



Aplicación de IA para el control de calidad de los metadatos en revistas y catálogos de bibliotecas

Experiencias surgidas en el Curso Virtual del Caicyt
“Mejorando la calidad de metadatos en revistas científicas con
herramientas IA” dictado en Agosto del 2025

Horacio Degiorgi

Director del SID

Adrián Mendez

Coordinador Biblioteca Digital



Revisión de metadatos en revistas científicas

Revisar los metadatos de los artículos científicos es un proceso clave para garantizar la calidad, trazabilidad y accesibilidad de la investigación.

Trazabilidad: Los metadatos permiten identificar de forma única un artículo (autoría, título, DOIs, fechas de publicación).

Rastreo de derechos y accesos: Los metadatos pueden incluir información sobre derechos de uso, licencias y restricciones de acceso (Open Access, embargo, etc.)

Interoperabilidad y reutilización: Permite la extracción automática de información para métricas, repositorios de acceso abierto y agregadores.

Calidad de datos: La revisión ayuda a detectar errores (autores mal escritos, afiliaciones incorrectas, fechas inconsistentes) y a corregir inconsistencias.



Revisión de metadatos en catálogos de bibliotecas

Revisar los metadatos de los registros de los catálogos, nos va a permitir:

Calidad de la información: evitar duplicados, errores de catalogación y inconsistencias entre registros, lo que mejora la confiabilidad del catálogo.

Recuperación y descubrimiento: metadatos bien gestionados permiten búsquedas más precisas y eficaces, facilitando que usuarios encuentren recursos relevantes.

Interoperabilidad: estándares y formatos de metadatos (por ejemplo, MARC, Dublin Core, RDA) facilitan la compartición de registros entre sistemas y bibliotecas.

Gestión del ciclo de vida: El control asegura que los metadatos se actualicen ante cambios en la edición, conservación digital, acceso y derechos, manteniendo la trazabilidad.



¿Qué aplicación usamos para esto?

Gemini, versión 2.5 Pro.

- ❖ Para trabajar armamos una Gema que se puede definir como “protocolo de comunicación y un ecosistema orientado a la publicación, la lectura y la interacción con contenidos en la web”.
- ❖ Las Gems suelen estar escritas en un formato de texto plano específico de Gemini llamado Gemtext.
- ❖ Este formato es ligero y está pensado para ser leído sin necesidad de popups, anuncios u otros elementos modernos de la web.



¿Cómo obtenemos los metadatos?

- ❖ Aplicación de recuperación de metadatos por OAI/PMH

Gestor de Revistas Electrónicas OAI/PMH

Añade revistas, cosecha registros y navega por la información.

Los metadatos obtenidos serán aquellos que soporte la revista:

- Dublin Core (pocos datos , sin granularidad)
- Marcxml (mayor granularidad, más registros completos)



Proceso





Podemos escribir prompt básico para extraer metadatos DC

Tengo que extraer de un archivo en formato PDF, que es un artículo científico de una revista digital, los metadatos respetando el esquema Dublin Core. Solamente me interesa los siguientes campos:

- Título
- Autor
- Materia
- Descripción: un resumen que no supere las 200 palabras.
- Fecha
- Identificador (si tuviera DOI)
- Idioma



Qué es una Gema ?

Características principales de las Gemas

- **Personalización:** Son completamente **personalizables**. Tú proporcionas las instrucciones detalladas (el *prompt*) que definen el rol que debe cumplir el Gem, el tono, estilo, o las limitaciones que debe tener.
- **Expertos en IA:** Pueden actuar como cualquier cosa que necesites: un *coach* profesional, un asistente de programación o un revisor de código.
- **Eficiencia para tareas repetitivas:** Son especialmente útiles para **tareas que se repiten** frecuentemente. En lugar de escribir las mismas instrucciones detalladas cada vez, la Gem ya las tiene guardadas.
- **Contexto Adicional:** Puedes subir **archivos** para darle a tu Gem contexto y recursos específicos que necesita para ayudarte mejor, como documentos o preferencias al momento de catalogar.
- Se pueden compartir, se pueden probar con distintos modelos. Se pueden replicar con pequeños cambios para distintos objetivos.



Descripción de la Gema

Soy curador de contenidos y tengo un archivo en formato PDF que contiene un artículo científico de una revista digital. Necesito que realices un análisis de metadatos siguiendo el esquema Dublin Core, extrayendo y estructurando la siguiente información:

- Título (Title): título principal del artículo.
- Creador (Creator): nombre(s) del autor o autores.
- Sujeto (Subject): palabras clave o tema principal del artículo.
- Descripción (Description): elaborar un resumen conciso del contenido del artículo, con una extensión máxima de 200 palabras.
- Fecha (Date): fecha de publicación.
- Identificador (Identifier): DOI, si está disponible.
- Idioma (Language): idioma principal del artículo.



Podemos escribir prompt para consultar calidad de un catálogo

Tengo un catálogo bibliográfico en MARC21 y necesito revisar la calidad de la los datos cargados en los campos

- Campo 100: Autor principal
- Campo 245: Título
- Campo 500: Notas generales
- Campo 650: Temas o descriptores
- Campo 700: Autores secundarios
- Campo 856: Localización electrónica



Descripción de la Gema

Sos un experto en catalogación, metadatos bibliográficos y MARCXML. Tu tarea es analizar un catálogo de una biblioteca, identificar sus componentes clave, evaluar su calidad y proponer mejoras. Te adjunto un documento de reglas de catalogación que usamos en la biblioteca

Campos elegidos:

Campo 100: Autor principal

Campo 245: Título

Campo 500: Notas generales

Campo 650: Temas o descriptores

Campo 700: Asiento secundario

Campo 856: Localización electrónica

En MARC, la calidad se mide por la consistencia en el uso de los subcampos y la precisión en los códigos de control:

Consistencia: ¿Se utilizan los subcampos correctamente (ej., \$a para el nombre, \$b para la inicial)?

Precisión: ¿Son los códigos de los indicadores (los dos dígitos que siguen al campo) los correctos? Por ejemplo, el indicador 1 del campo 245 (título) indica si el título es el principal.



Descripción de la Gema (cont.)

Detecta problemas comunes: El análisis debe identificar problemas específicos de los registros MARC. Algunos ejemplos son:

Uso incorrecto de subcampos: Por ejemplo, poner la fecha de publicación en un subcampo no designado para ello.

Datos duplicados: El mismo autor o tema repetido en diferentes campos.

Falta de información clave: Ausencia de campos importantes como el 856, que contiene el enlace al artículo.

Aporta soluciones y mejoras: Sugerí maneras de corregir o mejorar los registros MARC. Aconseja la validación de los datos con herramientas especializadas y la capacitación en estándares de catalogación como RDA



Problemas actuales derivados del uso de herramientas y modelos genéricos.

- Difícil de costear
- El resultado no siempre es el mismo: la cantidad de recursos asignados por las herramientas varía, la versión del modelo, el agente seleccionado por encarar el problema.
- La forma de armar los prompts cambia, por herramientas y versiones.
- Es difícil realizar un flujo de trabajo que sea reproducible.
- La “memoria” de los modelos no siempre ayuda.
- El costo de desarrollo de IA en local todavía es alto.



MUCHAS GRACIAS POR LA ATENCIÓN!!!

Contactos



/BIBLIOTECAS.UNCUYO



/BIBLIOTECASUNCUYO



/SIDUNCUYO

bdigital@uncuyo.edu.ar



UNCUYO
UNIVERSIDAD
NACIONAL DE CUYO

ACADÉMICA
SECRETARÍA
ACADÉMICA

Biblioteca Digital / UNCUYO



ecosistema | **SID**
UNIVERSIDAD NACIONAL DE CUYO