1° CONGRESO INTERNACIONAL DE INGENIERÍA APLICADA A LA INNOVACIÓN Y EDUCACIÓN

ASAMBLEA GENERAL
ISTEC 2019
20, 21 y 22 de Noviembre 2019
Córdoba - República Argentina

INGENIERÍA APLICADA a la Innovación y Educación

Herramientas para obtener, mapear y filtrar recursos académicos desde repositorios digitales

Soloaga, Ignacio
Lira, Ariel Jorge
Villarreal, Gonzalo Luján
Vila, María Marta
De Giusti, Marisa Raquel

PREBI SEDICI
prebi.unlp.edu.ar sedici.unlp.edu.ar

UNIVERSIDAD NACIONAL DE LA PLATA

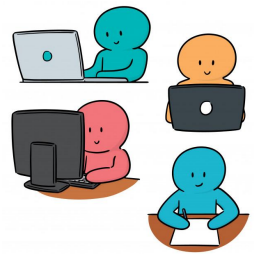
Introducción

SEDICI - Objetivos en relación al contenido

- Poblar continuamente el repositorio.
- Recuperar contenidos que deben estar por ley.
- Aumentar visibilidad del repositorio incorporando obras.
- Aumentar visibilidad e impacto de la UNLP

Vias de Ingesta

SEDICI
sedici.unlp.edu.ar



Múltiples formas de incorporar documentos al repositorio

- Autoarchivo
- Carga desde administración
- Interoperabilidad estable
  - SWORD
  - OAI-PMH

y también ...

- Importación: ingesta masiva de documentos

Importación masiva - Etapas

Etapa 1 Selección y obtención

Etapa 2 Limpieza y normalización

Etapa 3 Deduplicación

Etapa 4 Carga

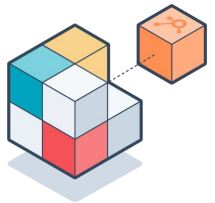
Etapa 1 | Etapa 2 | Etapa 3 | Etapa 4

Selección y obtención de documentos

- SCOPUS - API + Exportación
- Unpaywall - API
- Scholar - Web Scraping
- Repositorios - Harvesting sobre OAI-PMH

Harvester / Cosechador

- Herramienta que permite cosechar repositorios mediante OAI-PMH.
- Robusto ante errores.
- Modular.
- Permite programar cosechas.



Harvester (cont.)

- Módulos
  - Alta y planificación de cosechas
  - Recuperadores / cosechadores
  - Almacenadores

Etapa 1 | Etapa 2 | Etapa 3 | Etapa 4

Limpieza y normalización

- Mapeo de metadatos
  - Diferentes esquemas de metadatos
- Diferencias en la normalización de los datos
  - Estructuración de los datos

Crosswalk

- Herramienta que permite:
  - mapear metadatos entre distintos esquemas.
  - definir configuraciones.
  - aplicar filtros sobre los datos.

Deduplicación de documentos

- No debe subirse contenido ya cargado.
- Muy difícil detectar duplicados en grandes cantidades.
- Dificultades para desambiguar valores.
- Múltiples comparaciones.

Herramientas - Deduplicador

- Comparar grandes cantidades de documentos.
- Utiliza sistema de reglas.
- Comparaciones sintácticas sobre los metadatos.
- Establece porcentajes de similitud.

Deduplicador (cont.)

Permite:

- Encontrar documentos duplicados en repositorios.
- Definir nuevas reglas.
- Generar reportes de resultados.

Carga

- Gran cantidad de registros en una tanda.
- Comprimir en un archivo ZIP y en el formato que acepte el repositorio.
  - DSPACE (SEDICI) utiliza el formato SAF.

Resultados y Conclusiones

- Fortalecimiento del contenido del repositorio.
- Herramientas para la automatización de tareas.

- Ahora en **SEDICI**:
  - Importaciones desde SCOPUS
  - Más de 5000 documentos de autores UNLP de acceso abierto.

1° CONGRESO INTERNACIONAL DE INGENIERÍA APLICADA A LA INNOVACIÓN Y EDUCACIÓN

ASAMBLEA GENERAL
ISTEC 2019
20, 21 y 22
de Noviembre 2019
Córdoba - República Argentina

INGENIERÍA APLICADA
a la Innovación y Educación

¡Muchas gracias!

Consultas: {ignaciosoloaga, vilamm, alira, gonzalo, marisa.degiusti}

@sedici.unlp.edu.ar

PREBI SEDICI
prebi.unlp.edu.ar sedici.unlp.edu.ar

UNIVERSIDAD NACIONAL DE LA PLATA