

Trabajo Práctico Nº 4

Tema: Unidad 2 - Repositorios - Indexadores

Fecha Inicio: 13/05/2025 **Fecha de Entrega:** 27/05/2025

Actividades:

1. Consultar la documentación del protocolo OAI-PMH en el sitio [web oficial](#). Esto es para que al momento de emplear las consultas con los verbos en el punto siguiente se aclaren las dudas o se observe el modo de emplearlo, no es necesario la lectura y comprensión completa de dicho artículo, solo lo necesario para consultar con cada verbo.
2. Elegir un repositorio cualquiera que posea más de 100 registros. Consultar, averiguar, identificar la URL que implementa el protocolo OAI-PMH. Como dato, el sitio web [ROAR](#) ofrece un listado de repositorios que implementan el protocolo, uno puede buscar por país, por software de base o por tipo de repositorio, dentro del detalle del registro buscado aparece la URL identificada como (**OAI-PMH Interface**). Dicha URL es la utilizada para realizar las consultas.
3. El grupo de alumnos deberá realizar las consultas con los diferentes verbos, una por cada verbo (*Identify*, *GetRecord*, *ListMetadataFormats*, *ListIdentifiers*, *ListRecords*, *ListSets*) y hacer capturas de pantalla de cada uno de los resultados, indicando la URL utilizada. Por ejemplo, la consulta con el verbo **Identify** para el repositorio del SEDICI es <http://sedici.unlp.edu.ar/oai/request?verb=Identify>. Tener en cuenta que el formato de metadatos que usaremos es el *oai_dc*, con lo cual las consultas que requieran dicha indicación deben emplear el atributo **metadataPrefix**, ejemplo: http://sedici.unlp.edu.ar/oai/request?verb=ListIdentifiers&metadataPrefix=oai_dc
4. Diseñe una rutina en el lenguaje que desee y recolecte al menos 100 registros del repositorio CORA.Repositori de dades de recerca(<https://dataverse.csuc.cat/oai?verb=Identify>), con todos los metadatos y arme un archivo .csv con dichos registros. Adjuntar dicha rutina tanto en el informe como en el paquete a entregar. Deben almacenar los siguientes campos de metadatos: Título (Title), Resumen (Description), Autores(Creator), Palabras claves (Subject) y Fecha (Date).
5. Diseñe un indexador de documentos local, empleando la librería Lucene. La aplicación permitirá indexar una ubicación del disco rígido, previamente deberá definir el tipo de archivos a indexar (estos deben ser:.DOC, DOCX, PDF, .TXT, .HTML), luego se deberá emplear el índice generado realizando búsquedas textuales mostrando los resultados en una grilla o listado de fácil acceso. El índice puede almacenarse o generarse en memoria de acuerdo a la cantidad de documentos a indexar (mínimo 50 documentos de al menos 10 páginas, sino 100 documentos más pequeños).