. Importar entidades normalizadas en nuevo sistema
6. Desarrollar conector en SEDICI
7. Actualizar referencias a autoridades en SEDICI



Etapa 1 - Relevar estado de CelsiusDL

Autoridades gestionados en CelsiusDL.

- Desarrollada en Java en 2003
- Gestiona personas, instituciones, tesauros y sistemas de clasificación.

Tabla	Entidad representada
<u><i>jerarquias_termino</i></u>	Instituciones de Argentina Grado a alcanzar
<u><i>tesaruro_termino</i></u>	Tesoro de UNESCO Descriptores SeDICI Materias Tesoro de Eurovoc Tesoro DeCS



Etapa 1 - Relevar estado de CelsiusDL

Problemas de la solución actual

- Software antiguo, muy difícil de actualizar.
- Falta de mantenimiento y actualizaciones
- Problemas de diseño
- Redundancia de datos.
- Duplicación de información
- UI carece de chequeos de validez
- Problemas de integridad de datos
- Gran porcentaje de información almacenada que no es utilizada
- Ausencia de roles y permisos

Etapa 1 - Relevar estado de CelsiusDL

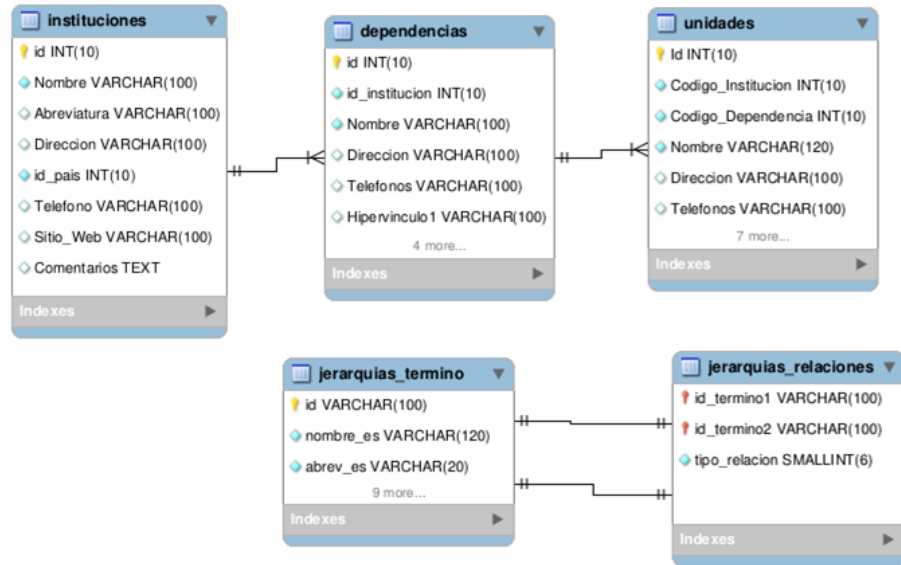
Entidades duplicadas

Institución de las personas representadas en las tablas:

- países
- instituciones
- dependencias
- unidades

Más Instituciones en tabla

- jerarquia_terminos





Etapa 2 - Análisis y elección de herramientas



VocBench



TEMATRES



Aspectos analizados

- Integración de vocabularios
- Interoperabilidad
- Casos de uso



Etapa 2 - Análisis y elección de herramientas

VocBench:

VocBench

- Usuarios deben tener conocimientos sobre OWL, SKOS(XL) y RDF
- El propósito y las funcionalidades de VocBench son demasiado específicas para SEDICI.
- No permite generar instancias de la información a migrar

TemaTres:

- Vocabularios gestionados como un tesoro y posee estados que no interesan en el contexto de nuestra aplicación.
- No permite representar entidades de todo tipo
- No dispone de algún módulo que permita importar los datos de CelsiusDL

TEMATRES 



Etapa 2 - Análisis y elección de herramientas

Drupal permite:

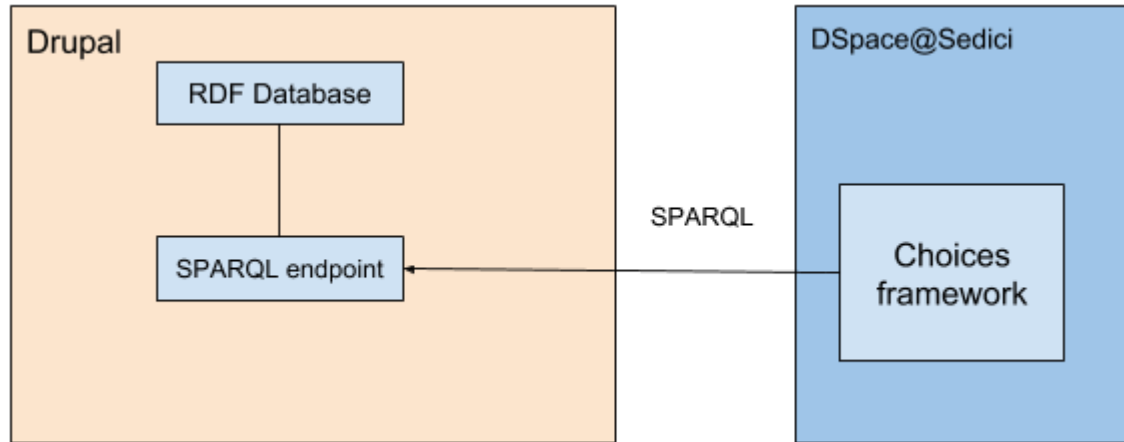
- Definir tipos de nodos con estructura dinámica
- Disponibilidad de módulos reutilizables
- Importar contenidos con sus relaciones a través de módulos contrib
- Exponer su contenido a otros sistemas a través de SPARQL
- Definir roles para administrador y editor



Etapa 2 - Análisis y elección de herramientas

Solución propuesta

- Almacenar autoridades como nodos en una instalación de Drupal 7
- URI del recurso identifica la autoridad
- Mapear nodos a RDF
- Exponer RDF a través de un endpoint SPARQL





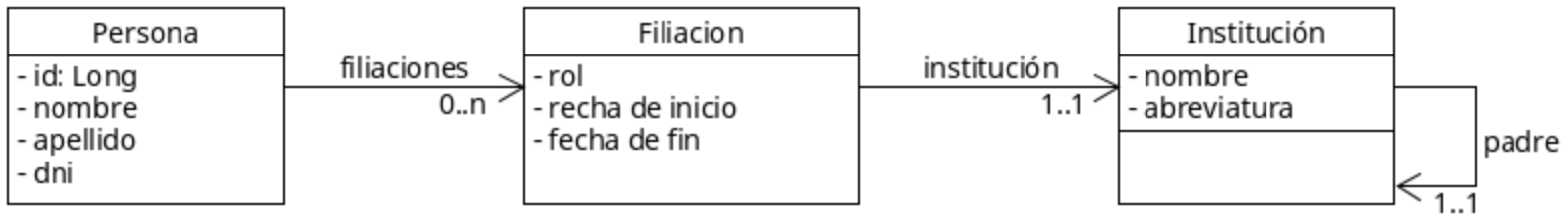
Etapa 3 - Desarrollo de nuevo módulo de gestión de Vocabularios Controlados

AuthVoc compuesto por:

- AuthCTypes
- Auth SPARQL
- Auth Importer

Módulo AuthCTypes

El modelo propuesto permite relacionar a una persona con una o más instituciones, registrando también la fecha de inicio, fecha de fin y rol que cumple en dicha institución.





AuthSPARQL

- Exponer los campos de las entidades Persona e Institución
- Utilizar los módulos contrib RDF indexer y SearchAPI

ARC SPARQL+ Endpoint (v2011-12-01)

[This interface](#) implements [SPARQL](#) and [SPARQL+](#) via [HTTP Bindings](#).

Enabled operations: select, construct, ask, describe

Max. number of results : 500

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX sioc: <http://rdfs.org/sioc/ns#>
SELECT ?institution ?label ?initials
WHERE {
  ?institution a foaf:Organization ; foaf:name ?label .
  OPTIONAL { ?institution sioc:id ?initials}
  FILTER(REGEX(?label, "universidad", "i") || REGEX(?initials, "universidad", "i"))
}
ORDER BY ASC(?label)
```

Change HTTP method: [GET](#) [POST](#)

Send Query

Reset



AuthSPARQL

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX sioc: <http://rdfs.org/sioc/ns#>
SELECT ?institution ?label ?initials
WHERE {
  ?institution a foaf:Organization ; foaf:name ?label .
  OPTIONAL { ?institution sioc:id ?initials}
  FILTER(REGEX(?label, "universidad", "i"))
}
ORDER BY ASC(?label)
```

```
<?xml version="1.0"?>
....

<binding name="institution">
  <uri>http://voc.sedici/node/46624</uri></binding>
<binding name="label">
<literal>Universidad Nacional de La Plata</literal>
  </binding>
  <binding name="initials">
    <literal>UNLP</literal>
  </binding>
</result>
</results>
</sparql>
```




Etapa 4 Normalización y exportación de entidades

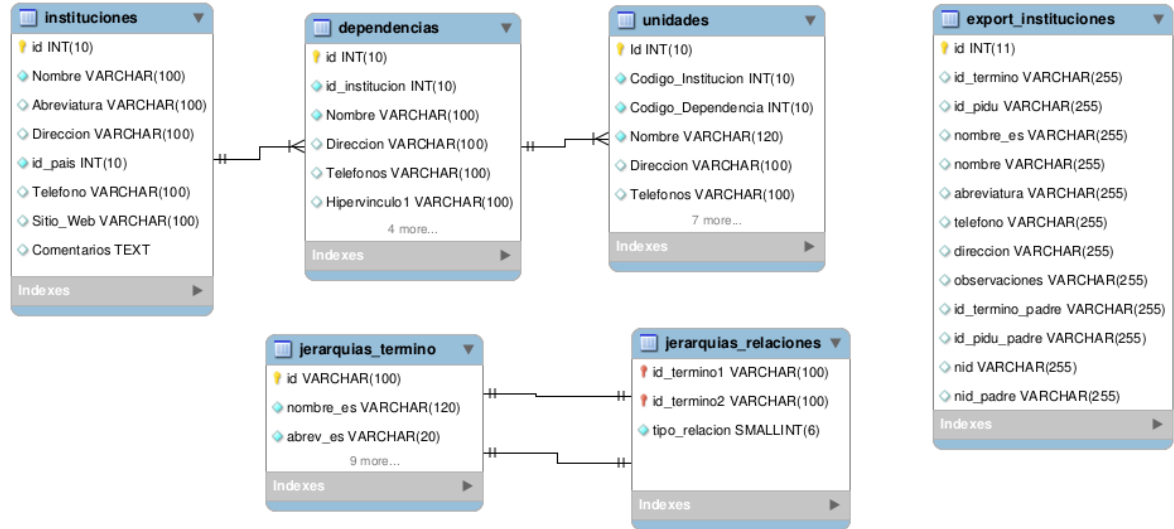
Sub-etapas:

1. Normalización de información en CelsiusDL
 - a. Unificación de instituciones.
 - b. Corrección de errores de sintaxis.
2. Depuración de autoridades innecesarias
3. Exportación de autoridades a formato CSV
4. Importación de autoridades en AuthVoc

Normalización

Se creó una tabla intermedia, que fue utilizada para unificar las instituciones.

Mientras se realizaba la unificación de instituciones se corrigieron los errores encontrados.





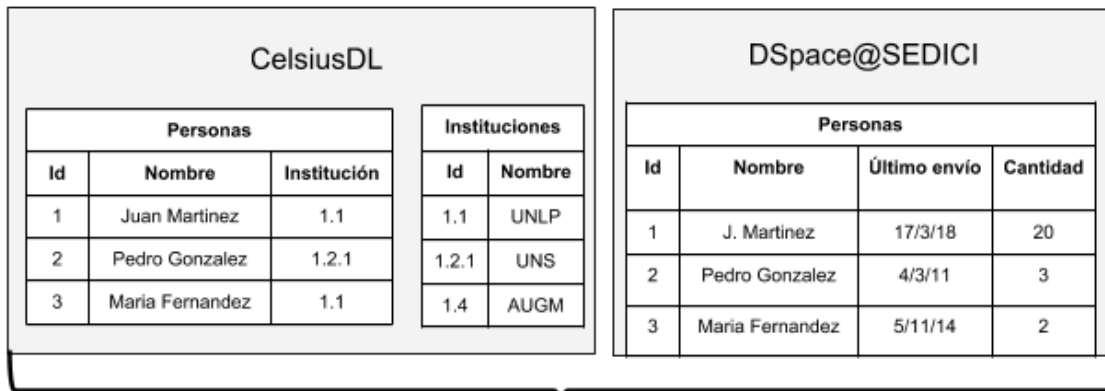
Normalización

Para definir la importancia de una institución se tuvo en cuenta:

1. País de origen de la institución
2. Si forma parte de la UNLP
3. Cantidad de publicaciones

Para definir el peso de un autor se hizo hincapié en:

1. Institución a la que pertenece
2. Cantidad de publicaciones
3. Fecha de su última publicación





Depuración de autoridades

Reglas de eliminación

Personas :

- Con sólo 1 ítem y no pertenecían a la UNLP.
- No son de Argentina y que tenían 2 ítems o menos y cuyo último haya sido 2017 o antes.
- No son de Argentina y que tenían participación en 3 o menos ítems, cuyo último envío fuese en 2013 o antes.
- No argentinos que pertenecen a instituciones que hayan hecho 10 o menos publicaciones.
- No pertenecientes a la UNLP con 1 solo ítem cuyo envío tenga fecha anterior al 2011.

Instituciones

- Argentinas que no pertenecían a la UNLP, CONICET y CIC con menos de 10 publicaciones



Depuración de autoridades

Se redujo en un 41 % el número autoridades

- 48436 a 20028 autoridades
- 46900 a 18600 autores
- 1536 a 1428 instituciones



Exportación en CSV

- Registros propios de cada entidad
- Claves de las autoridades en CelsiusDL

nid	old_id	apellido	nombre	dni	cargo	filiacion
48368	12649	González	Carlos		Becario	48024
48369	12650	Mármol	Juan			48024

nid	nid padre	nombre	abrev	id termino	id termino padre	id pidu	id pidu padre
46940		Universidad Nacional de La Plata	UNLP	1.1	1	1	1
46941		Comisión Nacional de Energía Atómica	CNEA	1.15	1	5	1

Etapa 5 - Importación

AuthImporter

- Importar desde un CSV la información normalizada del sistema de autoridades CelsiusDL.
- Basado en el módulo Feeds

Mapping for Node processor

Define which elements of a single item of a feed (= Sources) map to which content pieces in Drupal. A unique target means that a value for a target can only occur once. E. g. only one item with the UF

SOURCE	TARGET
+ nid	Node ID (nid)
+ nombre	Nombre (field_nombre)
+ apellido	Apellido (field_apellido)
+ dni	DNI (field_dni)
+ filiacion	-- (Entity reference by Entity ID) (field_filiaci_n:etid)
<input type="text"/>	- Select a target -
The name of source field.	The field that stores the data.

▶ LEGEND

Guardar



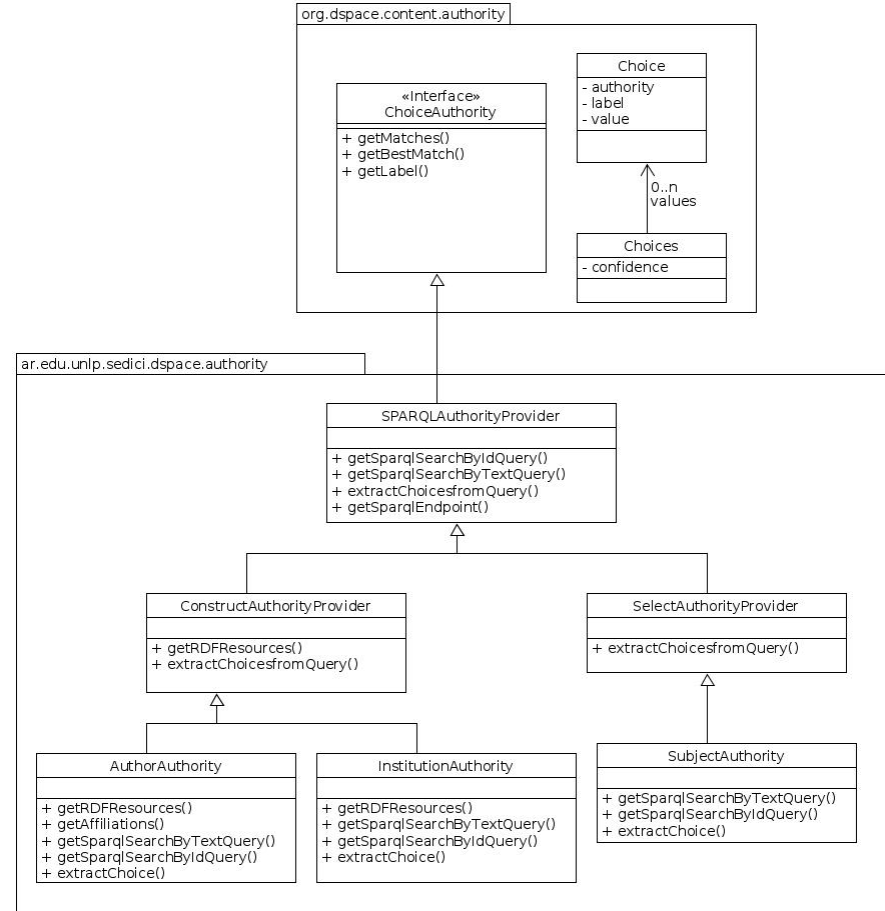
Etapa 6 - Desarrollo de Conector en DSpace

Clases provistas por DSpace

1. [org.dspace.content.authority.Choice:](#)
 - a. Representa un término controlado por una autoridad. Contiene los atributos authority, label, confidence y value.
2. [org.dspace.content.authority.Choices](#)
3. [org.dspace.content.authority.ChoiceAuthority:](#)
 - a. Esta interfaz, se define para tomar valores de una autoridad que puede ser provista de cualquier fuente, ya sea una base de datos local o no, un servicio prestado por un tercero, documentos en cualquier formato, algo que provea términos controlados.

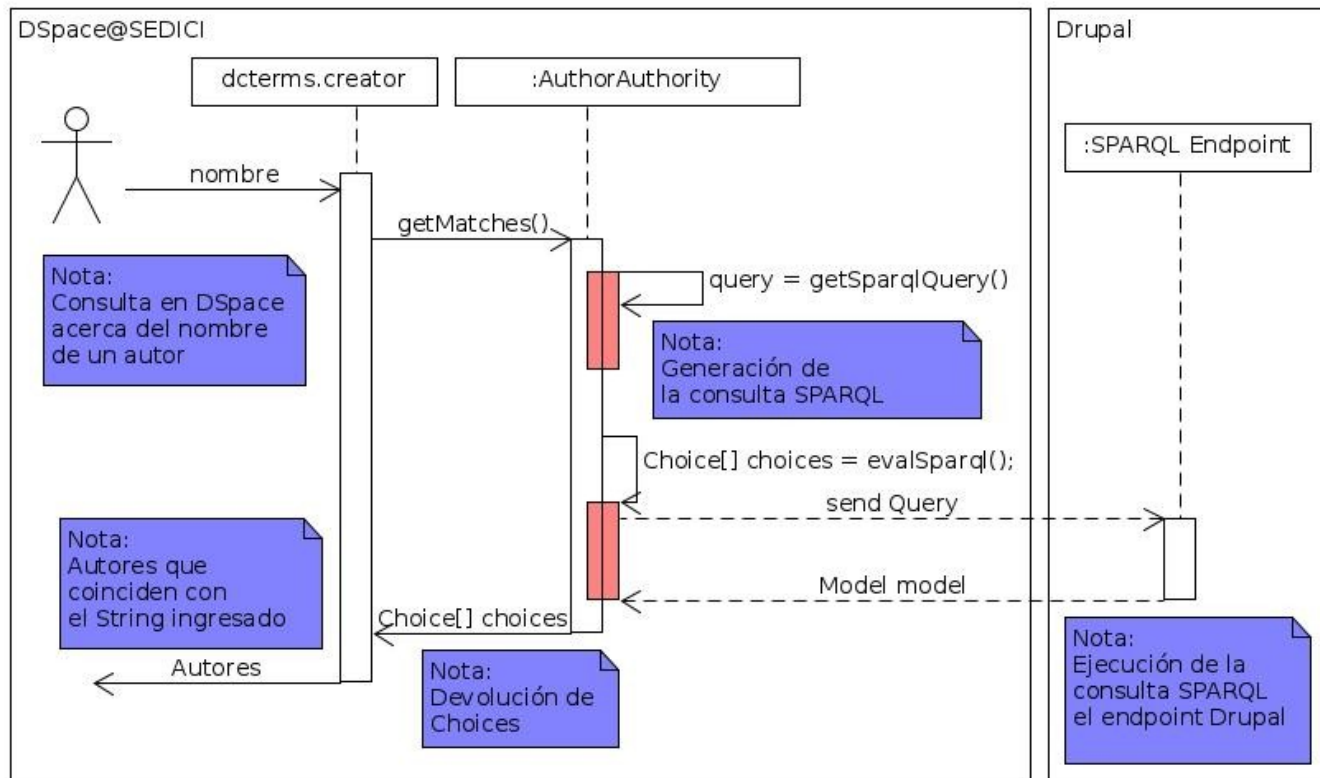
Conector en DSpace

- [SPARQLAuthorityProvider.java](#)
 - extiende ChoiceAuthority que sirve de base para definir fuentes de autoridades específicas.
- [AdvancedSPARQLAuthorityProvider.java](#)
 - Procesar consultas SPARQL con CONSTRUCT
- [SimpleSPARQLAuthorityProvider.java](#)
 - Clase que contiene la funcionalidad para procesar consultas SPARQL con SELECT
- [AuthorAuthority.java](#) e [InstitutionAuthority.java](#)



Conector en DSpace

Recuperación de un valor de autoridad



Resultados finales

GRADOS

- Facultad de Arquitectura y Urbanismo
 - Doctor en Arquitectura
 - Especialista en Arquitectura y Hábitat Sustentable
 - Especialista en Ciencias del Territorio
 - Especialista en Higiene y Seguridad Laboral en la Industria
 - Magister en Arquitectura y Hábitat Sustentable
 - Magister en Ciencias del Territorio
 - Magister en Conservación, Restauración e Intervención del
 - Magister en Paisaje, Medioambiente y Ciudad
- Facultad de Bellas Artes
 - Diseñador en Comunicación Visual
 - Doctor en Artes
 - Especialista en Lenguajes Artísticos
 - Licenciado en Artes Plásticas
 - Licenciado en Artes Plásticas (orientación Cerámica)
 - Licenciado en Artes Plásticas (orientación Dibujo)
 - Licenciado en Artes Plásticas (orientación Escenografía)

MATERIAS

- Arquitectura
- Bellas Artes
- Ciencias Agrarias
 - Ingeniería Agronómica
 - Ingeniería Forestal
- Ciencias Astronómicas
- Ciencias de la Educación
- Ciencias Económicas
 - Administración
 - Contabilidad
 - Cooperativismo
 - Economía
 - Turismo
- Ciencias Exactas
- Ciencias Informáticas
- Ciencias Jurídicas
- Ciencias Médicas
- Ciencias Naturales
- Ciencias Sociales
- Ciencias Veterinarias
- Comunicación Social
- Desarrollo Regional
- Humanidades
- Ingeniería
- Odontología
- Trabajo Social

Editar Persona Riesco, Daniel Eduardo

Apellido

Riesco

Nombre

Daniel Eduardo

DNI

MAIL

+

Añadir otro elemento

FILIACIÓN

Institucion

Departamento de Informática (48265)

Rol

FECHA DE INICIO

Format: 2018

FECHA DE FIN



Resultados finales

The quantum-classical transition as an information flow

Autores: Kowalski, Andrés | Martín, María Teresa | Zunino, Luciano José | Plastino, Ángel Luis | Casas, Montserrat

Tipo de documento: Artículo

Resumen

We investigate the classical limit of the semiclassical evolution with reference to a well-known model that represent given field. This is done by recourse to a special statistical quantifier called the "symbolic transfer entropy". W transition gets thereby described as the sign-reversal of the dominating direction of the information flow between clas

Información general

Fecha de publicación: 2010

Idioma del documento: Inglés

Revista: Entropy; vol. 12, no. 1

Institución de origen: Facultad de Ciencias Exactas

ISSN: 1099-4300

Extensión: p. 148-160

Materias: Ciencias Exactas ; Ingeniería ; Física

Palabras clave: Bandt and Pompe method; quantum chaos; semiclassical theories; transfer entropy

Refine su búsqueda

Autor

Riesco, Daniel Eduardo (45)

Montejano, German Antonio (36)

Pons, Claudia (36)

Fillottrani, Pablo Ruben (28)

Cechich, Alejandra (25)

Giandini, Roxana Silvia (25)

Daniele, Marcela (23)

Navegar por autor

Todos [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#)

O introducir las primeras letras:

Riesco, Daniel Eduardo [104]

Rimoldi, Marta Lidia [18]

Riesco, Gonzalo Ezequiel [1]

Rimoldi, Omar Jorge [5]

Rimorini, Laura [1]

Riestra, Dora [2]

Rinaldi, Carla Andrea [1]

Riesco, Daniel Eduardo [1]

Rinaldi, Emerson Valen



Conclusiones

1. Uso de CMS como base de herramienta de gestión de vocabularios
 - a. Drupal
 - b. Uso de módulos y personalización
 - c. Buenas prácticas
2. Proceso de migración
 - a. Relevamiento de CelsiusDL
 - b. Normalización
3. Acercamiento hacia el enfoque propuesto por los principios de Linked Data
4. Implementación de conectores en DSpace
5. Se redujo en un 41 % el número autoridades



Trabajo a futuro

- Migrar tesauros y la jerarquía de materias y grados alcanzados de CelsiusDL
- Implementar una distribución en Drupal
- Replicar módulo nuevo en CIC-DIGITAL
- Modelar y la gestionar variantes para las autoridades en Drupal
- Implementar en DSpace nuevos conectores que consuman información de otros sistemas



iGracias!

Preguntas

de Albuquerque Pablo César