

Tipo de artículo: Artículo original

Catpal: Herramienta para el manejo de metadatos de documentos del gestor bibliográfico Mendeley

Catpal: Mendeley bibliographic manager document metadata management tool

José Carlos Mendoza Sosa¹ , <https://orcid.org/0000-0003-2421-5260>,
Arturo Orellana García² , <https://orcid.org/0000-0002-3652-969X>

¹ Centro de Informática Médica (Grupo de Investigación de Informática Médica., Facultad 1, Universidad de las Ciencias Informáticas). josecms@estudiantes.uci.cu

² Centro de Informática Médica (Grupo de Investigación de Informática Médica., Facultad 1, Universidad de las Ciencias Informáticas). aorellana@uci.cu

* Autor para correspondencia: josecms@estudiantes.uci.cu

Resumen

La investigación científica es un proceso dinámico que se caracteriza por ser riguroso y por conducir a la adquisición de nuevos conocimientos. En el proceso de producir nuevos conocimientos se consume y se genera abundante documentación. Con el objetivo de combatir la infodemia y de lograr organización en la documentación bibliográfica el centro de neurociencias de Cuba y la universidad de ciencias informáticas UCI colaboran juntos para el desarrollo de una herramienta que permita lograr el objetivo con la menor interacción humana posible y reducir el tiempo que podría llevar depurar toda la información.

Palabras clave: Django, documentos, etiquetas, Mendeley, metadatos, Python

Abstract

Scientific research is a dynamic process characterized by being rigorous and by leading to the acquisition of new knowledge. In the process of producing new knowledge, abundant documentation is consumed and generated. In order to combat the infodemic and to achieve organization in bibliographic documentation, the Cuban Neuroscience Center(CNEURO) and the Computer Science University(UCI) collaborate together to develop a tool that allows achieving the objective with the least possible human interaction and reducing the time it could take to purge all the information.

Keywords: Python, metadata, Django, tags, Mendeley, documents.

Recibido: 20/12/2021

Aceptado: 18/03/2022

En línea: 01/06/2022



Esta obra está bajo una licencia *Creative Commons* de tipo **Atribución 4.0 Internacional**
(CC BY 4.0)

Introducción

Un gestor bibliográfico o gestor de referencias bibliográficas es software que recoge las referencias bibliográficas de las bases de datos de investigación (catálogos de bibliotecas, índices, bases de datos de revistas científicas, etc.) y permiten organizar las citas y la bibliografía para tus proyectos. Además ayudan en la citación de los trabajos de investigación permitiendo la creación automática de citas, bibliografía y notas al pie. Permitiendo una mejor gestión de la información y del tiempo, ayudar a evitar errores, y facilita la cita y la creación de las referencias en múltiples estilos.

A principios de la década de los ochenta surgieron los primeros gestores bibliográficos, creados como aplicaciones aisladas que podían ser instaladas en los ordenadores de los usuarios, con el objetivo de atender la necesidad de los investigadores con una base de datos personal de colecciones de documentos impresos (Silva Hernández et al., 2013).

Con la generalización de las revistas y recursos digitales accesible a través de Internet, el navegador se ha convertido en al herramienta natural para compilar información, por ello es lógico que los gestores de referencias hayan incorporado el acceso a través de Internet. Algunos de ellos disponen de versiones locales y otros no, también es distinta la manera en la que se integran en el navegador (Reyes Pérez et al., 2020), (Alarc et al., 2016).

Permite tanto el trabajo remoto a partir de una aplicación en la red, así como también la instalación local y el trabajo sin necesidad de una conexión a la web. Estos gestores unen a las funciones tradicionales las capacidades de las redes sociales para descubrir y compartir información bibliográfica. Algunos de ellos son: Mendeley, Zotero, Papers o Bibloscape.

Con el acelerado ritmo que tienen las tecnologías de la información, es claro que el proceso de informatización de la sociedad es clave para poder lograr nuevas metas en todos los ámbitos. En tiempos en que todo suele estar digitalizado es preciso poder almacenar la información adecuadamente y evitar caer en la infodemia.¹

La continua evolución del desarrollo científico y la presentación de trabajos de investigación es vital mantener organizada la bibliografía de los centros de investigación. Una forma estupenda de hacerlo es a través de gestores

¹ El término infodemia se refiere a un gran aumento del volumen de información relacionada con un tema en particular, que puede volverse exponencial en un período corto debido a un incidente concreto como la pandemia actual. En esta situación aparecen en escena la desinformación y los rumores, junto con la manipulación de la información con intenciones dudosas. En la era de la información, este fenómeno se amplifica mediante las redes sociales propagándose más lejos y más rápido, como un virus. (4)



bibliográficos, incluso algunos incluyen API's que permiten el desarrollo de aplicaciones y/o *plugins* para extender las capacidades del software (Carreño, 2014), (Alonso-Arévalo et al., 2012).

El Centro de Neurociencias de Cuba (CNEURO) forma parte de una alianza con la Universidad de Ciencias Informáticas (UCI) en la que ambos centros colaboran en pos de alcanzar beneficios mutuos. El Centro de Desarrollo CESIM de la UCI en intercambios con CNEURO ha comenzado proyectos que ayudarán en el proceso de informatización de CNEURO.

Las investigaciones que actualmente lleva el centro, se nutren de documentación previamente descargada y el centro además quisiera poder tener independencia tecnológica creando un repositorio local capaz de almacenar la bibliografía de manera que siempre esté accesible para sus investigadores. Teniendo esto en cuenta, este trabajo busca desarrollar una herramienta que emplee la API desarrollada por Mendeley con dos objetivos en mente:

1. Autcompletamiento de metadatos corrompidos, incompletos o ausentes.
2. Facilitar el proceso de organización, separando el proceso de calificación de los documentos del almacenamiento de los mismos.

Materiales y métodos

En la investigación como métodos científicos de investigación utilizados se destacan los siguientes:

Métodos teóricos:

Método hipotético-deductivo para la elaboración de la hipótesis central de la investigación y plantearse objetivos específicos a partir del problema concreto, para proponer nuevas líneas de trabajo a partir de los resultados parciales.

Método sistémico para lograr que los elementos que forman parte de la estrategia sean un todo que funcione de manera armónica. Método histórico-lógico y el dialéctico para el estudio crítico de los trabajos anteriores, para utilizar éstos como punto de referencia y comparación de los resultados alcanzados.

Métodos lógicos:

Método analítico-sintético al descomponer el problema de investigación en elementos por separado y profundizar en el estudio de cada uno de ellos, para luego sintetizarlos en la solución de la propuesta.

Métodos empíricos: Método experimental para comprobar la utilidad de los resultados obtenidos a partir de la estrategia definida. Método de la entrevista y la encuesta fueron vitales para el diagnóstico de la organización relacionado con las pruebas de software. Captar la información cualitativa y cuantitativa del fenómeno, conocer los criterios sobre la forma en que se organiza y se lleva a cabo la calidad y pruebas de software, así como las posibles soluciones que se proponen en la investigación. Para ello se entrevistaron y encuestaron personas involucradas en la producción de software en CNEURO.



En CNEURO se empleaba Python como lenguaje de desarrollo para las tareas similares a la que se propone esta investigación, teniendo en cuenta esto se trabajó con el SDK de Mendeley para Python, para mantener el desarrollo con un lenguaje que era empleado y tenía personal calificado dentro del centro y, así se obtenía un proceso de desarrollo compatible con las tecnologías ya empleadas en el centro. Dicho esto, como entorno de desarrollo se emplea el framework de desarrollo web basado en Python: Django, para el trabajo de versiones se usó Git, empleando tanto GitHub como el propio GitLab del centro y para la gestión de los datos el sistema de gestión de base de datos PostgreSQL.

Para los datos del estudio se utilizaron los repositorios de los investigadores del Centro de Neurociencias de Cuba que participaron como consultantes en la implementación.

Resultados y discusión

En CNEURO se trataba de lidiar con la situación epidemiológica y con el abundante volumen de documentación que se generaba relacionada con el COVID-19. Se comienza planteando la necesidad de tener organización de la documentación científica y de tener una documentación con la mayor cantidad de metadatos completos posibles.

En primera instancia se comienza el trabajo en añadir la funcionalidad de completamiento automático de los metadatos de los documentos almacenados en los repositorios locales accesibles por los investigadores del centro.

Se desarrolló un script en lenguaje Python y se completó una funcionalidad para obtener automáticamente los metadatos faltantes de los documentos descargados (Caluña & Capelo, 2019).

Un documento se presenta en el gestor bibliográfico Mendeley con su título, tipo de publicación, autor, fecha de publicación, DOI entre otros metadatos.

[Espacio para foto de la representación de un documento en Mendeley]

El trabajo con el SDK de Mendeley nos permite entrar en contacto con la API de Mendeley y poder a partir de esto generar nuestra solución. Dicha API funciona sobre el protocolo HTTP para la comunicación entre el cliente y el servidor, siendo posible ejecutar las funcionalidades necesarias empleando verbos de HTTP como GET y POST. Para ello es preciso primero registrar la aplicación con una cuenta de desarrollador para poder tener credenciales válidas para el intercambio de operaciones con el API.

La idea principal en el desarrollo de la nueva funcionalidad del script era obtener los metadatos de un documento determinado, de lo que el SDK de Mendeley denomina como *catalogs*. Generalmente en los catalogs se encuentra



información relacionado a un documento, y puede ser obtenida si se busca con cualquiera de sus identificadores, por ejemplos su DOI (**Fig.1**). Se aloja el script en una máquina virtual en los servidores del centro, donde se corría el script de manera periódica para asegurar que la información estuviera actualizada y que se completaran los metadatos de los nuevos documentos agregados al repositorio (**Fig. 2**).

```
def review(doc):  
  
    if doc.identifiers is None:  
        doc.identifiers = {}  
  
    identifier_changed = False  
  
    # doc_keywords = doc.keywords  
  
    for key in keys:  
        if not visited[key]:  
  
            cat = None  
  
            if key in doc.identifiers.keys():  
                if key == 'doi':  
                    try:  
                        cat =  
session.catalog.by_identifier(doi=doc.identifiers['doi'], view='all')
```

Figura 1. Ejemplo de cómo se obtiene el objeto catálogo (cat), y empleando el patrón Singleton se usa el mismo objeto para el resto del código.



```
# . . .
# modifying the Keywords
if cat.keywords is not None:
    # print(cat.keywords)
    if doc.keywords is None:
        doc.keywords = []
        doc.keywords = list(set(doc.keywords + cat.keywords))

# modifying abstract
if doc.abstract is None and cat.abstract is not None:
    doc.abstract = cat.abstract
    pass

# modifying source
# In Mendeley2 source => Journal
if doc.source is None and cat.source is not None:
    doc.source = cat.source
    pass

# modifying year
if doc.year is None and cat.year is not None:
    doc.year = cat.year
    pass

# . . .
# if there is no new information to review then save the data
doc.update(identifiers=doc.identifiers, keywords=doc.keywords,
abstract=doc.abstract, source=doc.source,
year=doc.year, volume=doc.volume, issue=doc.issue, pages=doc.pages)
return doc
```

Figura 2. En este otro fragmento de código se puede apreciar cómo se asignan los metadatos al documento.

Aunque el script cumplía con las expectativas del centro, aún no era posible una organización adecuada ni guiada por las orientaciones del centro, pues a pesar de que el gestor bibliográfico Mendeley permite la edición de documentos ya sea desde su plataforma web o de la versión de escritorio, esta edición no es restringida y cualquier usuario que tenga acceso a un documento puede editar todos los metadatos y no existía manera de controlar que usuario debería ser capaz de editar la información relacionada con el documento.

Es cuando se comienza el desarrollo de una herramienta con interfaz gráfica de usuario para la gestión efectiva de los metadatos de los documentos almacenados en los repositorios locales de Mendeley. Este desarrollo ya emplea Django para crear una herramienta web que fuera accesible por todos los investigadores. En las primeras iteraciones del desarrollo la aplicación se buscaba cumplir los requisitos funcionales que se recogieron durante las reuniones con los investigadores involucrados (Corchuelo-Rodríguez et al., 2018). En la aplicación debía ser posible obtener todos los documentos de un grupo de Mendeley determinado por el usuario, y poder modificar los metadatos de estos documentos (**Fig. 3**). Ya en esta herramienta era posible restringir los metadatos que serían editados y que usuario tendrían los permisos para hacerlo, una característica que le interesaba al centro.



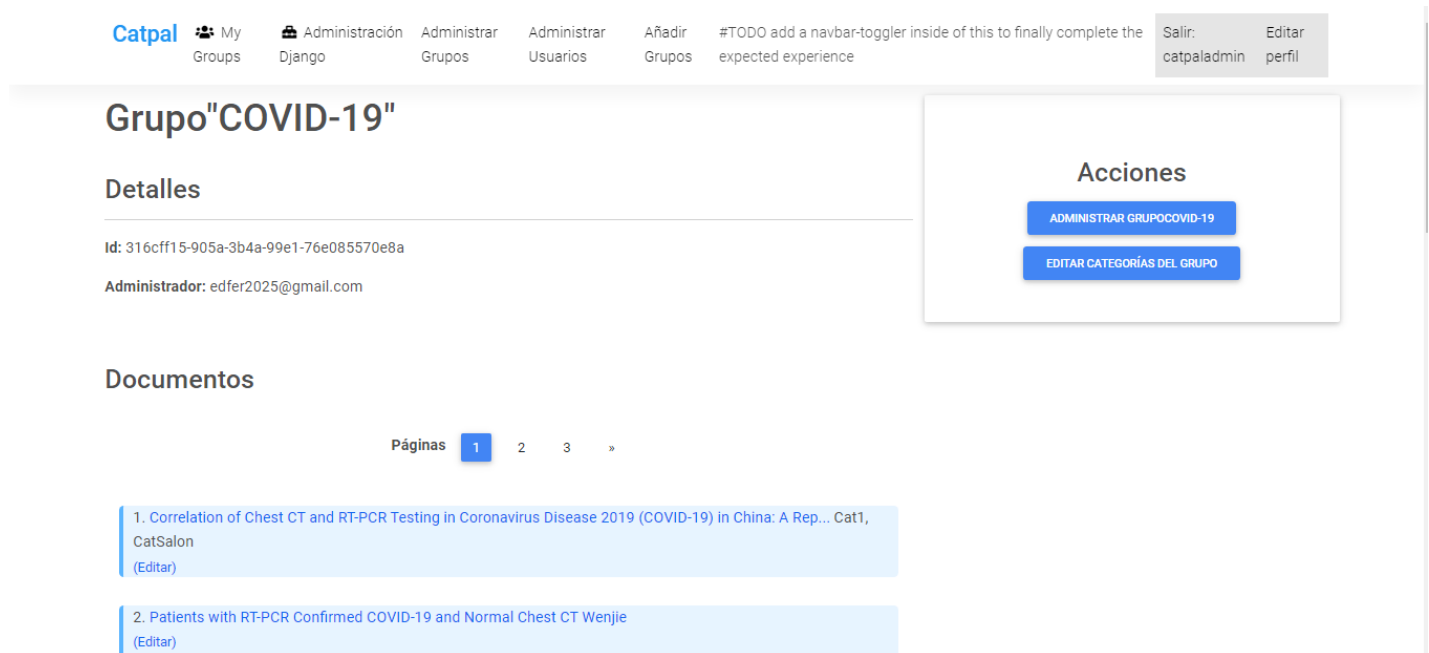


Figura 3. Vista de la página que muestra los documentos de un grupo.

En la tercera iteración del desarrollo se incorporan las categorías que podían ser asignadas a los documentos de un grupo determinadas y estas estaban estructuradas jerárquicamente (Fig. 4). Se añade además la capacidad de poder sincronizar los cambios hechos sobre cualquier documento con los repositorios online de Mendeley.

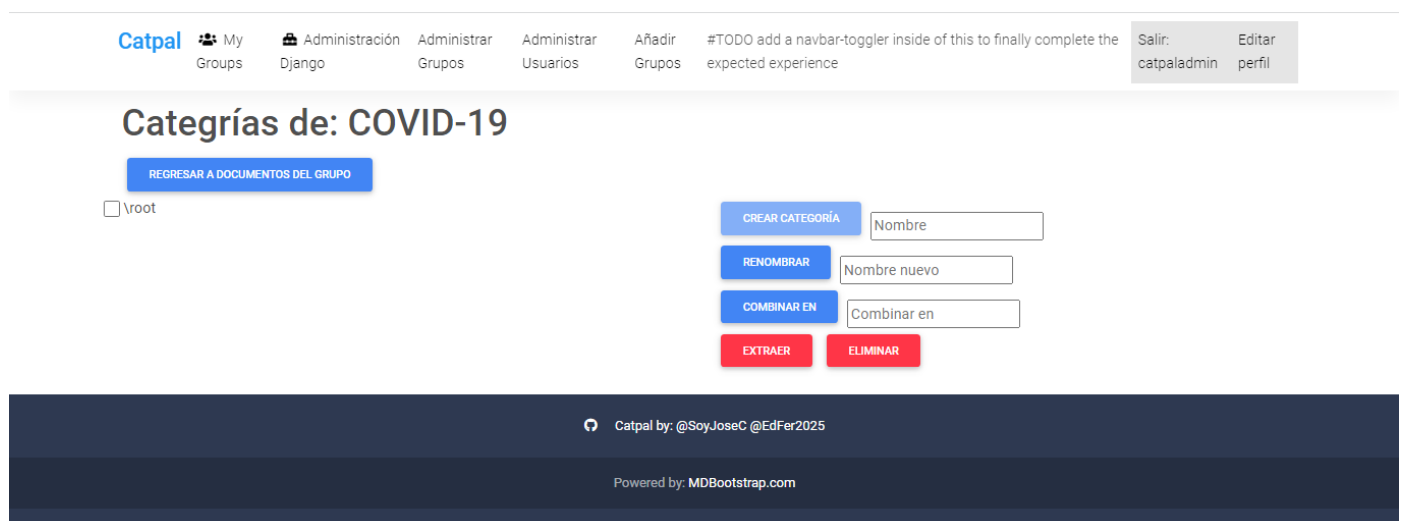


Figura 4. Página para agregar categorías a un grupo, en este caso el nombre del grupo es COVID-19



Conclusiones

El gestor bibliográfico Mendeley cuenta con flexibilidad para la creación de nuevas herramientas asociadas a este y mejorar su función.

La aplicación desarrollada en la presente investigación propicia el completamiento de metadatos de documentos, lo cual resuelve un problema existente para los usuarios de esta herramienta.

El Centro de Neurociencias de Cuba cuenta con una nueva herramienta capaz de facilitar el trabajo diario con la gestión documental afín a su quehacer investigativo.

Agradecimientos

A todos los que de alguna manera contribuyeron al desarrollo de esta investigación, los investigadores y trabajadores de CNEURO, los tutores y familiares, en especial a mi madre, Ania.

Referencias

- Alarc, R., Navas, A. B., Polo, W. M., Osorio, J. P., & Paola, S. (2016). Herramienta para el reconocimiento de referencias bibliográficas. *Investigación y desarrollo en TIC*, 7(1), 21-28.
<https://revistas.unisimon.edu.co/index.php/identific/article/download/2503/2404>
- Alonso-Arévalo, J., Cordón-García, J.-A., & Martín Rodero, H. (2012). Investigación 2.0 con gestores de referencias sociales: Mendeley y Biowizard.
<https://gredos.usal.es/bitstream/handle/10366/118456/GestoresSociales.pdf?sequence=1>
- Caluña, E. R. M., & Capelo, F. X. A. (2019). Utilización de Zotero como gestor bibliográfico web en estudiantes universitarios. *La Ciencia al Servicio de la Salud*, 10(Ed. Esp.), 397-405.
<http://revistas.esepoch.edu.ec/index.php/cssn/article/download/301/268>
- Carreño, M. L. (2014). Análisis comparativo de los gestores bibliográficos sociales Zotero, Docear y Mendeley: características y prestaciones. *Cuadernos de gestión de Información*, 4, 51-79.
<https://revistas.um.es/gesinfo/article/view/219511/171881>
- Corchuelo-Rodríguez, C.-A., Barreto Gallo, G. P., Cepeda González, J. D., Castro Marín, L. A., Segura Deaza, E. S., Rodríguez Gutiérrez, D. F., Fonseca Garavito, J. P., Montero, J. S., Flórez Chaves, M., & Díaz, Y. A. (2018). Propuesta de organización del conocimiento científico de la Universidad de La Salle en el Marco del SCTI.
<http://eprints.rclis.org/33395/1/Propuesta%20de%20organizacio%CC%81n%20del%20conocimiento%20cienti%CC%81fico%20de%20la%20Universidad%20de%20La%20Salle%20en%20el%20Marco%20del%20SCTI.pdf>



- Reyes Pérez, J. J., Cárdenas Zea, M. P., & Aguirre Pérez, R. (2020). Los gestores bibliográficos, una herramienta de apoyo al proceso investigativo en los estudiantes de agronomía. *Revista Universidad y Sociedad*, 12(1), 232-236. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2218-36202020000100232
- Silva Hernández, D., Rodríguez Felizola, M. C., & del Campo Peña, A. D. (2013). Las referencias bibliográficas en artículos originales y de revisión en revistas biomédicas cubanas. *Revista Cubana de Salud Pública*, 39. <https://www.scielosp.org/article/rcsp/2013.v39n1/o8/>

