GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
UNIVERSIDAD DE MURCIA
FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)
Unión Europea
Una manera de hacer Europa

OAI-PMH, Carga de Datos y descubrimiento

Álvaro Palacios
RIAM Intelearning Lab – GNOSS
alvaropalacios@gnoss.com

HĒRCULES

gnoss
Deusto
Universidad de Deusto
Deustuko Unibertsitatea
University of Deusto

Introducción

- ❑ Preparación.
- ❑ Gestión de repositorios.
- ❑ Descubrimiento. Detección de enlaces.
- ❑ Descubrimiento. Reconciliación con identificadores.
- ❑ Descubrimiento. Detección de enlaces + reconciliación con fuentes externas.
- ❑ Descubrimiento. Desambiguación con intervención del usuario administrador.
- ❑ Descubrimiento. Detección entidades CV.
- ❑ Descubrimiento. CV Propio.

Preparación

En este taller simularemos el proceso de gestión de repositorios OAI-PMH y realizaremos el descubrimiento sobre varios 'CV' simulados.

Para ello utilizaremos:

1. FrontEndCarga (https://herc-as-front-desa.atica.um.es/carga-web/public/gnossdeustobackend/home): Administración web que nos permitirá realizar la gestión de repositorios y la resolución de los problemas de desambiguacaión.
2. API de Carga (http://herc-as-front-desa.atica.um.es/carga/swagger/index.html): API para cargar los RDF de ejemplo.
3. Linked Data Server (http://graph.um.es/graph/sgi): Que nos permitirá ver el funcionamiento del descubrimiento sobre los RDF.

Para el uso del API de Carga utilizaremos el siguiente token:

Bearer
eyJhbGciOiJSUzI1NiIsImtpZCI6Imw1MF9BQ2haQkJGQnFrc0VqVWQ0RlEiLCJ0eXAiOiJhdCtqd3QifQ.eyJuYmYiOjE2MTQ4MDA0NDUsImV4cCI6MTYxNDg4Njg0NSwiaXNzIjoiaHR0cDovL2hlcmMtYXMtZnJvbnQtZGVzYS5hdGljYS51bS5lczo1MTA4IiwiYXVkIjoiYXBpQ2FyZ2EiLCJjbGllbnRfaWQiOiJXZWIiLCJzY29wZSI6WyJhcGlDYXJnYSJdfQ.DgMldsTE-0jmFfNUNEg-Bc2Wa4XWyZ0iaZzQ5zF025jCmvVuEjdZsxKUyVhFh7vZKAXRHkL-F7pvrexY_Elzj_nXU7EGLwBoJGNLTN5K2MzmQrr6p1bRnlE0AsL3yYfNXwN4bh7vQKaH6iGXXddA4QAvBGxkGCt_Bq_nbxlEBYUXUvIgjOyyZTy37h2QsD8tEQyYHfLmfYE1Ps1ml99OKLhtTQEp8JdU-a9CIHQ8J9KVYk7gkhMqyvNff9Y01lHOrrhNrjNLRXH3ouf98uJKhCiwrfCnUd2LynlDNvoyStH6zRvBkVZ6Zz6w_hwAKTOCI5YcZTYvA0RyaH5hvRQgHA

Preparación

En este taller en lugar de utilizar repositorios OAI-PMH, simularemos las cargas de los RDF para ver el proceso del descubrimiento.

Para ello, disponemos de 5 carpetas, numeradas del 1 al 5. En el interior de cada una de ellas nos encontraremos 7 versiones del mismo CV, con diferentes identificadores, listados con una letra, de la A a la G.

Cada uno de los participantes deberá trabajar con los CV correspondientes con una letra.

Gestión de los repositorios

Cada uno de los participantes dará de alta un repositorio con su nombre (la url es indiferente) en la página de administración de repositorios: https://herc-as-front-desa.atica.um.es/carga-web/RepositoryConfig

Una vez creado, pulsaremos sobre 'sincronizar' y se creará una tarea de sincronización, que fallará porque el repositorio no es válido.

Ahora procederemos a subir diferentes RDF a través del API de Carga, para ello accederemos a la URL http://herc-as-front-desa.atica.um.es/carga/swagger/index.html y utilizaremos el método /etl/data-publish con el jobId de nuestra tarea y discoveredProcessed=false

Descubrimiento. Detección de enlaces

Simula el CV de un investigador (Diego López de Ipiña) que contiene sus publicaciones y los coautores de sus publicaciones (Diego Casado-Mansilla).

En este caso se enriquecen los identificadores de los elementos encontrados en las fuentes externas de información, no se realiza reconciliación porque no hay datos cargados en la BBDD RDF

Descubrimiento. Reconciliación con identificadores.

Simula el CV de un investigador (Esteban Sota Leiva) que contiene sus publicaciones y los coautores de sus publicaciones (Diego Casado-Mansilla).

En este caso las publicaciones del RDF son inventadas por lo que no se detectan en las fuentes externas de información. Sin embargo el investigador Diego Casado-Mansilla tiene ORCID en el RDF y se realizará la reconciliación el investigador Diego Casado-Mansilla cargado en el RDF anterior porque está enriquecido en la BBDD RDF. Se puede ver como al Diego Casado que ya había cargado se le ha añadido el documento ‘Este es un documento de prueba con título inventado para probar más tarde’.

Descubrimiento. Detección de enlaces + reconciliación con fuentes externas.

Simula el CV de una investigadora (Oihane Gómez-Carmona) que contiene sus publicaciones y los coautores de sus publicaciones (Diego Casado-Mansilla).

En este caso se enriquecen los identificadores de los elementos encontrados en las fuentes externas. Se logra la reconciliación de Diego Casado porque aunque no tenga ninguna publicación en común con el Diego que está cargad en la BBDD RDF, hemos detectado que el investigador existe en fuentes externas con algún documento del RDF que estamos cargando y algún documento de los que ya estaban cargados en la BBDD. Se puede ver como al Diego Casado que ya había cargado se le ha añadido el documento ‘SmartWorkplace: A Privacy-based Fog Computing Approach to Boost Energy Efficiency and Wellness in Digital Workspaces.’.

Descubrimiento. Desambiguación con intervención del usuario administrador.

Simula el CV de un investigador (Álvaro Palacios) que contiene sus publicaciones y los coautores de sus publicaciones (Diego Casado-Mansilla).

En este caso las publicaciones del RDF son inventadas y no están cargadas previamente en el grafo ni se encontrarán en las fuentes externas, por lo que se producirá un problema de desambiguación porque ya hay una persona en el sistema que se llama Diego Casado-Mansilla.

Este error se producirá a todos, lo resolverá sólo uno y el resto le dará a reintentar. De esta forma ahora no se producirá el error porque ya detectará al usuario con la publicación en el RDF Store.

Descubrimiento. Detección entidades CV.

Simula el CV del investigador Diego Casado-Mansilla que contiene sus publicaciones y los coautores de sus publicaciones (todos los utilizados en los RDF anteriores).

En este caso las publicaciones y autores que ya estaban cargados se reconocen y se desambiguan automáticamente y las publicaciones nuevas se cargan.

Además, cabe destacar que en este caso 'Diego López de Ipiña’ se cita como 'D. López de Ipiña' y también es reconocido y renombrado.

Descubrimiento. CV Propio.

En este caso utilizaremos el fichero ‘CV Propio.xml’ para probarlo con los datos de los participantes. Si alguno tiene participantes puede probar a cambiar el nombre del CV por el suyo y añadir el título de varios documentos para probar.

En caso de que el participante no disponga de datos puede probar a buscar en ORCID https://orcid.org/orcid-search/search algún autor y añadirlo al CV

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
UNIVERSIDAD DE MURCIA
FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)
Una manera de hacer Europa
Unión Europea

GRACIAS

gnoss
Deusto
Universidad de Deusto
Deustuko Unibertsitatea
University of Deusto