

Control bibliográfico impulsado por IA: automatización de la catalogación, la estandarización y la captura de datos

María Paz Rioseco

Coordinadora Gestión de Colecciones y Análisis de Información

Claudio Henriquez

Subdirector Sistemas de Gestión de Información y Tecnologías CRAI





1

Proyecto

- a. Catalogación asistida
- b. Corrección y normalización
- c. Aplicación móvil

2

Prompt

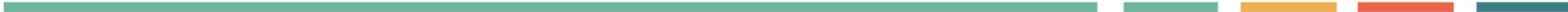
- a. Lo que funciona bien
- b. Lo que no funciona tan bien

3

Lecciones aprendidas

4

Conclusiones





Problema e inicio del proyecto

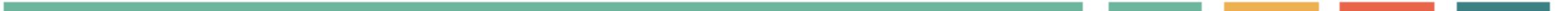
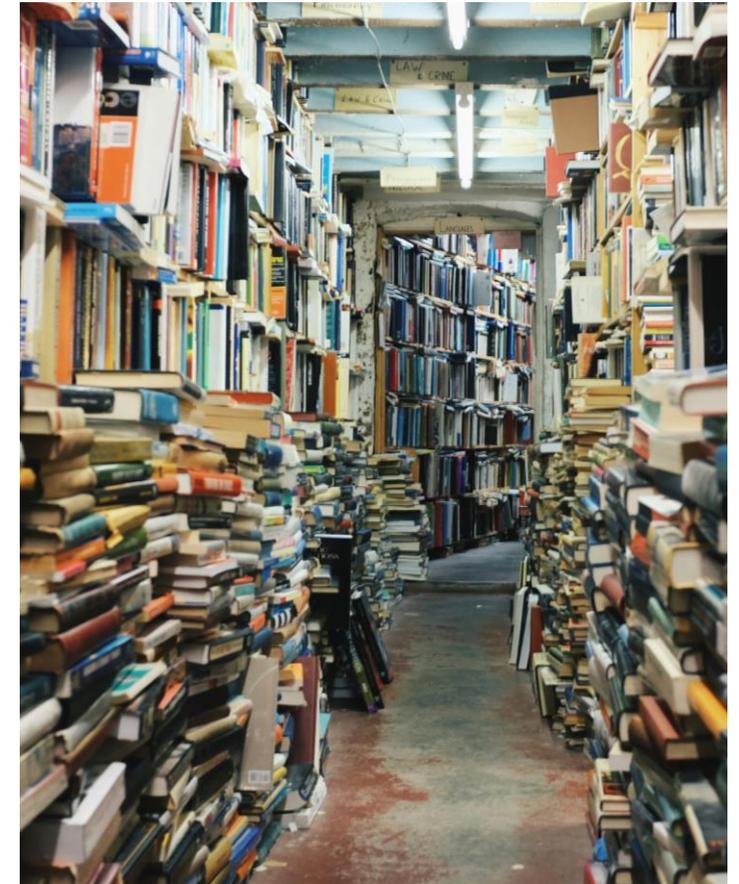
¿Por qué IA en la catalogación?

Situación inicial:

- Atraso significativo en la catalogación de libros y tesis (2.500 libros y 2.500 tesis).
- Falta de personal especializado para cubrir todo el volumen de trabajo.

Objetivo principal:

- Automatizar la catalogación para mejorar la productividad y reducir el tiempo de trabajo manual.
- Asegurar la calidad y la consistencia de los registros bibliográficos.





Flujo de Trabajo: De la Detección del ISBN a la Generación de MARC21

Proceso General



Búsqueda de metadatos
(Google Books API, otras fuentes)



Normalización y generación de campos (OpenAI API)



Validación y edición por parte del catalogador



Creación de registro MARC21



Envío al Sistema de Gestión Bibliotecaria (Alma de Ex Libris)





Diseño del formulario web e integración con el equipo de catalogación

Estructura del Formulario

Campos principales:

- ISBN (libros) o archivo PDF (tesis)
- Metadatos (título, autor, editorial, año, etc.)
- Campos fijos o semiautomáticos (008, 041, 336, 337, 338, etc.)
- Selectores para rol de autor (100\$a), indicación de si es traducción, si incluye índice, etc.

Validación y edición:

- El catalogador pueda revisar y corregir antes de enviar al sistema de gestión de biblioteca



Tecnologías clave utilizadas

Herramientas y Lenguajes

Lenguajes:

- Python (extracción de metadatos, conexión a APIs)
- PHP (backend del formulario web)

APIs y Modelos de IA:

- OpenAI API (generación de resúmenes, normalización de nombres, extracción de texto en imagen y lectura de PDF)
- Google Books API (obtención de información bibliográfica por ISBN)
- Ex Libris Alma API (creación de registros en el SIGB)

Procesamiento de PDFs:

- PyMuPDF (extracción de texto en PDFs digitales)
- Tesseract OCR (procesamiento de PDFs escaneados)



Con la tecnología adecuada, la escalabilidad se convierte en nuestro mejor aliado.





Prompts de Ejemplo y Resultados Obtenidos

PROMPT:

“Necesito que generes un registro MARC para el libro:

Título: Porth: Fundamentos de Fisiopatología

Autor: Carol Porth

ISBN: 9788417949723

Por favor, genera el formato MARC correspondiente para este libro. Asegúrate de que incluya campos como: ISBN, autor, título, edición, clasificación y editorial, entre otros detalles.”

¡Hola! Claro, con gusto te ayudo a generar el MARC para el libro "Porth: Fundamentos de Fisiopatología" con el ISBN 9788417949723.

Aquí tienes el formato MARC para este libro:

```
yaml Copiar
=LDR 01955cam a2200421 a 4500
=001 9788417949723
=003 ES-AL
=005 20250326000000.0
=008 210819s2021 sp a b 001 0 spa
=020 ## 9788417949723
=035 ## (OCoLC)1226715959
=040 ## ES-AL|bspa|cES-AL|dES-AL
=041 0# spa
=100 1# Porth, Carol,|eauthor.
=245 10 Fundamentos de fisiopatología /|cCarol Porth ; traducción, Laura Martínez Durán.
=250 ## 9a 9a ed.
=260 ## Madrid :|bElsevier,|c2021.
=300 ## 1 v. (varios pag.) :|bilustraciones ;|c29 cm.
=504 ## Incluye bibliografía e índice.
=650 #0 Fisiopatología|xTextos.
=650 #0 Fisiología humana|xTextos.
=655 #7 Manuales de consulta|2bdm
=710 2# Elsevier España,|eeditor.
=852 8# ES-AL|bCDU|c616.07
=999 ## Porth, Carol|cPorth: fundamentos de fisiopatología|d2021
```



Prompts de Ejemplo y Resultados Obtenidos

PROMPT:

“Necesito que generes un registro MARC para el libro:

ISBN: 9788417949723

Por favor, genera el formato MARC correspondiente para este libro. Asegúrate de que incluya campos como: ISBN, autor, título, edición, clasificación y editorial, entre otros detalles.”

Ejemplo de registro MARC:

swift

```
=LDR 02119nam a2200397 a 4500
=001 9788417949723
=005 20250326
=008 210625s2021 sp a b 000 0 spa d
=020 ## $a 9788417949723 $q (electrónico)
=040 ## $a ES $b spa $c ES $d ES
=041 0# $a spa
=100 1# $a Pérez, Juan, $e autor.
=245 10 $a El libro de los secretos / $c Juan Pérez.
=250 ## $a 1.ª ed.
=260 ## $a Madrid : $b Editorial Ficticia, $c 2021.
=300 ## $a 320 p. ; $c 21 cm.
=504 ## $a Incluye bibliografía e índice.
=650 #0 $a Literatura moderna $x Historia y crítica.
=650 #0 $a Filosofía $x Historia.
=653 #0 $a Secretos $x Filosofía.
=856 42 $u http://www.ejemplo.com/libro
=852 ## $a Biblioteca Central $b 12345 $c FIC
=999 ## $a LBD $b PRZ
```



Api de Google Books

La API de Google Books es un servicio web que permite consultar la extensa base de datos de Google Books de forma programática. A través de peticiones HTTP y respuestas en formato JSON, los desarrolladores pueden buscar libros por ISBN, título o autor, y obtener información clave como título, autores, editorial, fecha de publicación, descripción, entre otros metadatos.

<https://www.googleapis.com/books/v1/volumes?q=isbn:9788417949723>

```
https://www.googleapis.com/books/v1/volumes?q=isbn:9788417949723
Login Bibliotecas - UA Bibliografía Turnitin - Mis archiv... BestMARC CronoX - VL_Login
Impresión con sangría
{
  "kind": "books#volumes",
  "totalItems": 1,
  "items": [
    {
      "kind": "books#volume",
      "id": "7IlgzQEACAAJ",
      "etag": "vRFu7Yf0uX4",
      "selfLink": "https://www.googleapis.com/books/v1/volumes/7IlgzQEACAAJ",
      "volumeInfo": {
        "title": "Porth Fundamentos Fisiopatología 5",
        "authors": [
          "Tommie L Norris"
        ],
        "publisher": "LWW",
        "publishedDate": "2020-10-30",
        "description": "Porth. Fundamentos de fisiopatología cubre de forma concisa y completa los mecanismos de funcionamiento del organismo, y pone al alcance del estudiante los conocimientos necesarios sobre la enfermedad, así como los fundamentos que médicos, farmacólogos u otros profesionales de la salud requieren para un óptimo ejercicio de su profesión. Preparada como un destilado de la 10.a edición de Porth. Fisiopatología, esta 5.a edición contiene los aspectos más importantes y actuales en el área de la fisiopatología. Se abordan los temas tal como se realiza en una exploración de la fisiopatología, al relacionar la función normal del organismo con los cambios fisiológicos que tienen lugar como resultado de la enfermedad. Todo lo anterior, con el apoyo de un magnífico diseño de las ilustraciones y el énfasis en facilitar la comprensión de la materia.",
        "industryIdentifiers": [
          {
            "type": "ISBN_10",
            "identifier": "8417949720"
          },
          {
            "type": "ISBN_13",
            "identifier": "9788417949723"
          }
        ],
        "readingModes": {
          "text": false,
          "image": false
        },
        "pageCount": 1208,
        "printType": "BOOK",
        "maturityRating": "NOT_MATURE",
        "allowAnonLogging": false,
        "contentVersion": "preview-1.0.0",
      }
    ]
  }
}
```



Prompt

```
define('PROMPT_RESUMEN', "Necesito un resumen de 75 palabras del libro: %s. Redáctalo como lo haría un experto en catalogación, resaltando de manera precisa la temática, estructura y elementos esenciales del contenido, sin incluir información adicional.");
```

```
define('PROMPT_AUTOR_NORM', "Por favor, extrae el nombre completo del primer autor del siguiente texto: %s. Una vez identificado, devuelve únicamente el nombre en el formato 'Apellido1, Nombre1' (primer apellido, seguido de una coma y el primer nombre)");
```

```
define('PROMPT_PALABRAS_CLAVE', "Eres un experto en catalogación bibliográfica. Dado el libro: %s, asigna las materias adecuadas para el campo 650 en un registro MARC21. Utiliza vocabularios controlados como el Tesauro de la Biblioteca del Congreso (LCSH) o el encabezamiento de materia adecuado en español.");
```



Consulta por ISBN

Tipo de recurso:

ISBN:

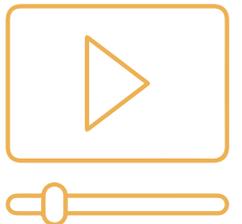
Consultar ISBN

Consulta por Título

Título del Libro:

Consultar Título

La educación es el arma más poderosa que puedes usar para cambiar el mundo. - Nelson Mandela





Prompts de Ejemplo y Resultados Obtenidos

PROMPT:

“Actúa como un experto catalogador. Analiza el PDF adjunto que corresponde a una tesis y extrae la metadata bibliográfica completa. Debes identificar y extraer los siguientes elementos:

- Título de la tesis
- Nombre del autor
- Nombre del director de tesis
- Institución y departamento (o facultad)
- Lugar y fecha de publicación
- Resumen y/o abstract
- Palabras clave”

A continuación se presenta la metadata bibliográfica extraída del PDF de la tesis:

Título de la tesis:

Análisis de las propiedades mecánicas y químicas de la arena de mar para su validación como un componente en la elaboración de hormigones en zonas costeras de la región del Maule

Nombre del autor:

Leonardo Pareja Arcos

Nombre del director de tesis:

Dr. Oscar Iván Candía Avello

Institución y departamento:

Facultad de Arquitectura, Construcción y Medio Ambiente – Carrera de Ingeniería en Construcción
(La portada no especifica el nombre completo de la institución, pero se deduce que se trata de una entidad académica ubicada en Talca.)

Lugar y fecha de publicación:

Talca, Chile – noviembre 2023

Resumen / Abstract:

El documento investiga la factibilidad de utilizar arena de playa, extraída de diversas costas de la región del Maule (Pelluhue, Constitución y Duao), como sustituto de la arena convencional en la fabricación de hormigón. Para ello, se realizaron ensayos que permitieron caracterizar tanto las propiedades mecánicas (especialmente la resistencia a compresión) como las químicas (contenido de cloruros, sulfatos, etc.) de la arena de mar. La tesis también aborda aspectos normativos, económicos y medioambientales, y propone recomendaciones para optimizar el uso de este recurso en construcciones costeras, donde la distancia a plantas de áridos convencionales encarece el costo de la materia prima.

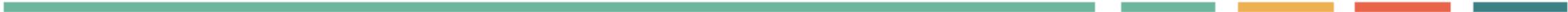
Palabras clave:

Hormigón; áridos; arena de mar; supramareal; densidad; granulometría; conductimetría; compresión; durabilidad



The screenshot shows the Marc-IA website interface. At the top left is the Marc-IA logo. The main header is blue with the text "INGRESO DE REGISTROS MÍNIMOS". Below the header is a navigation menu with "Inicio", "Libros", "Tesis", and "Contacto". The central content area features a white box titled "Subir Tesis en PDF" with a file selection field containing "tesis_ua.pdf" and a "Subir Tesis" button. The footer contains the copyright notice "© 2024 Marc-IA. Todos los derechos reservados."

The slide displays the thesis title and author information. At the top right is the logo of the Universidad Autónoma de Chile, Facultad de Arquitectura, Construcción y Medio Ambiente, Ingeniería en Construcción. The title is "ANÁLISIS DE LAS PROPIEDADES MECÁNICAS Y QUÍMICAS DE LA ARENA DE MAR PARA SU VALIDACIÓN COMO UN COMPONENTE EN LA ELABORACIÓN DE HORMIGONES EN ZONAS COSTERAS DE LA REGIÓN DEL MAULE". Below the title is "TESIS PARA OPTAR AL TÍTULO DE INGENIERO CONSTRUCTOR" and the author "LEONARDO PAREJA ARCOS". At the bottom, it lists "Dr. OSCAR IVÁN CANDIA AVELLO" and the date "Talca, noviembre 2023". A small dimension label "215,9 x 279,4 mm" is visible at the bottom left of the slide area.



MARC-IA Login

Usuario

Contraseña



Alma
Producción

Adquisiciones

Recursos

Descubrimiento

Servicios al
usuario

Admin

Analytics

Ocultar MDE

Configuración



Todos los títulos

Palabras clave

991000906



Biblioteca Providencia -
Mesón 1er Piso



Importar

Todos los títulos

Palabras clave Contiene 991000906195207491



Registros

Plantillas

Reglas

Guardar

Nuevo

Acciones de

Acciones de

Añadir inventario

Ver datos relacionados

Ordenar

Buscar y navegar

Enlaces rápidos

MARC21 Bib

Existencias

CLSAPUC

RVK

UADC

LCSH

LCNAMES

MESH

DC



Lista de filtros



No hay resultados para mostrar



Seleccionar una acción o un registro

Cargar registros de la comunidad





Lecciones aprendidas

Recomendaciones

1. Diseñar prompts cuidadosamente: Ser específico y claro en la instrucción.
2. Involucrar al equipo de catalogación desde el inicio: Su retroalimentación mejora la precisión de la IA.
3. Mantener un flujo de validación humana: La IA reduce el trabajo repetitivo, pero no reemplaza la experiencia bibliotecaria.
4. Iterar y mejorar constantemente: Ajustar prompts y flujos con base en los resultados.
5. Monitorear el desempeño: Uso de dashboards (ej. Power BI) para medir el impacto en la productividad y la calidad.





Lo que funciona bien (Éxitos)

Principales logros

Aumento de la velocidad de catalogación: Se pasó **de 15 libros diarios a 45 libros diarios** en picos de trabajo.

Normalización más uniforme: Menos variaciones en la forma de los nombres de autores, títulos, etc.

Generación de resúmenes: Facilita la indización y la recuperación de la información.

Reducción de errores tipográficos: Al delegar la estandarización a la IA, se minimizan errores de tipeo y capitalización.

Integración con Alma: Los registros se crean en modo “oculto” y luego se revisan, lo que mantiene el control de calidad.

300% *de incremento*





Lo que funciona bien (Éxitos)

Principales logros

Corrección de registros con IA.

- Para determinados registros se extraen de Alma a través de API son procesados por la IA y se corrigen para luego ser devueltos a Alma.
- Incorporación de las autoridades

Actúa como un experto en catalogación MARC21. A continuación, se presenta un registro MARC21 en formato MARCXML que necesita corrección. Realiza lo siguiente:

1) Capitalización de títulos, autores y materias:

- Corrige la capitalización de los nombres de autores, títulos y cualquier otro campo relevante (incluidas las materias).
- Los títulos deben seguir las convenciones de capitalización (por ejemplo, solo la primera letra de las palabras significativas en mayúscula y no todo en mayúscula).
- Los nombres de autores deben estar en formato adecuado, con la primera letra de cada nombre y apellido en mayúscula (por ejemplo, "Neisser, Ulric").
- Cualquier sigla, acrónimo, inicialismo o abreviatura debe conservarse en mayúsculas (p. ej.: "IBM", "UNESCO").

2) No cambiar ciertos campos:

- Mantén los campos 040, 090 y 999 en mayúsculas, sin alterar su contenido ni formato.
- Los códigos de idioma en el campo 041 deben permanecer en minúsculas (ej.: "spa").

3) Revisión de otros campos:

- Asegúrate de que campos como 245 (título), 500 (notas), 650 (temas) y otros campos de contenido textual sigan las convenciones de capitalización.
- ****No elimines**** ni modifiques la etiqueta 082, 245 o cualquier otra, excepto en los puntos que se indican a continuación.

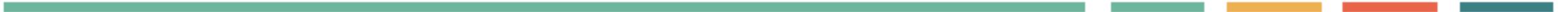
Autores:

Paulo Winterle

Autor sugerido IA:

Winterle, Paulo

No se agregó el campo Autoridades porque no se encontró coincidencia.





Lo que no funciona tan Bien (Desafíos)

Aspectos a mejorar:

- Clasificación Dewey
- Materias
- Resúmenes.

Dependencia de los costos de la API de ChatGPT:

- Si falla la API de OpenAI o Google Books, el proceso se detiene.
- Altos costos en modelos de mejor rendimiento

Calidad de los datos de origen:

- Google Books API a veces devuelve información incompleta o errónea.
- Se requiere verificación manual.





Lo que no funciona tan Bien (Desafíos)

Referencias bibliográficas:

- Extracción de bibliografías desde PDF-Word
- Sugerir bibliografías

A continuación se presenta una selección de 15 títulos en español (publicados o reeditados a partir de 2010) que abordan la "Psicopedagogía en Contexto", destacando la aplicación de teorías y metodologías psicopedagógicas en diversos escenarios y etapas del ciclo vital, así como el diagnóstico, la intervención y la evaluación en contextos educativos, comunitarios y organizacionales, con énfasis en la inclusión y la adaptación contextual.

Autor(es)	Título	Editorial	Edición	Año	ISBN
1. Coll, César & Palacios, Jesús (Coords.)	<i>Desarrollo psicológico y educación: Psicología de la educación escolar (Vol. 2)</i>	Alianza	8.ª	2014	978-84-206-8718-5
2. Bassedas, Eulàlia	<i>Aprender y enseñar en la escuela inclusiva</i>	Graó	3.ª	2011	978-84-7827-908-2
3. Monereo, Carles (Coord.)	<i>La práctica del asesoramiento psicopedagógico en contextos educativos</i>	Graó	2.ª	2019	978-84-9980-544-6
4. Marchesi, Álvaro & Coll, César (Coords.)	<i>Desarrollo, aprendizaje y enseñanza en contextos educativos</i>	Alianza	1.ª	2015	978-84-206-7895-4

The screenshot shows a Google search result for the book "Desarrollo psicológico y educación" by César Coll Salvador. The search bar contains the text "1. Coll, César & Palacios, Jesús (Coords.) Desarrollo psicológico y educación: Psicología de la...". The search results show the book title and author, along with a snippet of the text: "Tensiones y alternativas en psicología de la educación: las relaciones entre el conocimiento psicológico y la teoría y la práctica educativa...". The book is listed as 725 páginas. The search results also show the book's availability in PDF format from various sources like Scala Learning, website-editor.net, and Dialnet. A sidebar on the right shows the author's name "César Coll Salvador" and a "Personaje público" label, with buttons for "Noticias", "Imágenes", and "Videos".





Conclusiones Finales

Reflexiones

- Mejora continua de prompts: Ajustes finos para roles de autor, clasificación, resúmenes.
- Integración con más fuentes (ej. catálogos nacionales, bases de datos de tesis) para enriquecer metadatos.
- La IA es un aliado poderoso en la catalogación, permitiendo automatizar tareas repetitivas.
- La colaboración humano-IA es esencial: la supervisión y validación de los bibliotecarios asegura la calidad final.
- Es un proyecto en evolución, con áreas de mejora y grandes oportunidades para el futuro.

"Gracias a la IA, la barrera de lo difícil se ha desvanecido, y el horizonte de lo imposible está más cerca que nunca."



Control bibliográfico impulsado por IA: automatización de la catalogación, la estandarización y la captura de datos

María Paz Rioseco

Coordinadora Gestión de Colecciones y Análisis de Información

Claudio Henriquez

Subdirector Sistemas de Gestión de Información y Tecnologías CRAI

