



MARAFd

Un nuevo paradigma en
la gestión inteligente
de fondos documentales
digitales

OPEN SOURCE TECHNOLOGY

Open Source Project
Digital Transition

<https://www.safecreative.org/work/2502060809787?0>

 H.AI - R AI ASSISTED - REAL 2502060809787

MARAFd

Gestión inteligente de Fondos Documentales Digitales

**Guía de usuarios
y
administradores**

Revisión y data: **0.99e** – 14/05/2025

site oficial: <https://calentamientoglobalacelerado.net/MARAFd/>

[MARAFd. Gestión inteligente de Fondos Documentales Digitales](#) © 2025 by [Rafael Lomena Varo](#) is licensed under [Creative Commons Attribution-NonCommercial 4.0](#)

International 

Resumen: *MARAFd* puede definirse como un gestor de fondos digitales de segundo nivel que se sitúa entre el usuario y el servicio básico de hosting web, sin necesidad de sistemas intermedios de software ni la hiperreestructuración propia de los sistemas de bases de dato documentales, priorizando la preservación documental así como una accesibilidad en línea permanente e inmediata al cien por cien del contenido indizado y respetando los principios de procedencia y de orden como pilares fundamentales de la archivística.

Técnicamente, *MARAFd* consiste en una tecnología ecosistémica de código abierto diseñada para gestionar fondos documentales digitales mediante índice-tesauro automatizado y que aprovecha principalmente los recursos del sistema operativo para el indizado de archivos. Su arquitectura, basada en estándares abiertos como *PHP*, *Batch-script*, *Python*, *SFTP*, *JavaScript* y *HTML5*, elimina la dependencia de los complejos y costosos sistemas de bases de dato tradicionales, priorizando como decimos la preservación documental así como la accesibilidad, escalabilidad y, muy especialmente, la reducción máxima en cuanto a costes de implementación se refiere. El índice-tesauro, generado a partir de la propia estructura de carpetas y nombres de archivo, permite búsquedas instantáneas incluso en volúmenes de información del orden de varios TeraBytes y sin otro límite que el derivado del servicio de hospedaje contratado (*hosting-web*).

Junto a la posibilidad de coexistir sin conflictos con otro tipo de soluciones para la gestión documental digital, su carácter abierto y *Open Source* garantizan también una rápida y sencilla integración con herramientas en el ámbito de la inteligencia artificial y así lo demuestran casos reales ya funcionales como el *fondo documental sobre Melilla y el Rif* en el que el usuario puede ya acceder al contenido del fondo interactuando con un modelo entrenado *MyGPT®*.

La presente guía analiza en profundidad todos sus aspectos técnicos, sus ventajas propias y casos de aplicación real, así como la relevancia que su adopción puede suponer en la preservación del patrimonio digital en cualquier ámbito.

Palabras clave: *Gestión fondo documental, código abierto, tesauro automatizado, transición digital, escalabilidad, preservación digital.*

Abstract: *MARAFd* can be defined as a second-level digital fund manager that sits between the user and the basic web hosting service, without the need for intermediate software systems or the hyper-structuring typically associated with documentary database systems. It prioritizes document preservation as well as a permanent and immediate online accessibility to one hundred percent of the indexed content, while respecting the principles of provenance and order as fundamental pillars of archiving science.

Technically, *MARAFd* is an open-source ecosystemic technology designed to manage digital documentary collections through an automated thesaurus index. It primarily leverages the operating system's resources for file indexing. Its architecture, based on open standards such as *PHP*, *Batch-script*, *Python*, *SFTP*, *JavaScript*, and *HTML5*, eliminates dependence on complex and costly traditional database systems. Instead, it prioritizes document preservation, accessibility, scalability, and, most notably, the lowest possible implementation cost.

The thesaurus index, generated from the folder structure and file names, allows for instant searches even within volumes of information reaching several terabytes, with the only limit being the contracted hosting service.

Along with the ability to coexist seamlessly with other digital document management solutions, its open and *Open Source* nature also ensures fast and easy integration with artificial intelligence tools. This is demonstrated by real-world functional cases, such as the *documentary fund on Melilla and the Rif*, where users can already access the fund's content by interacting with a trained *MyGPT®* model.

This guide provides an in-depth analysis of all its technical aspects, its unique advantages and real-life use cases, as well as the potential impact its adoption has on the preservation of digital heritage in any field.

Keywords: *Documentary fund management, open-source, automated thesaurus, digital transition, scalability, digital preservation.*

1. Introducción

1.1. Objetivo de la guía

Esta guía intenta detallar todos los aspectos técnicos sobre la tecnología de código abierto **MARAFd** empleada en la gestión de fondos documentales digitales y está dirigida tanto a usuarios que deseen conocer todos sus entresijos para poder exprimirla al máximo (estudiantes, investigadores, archivistas, etc.) como a los administradores potenciales que requieran de una herramienta rápida, accesible y gratuita para la gestión de fondos y grandes repositorios de archivos en línea.

1.2. ¿Qué es **MARAFd** ?

La transición digital ha exacerbado la necesidad de soluciones innovadoras para gestionar un crecimiento exponencial de datos que puede acabar colapsando a muchas organizaciones.

MARAFd emerge como una respuesta disruptiva a este desafío que pretende situar al archivo o fondo documental temático como un centro de información integral, capaz de abarcar prácticamente cualquier área del conocimiento humano y científico y de abrir la puerta a ese poder infinito que el acceso a la información.

En este sentido, no es desacertado afirmar que **MARAFd** rompe los esquemas conceptuales de la archivística digital al combinar una simplicidad técnica inusitada con una eficacia contundente en la organización y accesibilidad a fondos documentales digitales.

En este sentido y conceptualmente **MARAFd** es una tecnología que nos permite:

- trabajar de forma segura y ágil, sin intermediación de software y bajo un control permanente del usuario, con miles de documentos, estructura de carpetas y archivos en línea,
- garantizar la preservación de repositorios completos al trabajarse siempre, al menos, con una copia exacta del repositorio completo en el ordenador local y otra en línea,

-
- la automatización de procesos como la generación del índice-tesauro (columna vertebral del fondo) y otros flujos de trabajo automatizados mediante sencillos y editables *batch-scripts* (ficheros **.cmd** ó **.bat**),
 - ofrecer mecanismos de búsqueda rápida y avanzada, con acceso inmediato a los contenidos indizados (rutinas *PHP* para búsquedas en servidor y otras tecnologías de rastreo local profundo dirigidas a la investigación),
 - una implementación y mantenimiento asumible prácticamente por cualquier usuario con un mínimo de práctica y sin ninguna formación técnica específica,
 - integrar y conectar nuestro fondo documental con herramientas abiertas de inteligencia artificial que faciliten el análisis e interpretación de datos (opción de accesibilidad al fondo mediante modelos abiertos de inteligencia artificial entrenados como *MyGPT*®)
 - una escalabilidad segura, rápida y sencilla a servidores de mayor capacidad. Al trabajar siempre sobre rutas relativas (path) de carpetas y ficheros y una sencilla infraestructura de *mirror local* (duplicado), la migración del fondo completo a otro servidor se reduce prácticamente a volcar el repositorio local sobre el nuevo servidor.

Y lo más importante, lograr todo ello:

- sin recurrir a complejos sistemas de software de base de datos,
- sin pagar ni una sola licencia privativa,
- sin necesidad de contratar personal técnico ni externalizar el servicio o crear dependencias con terceros
- sin costes estructurales de soporte ni de mantenimiento.

Para lograrlo, **MARAFd** utiliza principalmente recursos nativos del propio sistema operativo (ya sea *Microsoft Windows*® ó *Linux*) que automatizan el indizado de ficheros del fondo. Mediante un mismo **índice-tesauro** compartido entre equipo local y servidor web (*web server* ó *web hosting*) así como las tecnologías abiertas *HTML* (para un *frontend* personalizado) ó sencillos guiones en *PHP* (*backend*) para ofrecer mecanismos de búsqueda y la generación dinámica de hiperenlaces, un único administrador puede publicar en línea y en tiempo récord cualquier volumen imaginable de archivos de forma casi automática,

siendo únicamente necesaria la intervención del administrador para la transferencia o sincronización de carpetas entre el equipo local y el servidor.

Es fundamental comprender que es sobre el **índice-tesauro**, auténtico eje vertebrador del fondo, sobre el que se realizarán las búsquedas rápidas y a partir del cual también se generarán los hiperenlaces. Por ello, el tesauro es creado siempre de forma automática mediante la ejecución de un sencillo fichero **.bat** ó **.cmd**, y dicho tesauro, dada la análoga estructura de archivos y carpetas del fondo tanto en la copia local como en la remota (alojada en el servidor web), es compartido en ambos lados. Esta es la idea principal sobre la que podemos afirmar que pivota el funcionamiento básico del fondo.

Redundando en lo anteriormente descrito, podemos decir que su enfoque básico se centra en lo que viene a denominarse un *mirror local-servidor*.

También considero importante constatar que las tradicionales fichas bibliográficas y los sistemas convencionales de bases de datos brillan aquí por su ausencia. **MARAFd** prescinde así de cualquier tipo dependencia software y aprovecha como digo los recursos que el propio sistema operativo ofrece de forma nativa para mantener, perfectamente ordenado y bajo control absoluto, cualquier repositorio por grande que sea, sin otro límite que la capacidad de almacenamiento disponible.

En cuanto al **índice-tesauro** se refiere, podemos entenderlo como una suerte de índice “ampliado” que contiene metadatos o información embebida en el propio nombre del archivo y las carpetas, ofreciendo así accesibilidad inmediata a los ficheros a través del propio *path* y nombre completo de cada fichero. Dicho índice-tesauro, como ya hemos apuntado, es compartido en todo momento por el cliente y el servidor.

1.3. Marco conceptual y teórico de la tecnología y ámbito de aplicación

Este sencillo y productivo modelo en el que un único tesauro común “ampliado” vertebrata y gestiona ambos repositorios (local y remoto), podemos decir que redefine conceptualmente la estructuración y explotación de la información tal y como la conocemos, escapando, especialmente, a la complejidad y costes inherentes de los tradicionales sistemas dedicados a la gestión de fondos digitales.

MARAFd representa un paradigma alternativo en lo que a gestión documental digital se refiere que apuesta por la sencillez y el bajo coste de implementación sin renunciar al control absoluto y transparente de la información.

Por otro lado, su adopción es viable tanto en el ámbito de fondos documentales de archivos públicos como privados, pasando por PYMEs, instituciones públicas, equipos científicos y multidisciplinares, etc., ofreciendo un valioso equilibrio entre escalabilidad, preservación, accesibilidad y costes. Además, su carácter *Open Source* garantiza futuras mejoras y una fácil integración con nuevas herramientas (*toolkits*) y/o tecnologías (*ChatGPT*®, etc) que amplíen sus funcionalidades sin comprometer su esencia minimalista.

1.4. Uso de tecnologías estándares subyacentes abiertas

El principio *Open Source* de **MARAFd** nos lleva al uso obligado de herramientas libres y nos invita siempre a una coexistencia dinámica y eficiente entre diversas tecnologías abiertas.

Así pues, el ecosistema **MARAFd** hace un uso exclusivo de herramientas libres de desarrollo como son:

- **PHP:** lenguaje para la ejecución de guiones y procesos en el servidor tales como adaptación de hiperenlaces, gestión de búsquedas, navegación por carpetas (*backend*), etc. Los ficheros con código *PHP* son editables con cualquier editor *ASCII*. Recomendado *NotePad++*
- **Batch-script:** Con un potencial a veces subestimado, este recurso del sistema operativo es capaz, a través de ficheros de procesamiento por lotes, automatizar la generación y sincronización del tesoro entre el equipo local y remoto, además de otros muchos procesos. Los ficheros *batchscripts* son editables con cualquier editor *ASCII*. Recomendado *NotePad++*
- **HTML5/JavaScript:** Nos ofrece la posibilidad de implementar una web personalizada (*frontend*), adaptable y multiplataforma que posibilite el acceso total al fondo mediante cualquier navegador estándar y en cualquier dispositivo conectado a internet. Los ficheros con código *HTML/JavaScript* son editables con cualquier editor *ASCII*. Recomendado *NotePad++*
- **Conexión FTP:** Transferencia de archivos y/o sincronización automática de carpetas (*mirror*) mediante el uso de software cliente *FTP* de libre distribución tales como *FileZilla*,

WinSCP (recomendado), etc. Resumiendo, nos permite subir físicamente los ficheros desde el equipo local hasta el servidor.

- **Python:** utilizado para el desarrollo de sencillas aplicaciones *toolkits* para su ejecución en local, en nuestro caso lo utilizamos para la verificación del tesoro desde el equipo local (*localT.py*) y también para la adaptación del índice-tesoro desde un formato **.txt** al estándar **.json** (*tesauro_json.py*) para un posible uso o tratamiento por terceras aplicaciones. Los ficheros con código *Python* son editables con cualquier editor *ASCII*. Recomendado *NotePad++*
- **FreeBASIC (v1.08.1) GCC-9.3.0:** utilizado para compilar las *toolkits* *TIA.bas* y *CompactTIA.bas*, encargadas adaptan, normalizan y compactan el índice-tesoro original para facilitar su utilización por parte del modelo largo del lenguaje *MyGPT®*, optándose por este lenguaje y compilador por su alta velocidad de ejecución. Los ficheros con código *FreeBASIC* son editables con cualquier editor *ASCII*. Recomendado *NotePad++*
- **ChatGPT®-MyGPT®:** Entrenamiento de sistemas expertos basados en modelos largos de lenguaje. En el caso de **MARAFd** podemos realizar, siempre limitado por la capacidad del modelo, un procesamiento inteligente del índice-tesoro del fondo para lograr un análisis de contenidos mucho más flexible y con resultados cuando menos impactantes (recomiendo ver caso real funcionando en el *Fondo Documental Digital Melilla-Riff*). El entrenamiento de un *MyGPT®* (básicamente se trata de un *ChatGPT®* customizado) puede realizarse desde el propio navegador.

1.5. Estructura básica del ecosistema **MARAFd**

Pasamos a ver en detalle los distintos elementos que conforman el ecosistema de nuestra tecnología.

1.5.1. El *mirror* local o copia espejo. Preservación nativa

De forma resumida, podemos afirmar que la solución **MARAFd** se sustenta sobre un sistema de *mirror local* y, como tal, exige mantener una copia exacta y permanente del fondo documental digital completo en una unidad de disco del equipo local. Esta característica intrínseca de **MARAFd** implica ya una redundancia nativa que garantiza una rápida recuperación ante una pérdida accidental de datos. Obviamente, la estructura de *mirror local* impone que tanto el servidor destinado a hosting-web como el disco duro del equipo o sistema local, dispongan de una capacidad o espacio libre de almacenamiento sobrado para contener la totalidad del fondo documental.

1.5.2. El índice-tesauro. Columna vertebral del fondo.

El índice-tesauro es compartido tanto en el lado local como en el servidor y se genera a nivel del sistema operativo mediante la ejecución de un sencillo fichero *batch-script* actuando como auténtica columna vertebral del fondo. De él se extraen metadatos a partir del propio nombre de los archivos y su ruta que serán empleados por los mecanismos de búsqueda rápida y también para la generación de hiperenlaces a los documentos del fondo. Al compartir una estructura idéntica ambas copias del fondo, la local y la alojada en el servidor, el índice-tesauro, con las adaptaciones pertinentes, nos servirá para la publicación de cualquier fichero y su acceso en línea.

Aunque el tesauro consiste básicamente en un volcado de la estructura completa del fondo documental en un fichero de texto plano **.txt**, no debemos subestimar el potencial funcional que el tesauro puede ofrecer en lo que al ámbito de la explotación de la información se refiere.

Por ejemplo, en el siguiente nombre de apenas 65 caracteres, */PUBLICACIONES/2023_Informe_Ambiental_ONG_Greenpeace_Espana_ES.pdf* se codifica año, título, autor, idioma y tipo del documento, una información que nos permitirá localizar archivos mediante un sencillo motor de búsqueda *PHP*.

Además, a partir del tesauro y una sencilla rutina *PHP* podemos generar una página dinámica *HTML* clasificada por carpetas y con acceso directo a todos los ficheros indizados en el fondo, todo ello en tiempo real.

1.5.2.1 Restricciones técnicas del índice-tesauro

- Límite de **2000 caracteres** por nombre de archivo y ruta completa.
- Exclusión en el nombre de los ficheros (y ruta de directorios) de símbolos y **caracteres especiales** o **acentuados** (con tilde) para lograr cierta homogeneidad en la información y mayor eficacia en las búsquedas.

1.5.3. Flujo básico de trabajo de añadido de archivos

- **Renombrado** de archivo/s según lógica descriptiva siguiendo criterios preestablecidos y copia en carpeta local (principios de procedencia y de orden, etc).
- **Sincronización** con el servidor desde cliente *FTP* (mirror).
- **Indizado** automatizado mediante orden *batch-script* local (**indizar.cmd**) que genera y actualiza el tesauro en servidor. Tal y como veremos en detalle, este script de indizado se

aprovecha para ejecutar distintas herramientas (*toolkits*) que serán estudiadas en detalle en sección propia, y que aumentarán notablemente el potencial de nuestro fondo, validando por ejemplo la integridad del tesoro (*TIA.exe*), adaptando la estructura y formato de éste (*tesauro_json.py*) o incluso generando tesauros más ligeros (*compactIA.exe*) para poder ser procesados por agentes inteligentes, en nuestro caso *MyGPT*®.

1.6. Algunos beneficios destacables derivados de usar **MARAFd** frente a otros sistemas

Aunque a priori el uso de esta tecnología no es en absoluto incompatible con otras soluciones especializadas en la gestión documental de fondos digitales y en teoría podría ser posible que **MARAFd** coexistiera con otras tecnologías, es importante destacar en ella ciertas ventajas relevantes como pueden ser:

- **Reducción drástica de costes:** **MARAFd** elimina la compra de licencias costosas de software y hardware especializado así como costosos servicios externos y de mantenimiento, requiriendo, en principio, únicamente un servicio básico de *hosting-web* con posibilidad de correr código *PHP* y aceptar conexiones desde cliente por *SFTP*. Obviamente, tanto el equipo local como el espacio de hospedaje contratado deben contar con suficiente capacidad o espacio de almacenamiento para albergar el fondo documental completo.
- **Independencia y control interno pleno del fondo documental** por parte de la entidad o persona propietaria del fondo, sin costes añadidos ni la necesidad de externalizar servicios para la adaptación, gestión y mantenimiento del fondo digital.
- **Escalabilidad:** Casos como el *Fondo Mellila-Rif* (>100 GB, >9.000 archivos) han sido sometido a migración (de plan básico a plus del proveedor *IONOS*) de forma transparente y demuestran ya cierta solvencia para manejar volúmenes del orden de varios cientos de Gbs e incluso de varios TeraBytes de archivos sin sufrir degradación alguna en el rendimiento.
- **Integración con IA y otras tecnologías emergentes:** Su carácter abierto garantiza en principio una sencilla integración con modelos como *MyGPT*®, permitiendo al usuario llevar a cabo análisis avanzados sobre el fondo, potenciando investigaciones académicas y permitiendo prácticamente consultas semánticas a cualquier nivel. No obstante, debemos destacar que el uso e integración de agentes de inteligencia artificial en el fondo

documental es un servicio externalizado que, si bien puede no suponer ningún incremento de coste significativo para la entidad titular del fondo documental, debemos ser conscientes y cuidadosos con los resultados obtenidos a través de dicha tecnología.

- **Investigación:** La posibilidad del rastreo profundo en el interior de los documentos del fondo mediante búsquedas lógicas booleanas, si bien no es posible en línea dada su elevada carga computacional, sí puede realizarse en modo local sobre el equipo que alberga la copia del fondo, lo cual es sin duda una herramienta a tener en consideración en el ámbito investigativo.

1.6. Desventajas

- **Rigor en la nomenclatura de archivos:** exigible para lograr eficacia en las búsquedas a través del tesoro. **MARAFd** obliga mantener unos criterios mínimos de renombrado descriptivo de archivos para lograr eficacia en las consultas, lo cual puede limitar el indizado masivo de datos no estructurados y sin intervención humana.

- **Limitaciones en el alcance de las búsquedas:** aunque eficientes y bastante refinados (búsqueda fonética), los sistemas de búsqueda propios de **MARAFd** tales como la navegación por directorios o la búsqueda rápida (implementada en **finder.php**) no soportan consultas semánticas complejas.

La búsqueda semántica o búsquedas booleanas no es posible sin la integración de agentes de IA adicionales (*MyGPT®*, etc.) o software externo de terceros, necesario en el caso de los rastreos profundos en equipo local (*Acrobat Reader®*, etc.).

Otra alternativa en fase experimental y con cierto potencial pero también limitada, es el uso de tecnologías de motores de búsqueda externos tales como *Google®* ó *Bing®* atacados mediante sencillos comandos de búsquedas restrictivas del tipo:

```
site:calentamientoglobalacelerado.net/ebiblio_melilla_riff/
```

La principal limitación de esta última fórmula de búsqueda en el fondo es que dependeremos de cómo y cuándo se indizará el contenido del fondo documental en el buscador, algo que no podemos controlar.

- **Ausencia de fichas bibliográficas de los documentos:** La hiperestructuración propia de las bibliotecas y repositorios digitales que brilla aquí por su ausencia, si bien ofrece innegables ventajas en cuanto a sencillez de implementación y costes, puede ser un obstáculo en determinados casos para lograr el indizado de nuestro fondo en plataformas

externas regidas por rigurosos estándares, al carecer los documentos del fondo de una ficha bibliográfica básica.

MARAFd

2. Instalación y configuración.

2.1. Requisitos del sistema.

Si bien podemos afirmar que **MARAFd** se caracteriza por no exigir prácticamente requisitos de hardware, señalaremos muy brevemente las necesidades en este aspecto.

2.1.1. Hardware.

- Equipo local con hardware suficiente para correr con solvencia el sistema operativo *Microsoft Windows 10/11* y con una capacidad de almacenamiento en disco adecuado al tamaño total del fondo. Es recomendable la disponibilidad de una segunda unidad de almacenamiento, destinada exclusivamente al backup local del fondo documental completo. En este sentido, la escalabilidad necesaria para atender el crecimiento del fondo es absoluta y simple de implementar.

- En lo que al equipo *hosting-web* o servidor se refiere, es un servicio que deberá contratarse con un proveedor de internet y, además de ofrecer capacidad de almacenamiento suficiente para albergar el fondo documental completo, también deberá soportar la ejecución de páginas con código *PHP* y aceptar conexiones *FTP* con el servidor para poder mantener sincronizadas las copias del fondo entre el equipo local y el servidor contratado. Estos requisitos son por tanto los mínimos exigidos a la hora de contratar un *hosting-web* con cualquier *proveedor de servicios de internet* y son características muy básicas e incluidas siempre en este tipo de servicios, cuyo coste es muy bajo. A título meramente informativo, hago constar que el fondo documental digital Melilla-Riff se encuentra alojado en un servidor operado por el popular proveedor de servicios **IONOS** y, a día de hoy, (mayo-2025) el coste de alojamiento para soportar el *Fondo Documental Digital MELILLA-RIFF* completo apenas supone unos 15 euros mensuales. Dicho coste es orientativo y podría variar en función de nuestras necesidades.

2.1.2. Software.

El único requisito básico para la implementación consiste en un sistema operativo *Microsoft Windows® 10/11* instalado en el equipo local destinado a la administración del fondo. A partir de ahí, no se requiere licencia de software privativa alguna pues todo lo que utilizaremos serán soluciones gratuitas y de código abierto como:

-
- *VSCode*®, *Notepad++* , o incluso el *Bloc de Notas* incluido en *Microsoft Windows*®, etc.

El código fuente del programa **TIA.bas** (utilizado para la normalización y adaptación del índice-tesauro a formatos **.txt** ó **.csv** para lograr que sean algo más asimilables y legibles por el sistema experto que emplearemos para investigar el fondo y basado en *MyGPT*®) ha sido compilado a código ejecutable *x64* para *Microsoft Windows*® mediante:

- **Compilador GNU** para *FreeBASIC* (v1.08.1) *GCC-9.3.0*, también podría emplearse para ello cualquier otro lenguaje/compilador que genere código ejecutable lo suficientemente eficiente. En cualquier caso se trata de software libre y, si en cualquier momento necesitáramos editar, adaptar y volver a compilar cualquiera de estos archivos fuentes de tipo **.bas** podremos usar el compilador citado que está incluido en el fichero comprimido **ecosistema_MARAFd.zip**

Por otro lado, en lo que a tareas de transferencia al servidor y sincronización de carpetas se refiere, éstas pueden ser encomendadas con garantía a cualquier software libre de los muchos existentes:

- **Cientes FTP:** *Filezilla*, *WinSCP* (recomendado), etc.

El uso de algunas herramientas complementarias que se han desarrollado para tareas concretas como es el caso de **localT.py** , destinada a la carga, visualización y verificación del índice-tesauro en el equipo local, o **tesauro_json.py** que convierte el tesauro al formato estándar **.json** por si fuera necesario para un posible uso, requieren la instalación de software externo como como es el caso del intérprete *Python* , incluido igualmente en el paquete comprimido **ecosistema_MARAFd.zip**

En el caso de **localT.py** , nos permite verificar el correcto funcionamiento del tesauro desde el equipo local. Así podemos comprobar de forma manual los enlaces a los documentos del fondo o buscar incluso algún archivo concreto (tecla *F3*).

En relación con la instalación en nuestro sistema *Microsoft Windows 10/11*® del intérprete *Python* necesario para correr esta utilidad y al objeto de evitar posibles infecciones, incompatibilidades o problemas de otra índole, la descarga del instalador deberá hacerse siempre desde el servidor oficial:

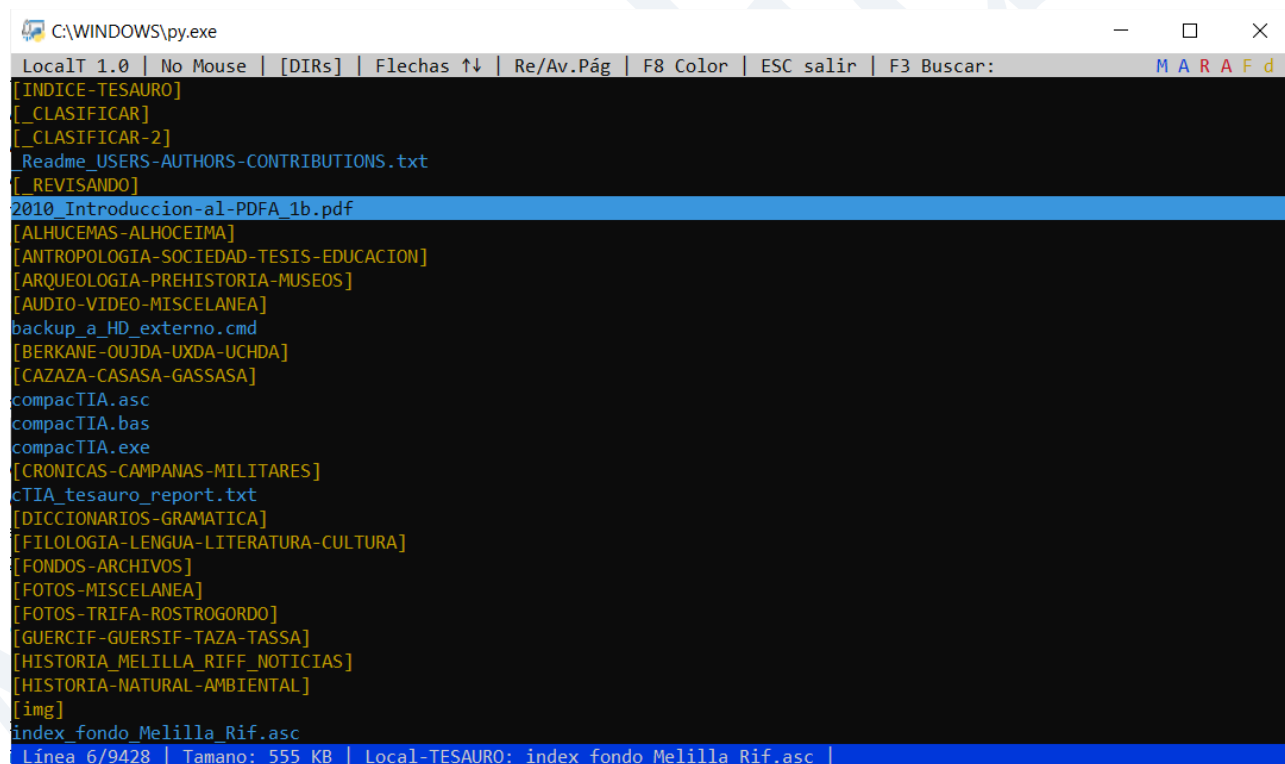
- Versión **3.12.2** del intérprete *Python*, disponible en la dirección:

<https://www.python.org/ftp/python/3.13.2/python-3.13.2-amd64.exe>

- O bien utilizar la incluida en el archivo comprimido **ecosistema_MARAFd.zip**

La vista de la utilidad **localT.py** puede alternarse entre pantalla completa y ventana pulsando la combinación de teclas **Alt + ENTER**

En la imagen siguiente se muestra una captura de la utilidad **localT.py** corriendo en una ventana de consola de *Microsoft Windows 10*®



```
C:\WINDOWS\py.exe
LocalT 1.0 | No Mouse | [DIRs] | Flechas ↑↓ | Re/Av.Pág | F8 Color | ESC salir | F3 Buscar: M A R A F d
[INDICE-TESAURO]
[_CLASIFICAR]
[_CLASIFICAR-2]
_Readme_USERS-AUTHORS-CONTRIBUTIONS.txt
[_REVISANDO]
2010_Introduccion-al-PDFA_1b.pdf
[ALHUCEMAS-ALHOCEIMA]
[ANTROPOLOGIA-SOCIEDAD-TESIS-EDUCACION]
[ARQUEOLOGIA-PREHISTORIA-MUSEOS]
[AUDIO-VIDEO-MISCELANEA]
backup_a_HD_externo.cmd
[BERKANE-OUJDA-UXDA-UCHDA]
[CAZAZA-CASASA-GASSASA]
compactIA.asc
compactIA.bas
compactIA.exe
[CRONICAS-CAMPANAS-MILITARES]
cTIA_tesauro_report.txt
[DICCIONARIOS-GRAMATICA]
[FILOLOGIA-LENGUA-LITERATURA-CULTURA]
[FONDOS-ARCHIVOS]
[FOTOS-MISCELANEA]
[FOTOS-TRIFA-ROSTROGORDO]
[GUERCIF-GUERSIF-TAZA-TASSA]
[HISTORIA_MELILLA_RIFF_NOTICIAS]
[HISTORIA-NATURAL-AMBIENTAL]
[img]
index_fondo_Melilla_Rif.asc
Línea 6/9428 | Tamaño: 555 KB | Local-TESAURO: index_fondo_Melilla_Rif.asc
```

2.2. Instalación del ecosistema **MARAFd**

Para instalar un ecosistema **MARAFd** y hacerlo funcionar con nuestro fondo documental, el que sea que tengamos preparado para montar, lo haremos a partir del fichero comprimido **ecosistema_MARAFd.zip** el cual contiene todos los archivos necesarios.

Antes de comenzar a trabajar, es importante comprender como se estructura el fondo y qué archivos vamos a necesitar a nivel local (en nuestro equipo) y en el servidor. En cualquier caso, son pocos ficheros los que deberemos configurar para preparar y hacer funcionar nuestro fondo documental.

Para ello vamos a tomar como ejemplo el *Fondo Documental Digital MELILLA-RIFF* que ya se encuentra en pleno funcionamiento y accesible en línea, de esta forma evitaré afirmaciones que podrían estar equivocadas.

Con un número de archivos superior a los 9000 ficheros y un volumen total de más de 100 *GBytes* de información, la estructura de este fondo documental, plenamente adaptable a nuestras necesidades, es lo que veremos con detalle a continuación.

2.2.1. Estructura básica de carpetas.

En un primer momento, descomprimido el archivo **ecosistema_MARAFd.zip** en una carpeta de nuestro disco duro, tendremos a nuestra disposición todos los ficheros necesarios que debemos editar y adaptar para poner en marcha nuestro fondo.

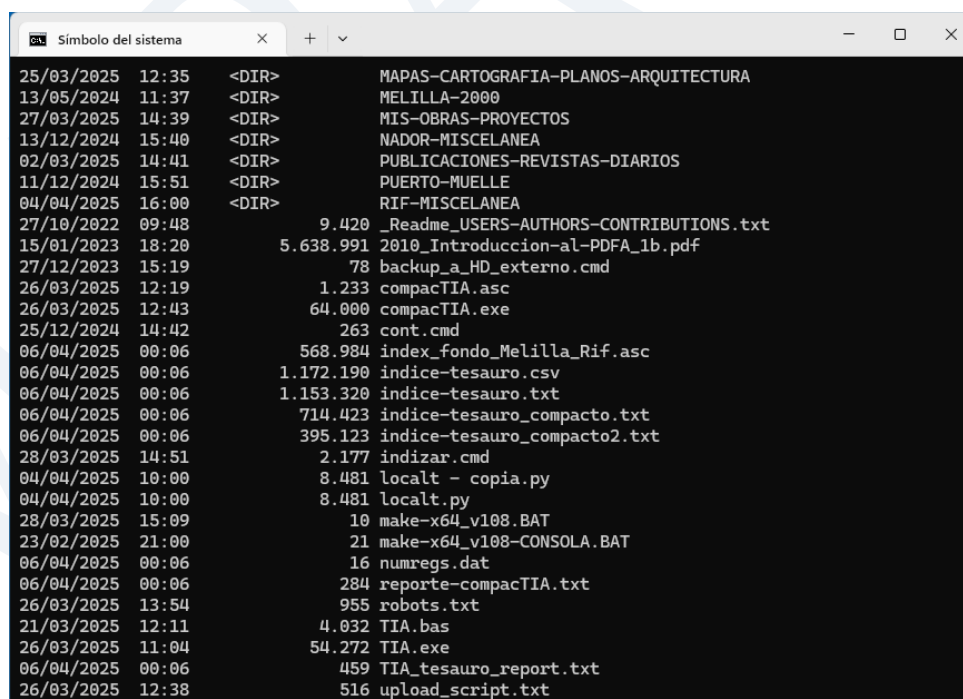
Lo siguiente será montar la estructura de carpetas y archivos del fondo en nuestro equipo local, para después transferirlo al servidor por FTP. Esta transferencia puede hacerse una vez que hayamos editado y revisado los ficheros *batch-script* y comprobado que los índices-tesauros del fondo se generan correctamente.

En nuestro modelo de ejemplo, basado como digo en el Fondo Documental Digital MELILLA-RIFF , tendríamos una estructura de carpetas similar a la que se muestra en la imagen de la página siguiente.

<p>Carpetas y archivos en disco duro local (archivos de la web¹ y fondo documental²)</p>	<p>Carpetas y archivos en servidor web (archivos de la web¹ y fondo documental²)</p>
<p>Réplica de la web (frontend) del fondo en equipo local con librería de archivos bootstrap (js, css) en carpeta distinta a la del fondo documental:</p> <pre> ../fondo_documental_Melilla_Riff/¹ ├── index.html (web de acceso principal frontend) ├── _Readme_USERS-AUTHORS-CONTRIBUTIONS.html ├── busqueda_avanzada.html (textos informativos) ├── finder.php (motor rastreo tesauro) │ ├── ../js │ └── ../css </pre> <p>Réplica espejo del fondo documental completo en equipo local con nombre coincidente en el servidor para la correcta generación del tesauro e hiperenlaces desde equipo local. Siempre se trabaja en local y la sincronización del contenido entre equipo local y servidor se realiza por sincronización FTP de carpetas (FileZilla, WinSCP, etc.)</p> <pre> ../_fondos/² (carpeta_contenedora_fondos_en_disco_local) ├── /ebiblio_melilla_riff │ ├── carpetas+archivos_fondo │ ├── ebiblio_historia_antigua_01 │ │ ├── carpetas+archivos_fondo │ │ └── ebiblio_manuales_programacion_01 │ │ └── carpetas+archivos_fondo </pre>	<p>Réplica espejo de la web (frontend) del fondo en equipo remoto con librería de archivos bootstrap (js, css) en carpeta distinta a la del fondo documental:</p> <pre> ../fondo_documental_Melilla_Riff/ ├── index.html (web de acceso principal frontend) ├── _Readme_USERS-AUTHORS-CONTRIBUTIONS.html ├── busqueda_avanzada.html (textos informativos) ├── finder.php (motor rastreo tesauro) │ ├── ../js │ └── ../css </pre> <p>Réplica espejo del fondo documental completo en servidor web. La sincronización del contenido entre equipo local y servidor se realiza por sincronización FTP de carpetas (FileZilla, WinSCP, etc.)</p> <pre> /ebiblio_melilla_riff² ├── index.php (genera enlaces del fondo completo) │ ├── lista_archivos.html (generada por index.php para navegación por carpetas) │ ├── indice-tesauro.csv (creado en Local) │ ├── indice-tesauro.txt (idem) │ ├── index_fondo_Melilla_Rif.asc (idem) │ └── resto_carpetas+archivos_fondo ├── ebiblio_historia_antigua_01 │ ├── carpetas+archivos_fondo │ └── ebiblio_manuales_programacion_01 │ └── carpetas+archivos_fondo </pre>
<p>Las carpetas y los ficheros referenciados en este apartado con los sobreíndices 1 y 2, deben ser una copia exacta (mirror local) de todos los ficheros y carpetas de nuestra web alojadas en nuestro disco local. Las carpetas indicadas como js y css contienen los ficheros necesarios para la ejecución del estándar Bootstrap que permitirá la compatibilidad visual del frontend de nuestra web tanto en dispositivos móviles y de escritorio mediante un interface dinámico.</p>	<p>Las rutas relativas (nombre de carpeta y archivos) de los fondos documentales deben coincidir en el equipo local y el servidor, lo cual facilita la generación y adaptación de hiperenlaces a partir del tesauro que se genera localmente.</p> <p>Los índice-tesauro, generado localmente y transferido al servidor mediante procesamiento por lotes automatizado (batch-script), son básicamente el mismo y servirán a las distintas herramientas como el agente de IA (MyGPT®) para rastrear y localizar información en el fondo. Algunos formatos como el .CSV pueden ser mejor reconocidos por determinados modelos de IA, de ahí que generemos varios tipos, básicamente para que satisfacer a la IA.</p>

2.3. Configuración inicial. Ajustes básicos y pruebas de funcionamiento tras instalación

Es importante saber que, tanto en el equipo local como en el servidor, la carpeta principal que alberga el fondo (subdirectorios y archivos) deben tener exactamente el mismo nombre tanto en el servidor como en el disco duro local, ahí radica una de las particularidades propias del potencial de **MARAFd** y eso nos permitirá aprovechar el tesoro ampliado tanto a la hora de generar enlaces dinámicos como a la hora de realizar búsquedas en línea de ficheros.



```
Símbolo del sistema
25/03/2025 12:35 <DIR> MAPAS-CARTOGRAFIA-PLANOS-ARQUITECTURA
13/05/2024 11:37 <DIR> MELILLA-2000
27/03/2025 14:39 <DIR> MIS-OBRAS-PROYECTOS
13/12/2024 15:40 <DIR> NADOR-MISCELANEA
02/03/2025 14:41 <DIR> PUBLICACIONES-REVISTAS-DIARIOS
11/12/2024 15:51 <DIR> PUERTO-MUELLE
04/04/2025 16:00 <DIR> RIF-MISCELANEA
27/10/2022 09:48 9.420 _Readme_USERS-AUTHORS-CONTRIBUTIONS.txt
15/01/2023 18:20 5.638.991 2010_Introduccion-al-PDFA_1b.pdf
27/12/2023 15:19 78 backup_a_HD_externo.cmd
26/03/2025 12:19 1.233 compactIA.asc
26/03/2025 12:43 64.000 compactIA.exe
25/12/2024 14:42 263 cont.cmd
06/04/2025 00:06 568.984 index_fondo_Melilla_Rif.asc
06/04/2025 00:06 1.172.190 indice-tesauro.csv
06/04/2025 00:06 1.153.320 indice-tesauro.txt
06/04/2025 00:06 714.423 indice-tesauro_compacto.txt
06/04/2025 00:06 395.123 indice-tesauro_compacto2.txt
28/03/2025 14:51 2.177 indizar.cmd
04/04/2025 10:00 8.481 localt - copia.py
04/04/2025 10:00 8.481 localt.py
28/03/2025 15:09 10 make-x64_v108.BAT
23/02/2025 21:00 21 make-x64_v108-CONSOLA.BAT
06/04/2025 00:06 16 numregs.dat
06/04/2025 00:06 284 reporte-compactIA.txt
26/03/2025 13:54 955 robots.txt
21/03/2025 12:11 4.032 TIA.bas
26/03/2025 11:04 54.272 TIA.exe
06/04/2025 00:06 459 TIA_tesauro_report.txt
26/03/2025 12:38 516 upload_script.txt
```

2.4. Aportaciones al fondo automatización de procesos. Entre *BatchScript* y *toolkits*.

2.4.1. Backup.

Empiezo por este apartado porque es más importante de lo que algunos puedan pensar a priori. En nuestro caso podemos optar por un proceso de backup redundante del tipo incremental, o lo que es lo mismo, mantener una copia exacta del fondo en distinta/s unidad/es de disco locales. Para lograrlo con solvencia y velocidad, recurrimos al uso del comando **robocopy** de *Microsoft Windows 10/11®* llamado desde un sencillo *BatchScript* como este que mostramos a continuación. Observe que puede modificar

fácilmente el contenido del mismo para efectuar tantas copias redundantes como desee y a las unidades disco que se indiquen:

backup a HD externo.cmd

```
@echo off
echo Proceso completo de backup incremental/sincronizado en unidad de disco distinta...
robocopy \ebiblio_melilla_riff M:\ebiblio_melilla_riff /MIR /NP
echo Fin de impresiOn
```

El proceso *batch* encargado de realizar la copia de seguridad, puede ser ejecutado directamente de forma manual o mediante una llamada desde dentro (añadir siempre al final de todos los procesos) del *batchscript* **indizar.cmd** de forma que con cada nuevo indizado en el fondo documental se realice un proceso de respaldo automático.

2.4.2. Copiado de archivos en equipo local y sincronización FTP con servidor.

Este proceso de aportación de archivos al fondo es prácticamente el único proceso que debe efectuarse de forma manual. Básicamente consiste en, una vez realizado el necesario e imprescindible proceso de renombrado descriptivo del fichero/s, depositar el mismo (o bloque de ellos) en la carpeta elegida de nuestro equipo local.

Tras este proceso básico de copiado o “arrastrar y soltar” se procederá a la conexión FTP con el *hostingweb* y situaremos ambos lados (cliente y servidor) en el directorio (carpeta) que vamos a sincronizar.

La sincronización de carpetas entre cliente (local) y servidor (*hostingweb*) es algo sencillo y puede hacerse de forma automática desde cualquier software cliente FTP (en nuestro caso veremos *WinSCP*) y garantiza la integridad del *mirror* en todo momento. En cualquier caso, recomiendo familiarizarnos antes con la aplicación FTP que utilicemos para evitar borrados accidentales.

Toda vez copiados los ficheros en ambos lados (cliente-servidor) y asegurados de que cuentan con el mismo nombre y se sitúan en la misma carpeta, pasamos a la siguiente tarea.

2.4.3. Indizado completo en local y sincronización del índice-tesauro en servidor

Este es un proceso completamente automatizado mediante un guión batch (también se denominan *batchscript* o *scriptbatch*). El archivo por lotes **indizar.cmd** recoge y ejecuta todos los procesos necesarios para el indizado y actualización del tesauro en local y remoto, además de realizar otras funciones auxiliares que iremos citando en esta manual.

El fichero que se muestra a continuación es un ejemplo que puede adaptarse a la ruta de cualquier fondo documental de forma rápida y segura.

```
[ indizar.cmd ]

@echo off
@cls
echo $ Contando registros y calculando espacio..
:: Carpeta local donde albergamos la copia completa del fondo
set "TARGET_DIR=D:\fondos\ebiblio_melilla_riff\"
:: Archivo donde se escribe el numero total de ficheros del fondo y tamaño en GBs
set "OUTPUT_FILE=numregs.dat"

if not exist "%TARGET_DIR%" (
    echo La carpeta especificada no existe: %TARGET_DIR%
    exit /b 1
)

:: Contar archivos
for /f %A in ('dir /s /a-d /b "%TARGET_DIR%" ^| find /c /v ""') do set "FILE_COUNT=%A"

:: Calcular tamaño total en GB con PowerShell
for /f "delims=" %A in ('powershell -command "$size = (Get-ChildItem -Path '%TARGET_DIR%' -Recurse | Measure-Object -Property Length -Sum).Sum / 1GB; [math]::Round($size, 2)'"') do set "SIZE_GB=%A"

:: Escribir resultados en el archivo
echo %FILE_COUNT% > "%OUTPUT_FILE%"
echo %SIZE_GB% >> "%OUTPUT_FILE%"

timeout /t 2
@chcp 1252

echo '
echo $ Please wait...
echo [ Indexando repositorio ... ]
dir /b /s /o:n >> index_fondo_Melilla_Rif.tmp
echo $ Proceso de indexado OK ...
echo $ Adaptando INDICE para enlaces URLs en servidor ...
timeout /t 4 /nobreak >nul
setlocal enabledelayedexpansion

echo INDICE-TESAURO > index_fondo_Melilla_Rif.asc

>>index_fondo_Melilla_Rif.asc (
    for /f "delims=" %i in (index_fondo_Melilla_Rif.tmp) do (
        set "line=%i"
        echo !line:D:\fondos\ebiblio_melilla_riff\=!
    )
)
```

```

)

timeout /t 2 /nobreak >nul
echo $ Revisando estructura INDICE-TESAURO y generando reporte de warnings ...
TIA.exe
timeout /t 2 /nobreak >nul
echo $ Eliminando archivo de INDEX temporal ...
del index_fondo_Melilla_Rif.tmp
echo $ Revisando y compactando INDICE-TESAURO y generando archivos compactos Tesauro y Tesauro-2 ...
compactTIA.exe
timeout /t 2 /nobreak >nul
echo $ Conectando al servidor mediante SFTP...
timeout /t 2 /nobreak >nul
echo $   subiendo archivos a directorio /fondo_documental_Melilla_Riff
timeout /t 2 /nobreak >nul
echo $   subiendo archivos a directorio /ebiblio_melilla_riff
"C:\Program Files (x86)\WinSCP\WinSCP.exe" /script=upload_script.txt

if %ERRORLEVEL% neq 0 (
    echo $ Error: No se pudo subir el/los archivo/s al servidor.
) else (
    echo $ Archivo/s subido/s correctamente.
)

echo $ FIN DE IMPRESION
pause

```

Si analizamos el *scriptbatch* **indizar.cmd** observamos que, además de generar el **índice-tesauro** de todo el fondo documental a partir del contenido de una carpeta local (donde debemos tener una copia local y exacta del fondo completo) y volcarlo en el fichero **index_fondo_Melilla_Rif.asc** , también se encarga, entre otras cosas, de contar el número total de ficheros contenidos en el repositorio y calcular su tamaño en *GBytes*, escribiendo estos resultados en el fichero **numregs.dat** que, junto a todos los tesauros generados (a partir de las *toolkits* complementarias que más adelante estudiaremos), son subidos automáticamente al servidor gracias a las indicaciones dadas en el fichero **script_load.txt** y que son ejecutadas por nuestro software cliente FTP *WinSCP*. Todo esto y mucho más desde el archivo **indizar.cmd**

Aquí vemos un ejemplo del archivo encargado de realizar la conexión FTP con el servidor y ejecutar los comandos necesarios para subir los ficheros generados al espacio *hostingweb*:

[upload_script.txt]

```

# Contenido de upload_script.txt
# Subida redundante de los tesauros a carpetas del servidor
#   para que puedan utilizarse desde cualquiera de ellas

open sftp://usuario:password@servidor.servidor.servidor:22
cd ebiblio_melilla_riff
put index_fondo_Melilla_Rif.asc
put numregs.dat
put indice-tesauro.txt

```

```
put indice-tesauro.csv
put indice-tesauro_compacto.txt
put indice-tesauro_compacto2.txt
cd ..
cd fondo_documental_Melilla_Riff
put index_fondo_Melilla_Rif.asc
put numregs.dat
put indice-tesauro.txt
put indice-tesauro.csv
put indice-tesauro_compacto.txt
put indice-tesauro_compacto2.txt
exit
```

El archivo **upload_script.txt** es, como decimos, un archivo de proceso por lotes reconocido y ejecutado por el software cliente FTP *WinSCP*.

MUY IMPORTANTE: Si bien en el caso del fondo documental digital MELILLA-RIFF la carpeta **fondo_documental_Melilla_Riff** contiene únicamente los archivos del *frontend* y la copia del fondo documental se alberga en otra carpeta diferente que es **ebiblio_melilla_riff**, esto no es realmente un criterio necesario ya que los archivos que conforman el *frontend* pueden albergarse igualmente en la carpeta principal o raíz del fondo, pero sí puede ser una forma interesante de modularizar la información y mantener por separado ambos conceptos (el *FrontEnd* por un lado y la carpeta contenedora del fondo documental por otro), especialmente en el caso de que debamos compartir el servidor *hosting-web* para alojar páginas ajenas al fondo.

No ocurre así sin embargo en el caso del Fondo Documental de Protohistoria e Historia Antigua de las culturas del Mediterráneo occidental, Iberia y norte de África, gestionado igualmente mediante tecnología **MARAFd** y accesible actualmente desde la dirección https://calentamientoglobalacelerado.net/ebiblio_historia_antigua_01/ en el cual se ha optado por albergar los archivos del *FrontEnd* en la misma carpeta del fondo documental.

Lo único importante del sistema *mirror* es, como ya se ha comentado, que el nombre de la carpeta local que albergue el fondo documental se corresponda siempre con el directorio remoto en el servidor.

Por otro lado, los programas **TIA.exe** y **compactIA.exe** que se detallarán más adelante, son pequeñas rutinas independientes escritas en *FreeBASIC* y diseñados para verificar y generar nuevos índices-tesauros mejor adaptados y mucho más manejables a partir del fichero **index_fondo_Melilla_Rif.asc** generado desde el *batchscript* **indizar.cmd**.

2.5. Herramientas complementarias (*toolkits*)

El ecosistema **MARAFd** incorpora una serie de pequeñas aplicaciones, todas de código abierto, muy específicas y destinadas a realizar cometidos igualmente especializados.

En la actualidad, a fecha de la elaboración del presente manual, se disponen de las que se describen a continuación:

- **TIA.exe** (aplicación de consola ejecutada desde **indizar.cmd**) revisa la estructura del tesoro original completo y reporta posibles fallos en los nombres en **TIA_tesoro_report.txt** ayudándonos a depurar y pulir el tesoro de forma más rápida y sencilla. No olvidemos que el tesoro, al prescindir de un sistema de base de datos, es la auténtica columna vertebral del fondo documental. Además, genera dos ficheros tesoro de contenido análogo a partir del archivo original como son:
 - **indice-tesoro.txt**
 - **indice-tesoro.csv**
- **compactTIA.exe** (aplicación de consola ejecutada desde **indizar.cmd**) se centra en aplicar un sencillo proceso de compactado al tesoro obtenido en **indice-tesoro.txt** mediante la eliminación de información redundante y selectiva que nos permite reducir el tamaño del tesoro a prácticamente la mitad del tesoro original generado por el fichero **indizar.cmd**, generando además un archivo de reporte llamado **ctIA_tesoro_report.txt** además de dos variaciones del tesoro completo original y que son:
 - **indice-tesoro_compacto.txt** con la exclusión del tesoro original de todos los archivos con ciertas extensiones. Estas extensiones eliminadas son indicadas a través del fichero **compactTIA.asc**
 - **indice-tesoro_compacto2.txt** en el que se elimina además la parte redundante del *path* o ruta base de todos los archivos del fondo como es el dominio principal, y que en el caso de nuestro fondo es: https://calentamientoglobacelerado.net/ebiblio_melilla_riff/

2.6. Ejecución *PHP* del lado del servidor

El ecosistema **MARAFd** hace uso también de la tecnología abierta *PHP*, la cual se ejecuta en el lado del servidor. Aunque lo veremos después con mayor detalle, estas dos son las tareas que se ejecutarán en el lado del servidor:

2.6.1. Acceso al fondo completo mediante navegación por carpetas.

Este código *PHP* (**index.php**) se encarga de generar una página *HTML* con todo el contenido del fondo y accesible mediante enlaces.



2.6.2. Búsqueda rápida (en tesauro).

El motor de búsqueda rápida consiste básicamente en una página web escrita en *PHP* (**finder3.php**) y dedicada exclusivamente al rastreo del índice-tesauro en busca de coincidencias y la generación del reporte en línea, con los enlaces dinámicos correspondientes.

· Para hacer búsquedas en el interior de varios documentos consulte con el administrador en eurocamsuite[arroba]yahoo.es o utilice el SISTEMA EXPERTO desde la web principal del fondo.

|ej ANUAL (2 palab./word MAXIMO)

Buscar

✓ ANTROPOLOGIA-SOCIEDAD-TESIS-EDUCACION/2017_TD_Estudio_medico_legal_donante_organos_VILLALBA_SORIA_Maria_Victoria.pdf

✓ HISTORIA_MELILLA_RIFF_NOTICIAS/2014_Medicos_cirujanos_sitio_Melilla_1774-1775.pdf

✓ _CLASIFICAR/2014_Memorias_medico_protectorado_espanol_mediados_siglo_XX.pdf

· BUSQUEDA: **medicos**

Coincidencias halladas: 3
(0.03%) de un total aproximado de 9271 archivos

3. Interfaz del usuario.

3.1. Descripción general.

Si bien las fórmulas para el diseño y estilo de la interfaz de cualquier fondo documental son absolutamente adaptables y prácticamente infinitas, en este caso nos centraremos en una experiencia propia completamente funcional y analizaremos los elementos principales de la interfaz y su disposición tomando ejemplo del actual *Fondo Documental Digital Melilla-Riff*.

El desarrollo de dicho proyecto, implementado de forma paralela junto con otro *Fondo Documental de Protohistoria Antigua en el Mediterráneo Occidental* de acceso restringido, ha servido para probar, pulir y mejorar en numerosos aspectos nuestra tecnología **MARAFd**, otorgándole a ésta, tras algo más de dos años de trabajo, la solvencia necesaria exigible a una solución de estas características.

En este sentido, pasamos a desglosar los diferentes apartados o secciones que conforman la web principal de acceso al *Fondo Documental Melilla-Riff (frontend)* desarrollada básicamente mediante *HMTL5*, con apoyo de herramientas de inteligencia artificial y cuya imagen se muestra en la siguiente captura.

La página principal que iremos desglosando a continuación, además del enlace opcional de envío de documentos, consta básicamente de 5 apartados centrados en el acceso a la información contenida en el fondo y que son:

1. Acceso al fondo completo (por carpetas)
2. Búsqueda rápida (en índice-tesauro)
3. Rastreo profundo en interior de documentos (investigación en modo local)
4. Sistema Experto IA (*MyGPT*® entrenado especialmente para cada fondo documental)
5. Búsqueda acotada con tecnología de terceros (buscadores externos como *Google*®, *Bing*®, etc)

Todas estas fórmulas de búsqueda están implementadas y son completamente funcionales en el *Fondo Documental Digital Melilla-Riff* para su testeo.

3.1.1. CABECERA de página web con imagen corporativa.

Área de cabecera en la que se puede visualizar la imagen corporativa identificativa del fondo documental, el título principal del mismo y cualquier otra información que se considere.

3.1.2. ENVIAR DOCUMENTO/S al fondo documental.

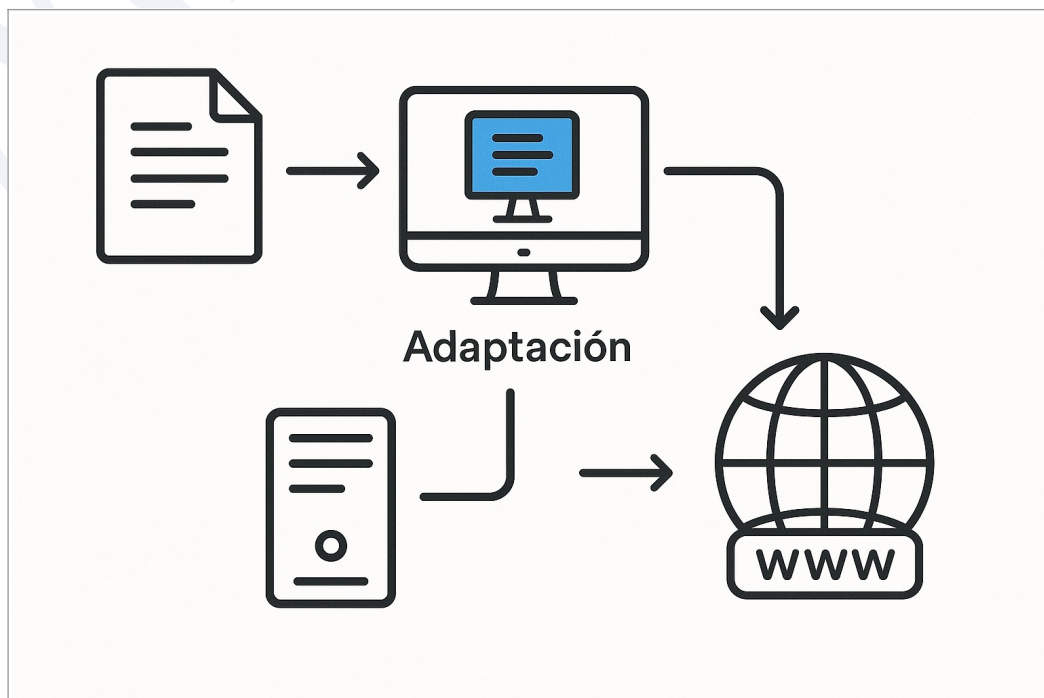
En esta sección podemos ofrecer meramente un enlace a texto informativo acerca del procedimiento básico a seguir para remitir archivos al fondo documental.

ENVIAR DOCUMENTO/S al fondo

El fondo documental digital Melilla-Rif no es un trabajo concluido sino un proyecto abierto y en curso con el que usted puede contribuir enviando documentos para su indizado.

Este apartado es obviamente opcional y servirá de apoyo a posibles colaboradores en el caso de fondos documentales abiertos. Es fundamental redactar un texto sencillo y claro que facilite al usuario los canales posibles para aportación de material al fondo así como el contacto con el administrador del mismo a través de un email.

Antes de poner en línea cualquier documento, el administrador del fondo deberá adaptar y supervisar finalmente el nombre y las propiedades del mismo para mantener los criterios de calidad y clasificación necesarios en todos los contenidos digitales indizados (Ver sección **GESTIÓN DE DOCUMENTOS** de la presente guía).



3.1.3. Acceso al FONDO COMPLETO (por directorios).

Acceso al fondo completo a través de la estructura de carpetas del mismo. Esta página es generada de forma dinámica mediante un guion *PHP* que se ejecuta de forma recursiva en el lado del servidor y genera una página *HTML* con todo el contenido del fondo.

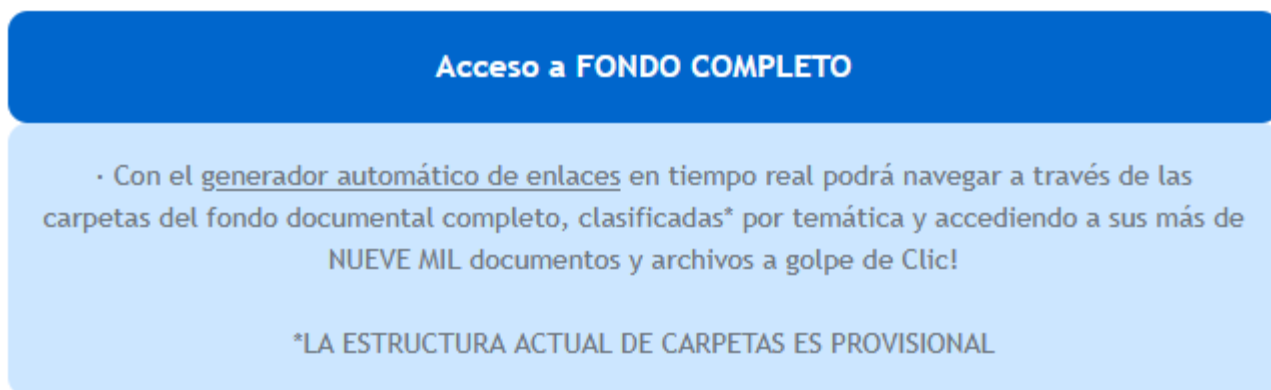
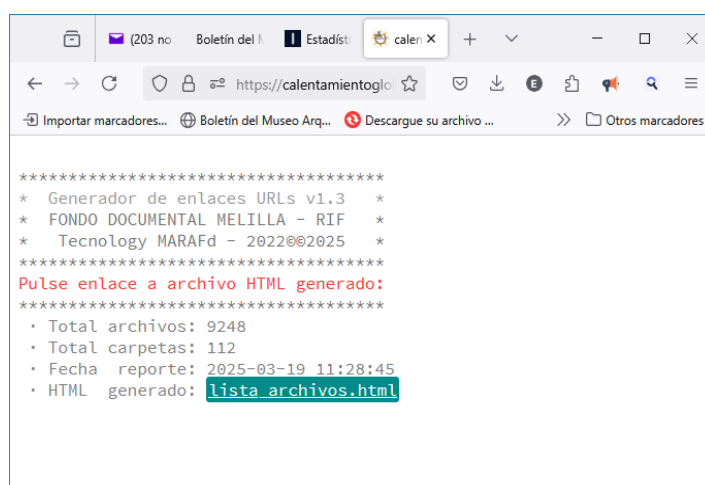


Imagen de la sección de ACCESO A FONDO COMPLETO tal y como se muestra en la página principal del Fondo Documental Digital Melilla-Rif.

Al clicar en Acceso a FONDO COMPLETO accedemos a una breve pantalla informativa del generador de enlaces (**index.php**) encargado de crear la página **lista_archivos.html** conteniendo el índice de todas las carpetas así como todos los archivos del fondo documental.

Esta que vemos a continuación es la pantalla previa a la generación de la página **lista_archivos.html**



Al clicar el texto **lista_archivos.html** accedemos ya al índice de todas las carpetas y archivos del fondo documental completo actualizado. Es una página *html* generada de forma dinámica por el código *PHP* contenido en **index.php** y en el caso del *Fondo Documental Melilla-Rif* es bastante extensa, por lo que puede a veces puede requerir un par de segundos de proceso antes de que podamos verla.

En esta página generada dinámicamente mediante *PHP* y dado que contiene la totalidad de archivos del fondo, nos resultará muy útil el uso de la herramienta de búsqueda por palabras integrada en los navegadores y accesible normalmente mediante la secuencia de teclas **CTRL+F**, un sencillo mecanismo que nos permite localizar rápidamente cadenas de palabras, contenidos y archivos de forma inmediata.

proyecto personal e independiente
NO-SUBVENCIONADO

Fondo Documental Digital MELILLA - RIF

thesaurus technology
MARAFd
gestión de fondos documentales

¡Más de **9000** documentos multimedia indizados!

COMO UTILIZAR EL CONTENIDO DE ESTA PAGINA

- Use el buscador de palabras de su navegador (Ctrl+F)
- Algunos navegadores pueden reproducir directamente los archivos multimedia
- Descargue a su disco duro cualquier documento que necesite

INDICE PROVISIONAL DE CARPETAS:

- ALHUCEMAS-ALHOCEIMA
- ANTROPOLOGIA-SOCIEDAD-TESIS-EDUCACION
- ARQUEOLOGIA-PREHISTORIA-MUSEOS
- AUDIO-VIDEO-MISCELANEA

3.1.4. BÚSQUEDA RÁPIDA (en índice-tesauro).

El sistema de **Búsqueda Rápida** consiste en un sencillo pero eficiente motor de búsqueda implementado en *PHP* y que rastrea el índice tesauro.

BÚSQUEDA RÁPIDA (en tesauro)

- El modo más rápido de encontrar lo que busca. Este motor de búsqueda filtra fácilmente documentos por palabras, tipo de archivo, año del documento, etc.

La información relevante codificada en el propio tesauro le permite acceso inmediato a más de 100 GB de datos!

Implementado a través del código *PHP* contenido en el fichero **finder3.php**, el buscador rápido realiza una búsqueda booleana del tipo *AND* en el texto completo del índice-tesauro con dos palabras como máximo, generándose con cada consulta un reporte básico con las coincidencias halladas y los hiperenlaces correspondientes a los ficheros encontrados.

· Para hacer búsquedas en el interior de varios documentos consulte con el administrador en eurocamsuite[arroba]yahoo.es o utilice el SISTEMA EXPERTO desde la web principal del fondo.

ej ANUAL (2 palab./word MAXIMO)

Buscar

✓ CRONICAS-CAMPANAS-MILITARES/ANUAL-ANUAL/1931_expediente_Picasso_Anuar_ES.pdf

✓ CRONICAS-CAMPANAS-MILITARES/ANUAL-ANUAL/1984_retazos_historicos_Expediente_Picasso.pdf

· BUSQUEDA: exped pica

Coincidencias halladas: 2
(0.02%) de un total aproximado de 9250 archivos

Con el objeto de dotar a este buscador de un mayor potencial, **MARAFd** implementa en el código *PHP* funciones de búsqueda avanzada que van más allá de la normalización de mayúsculas y minúsculas. Dichas funciones flexibilizan drásticamente las consultas permitiendo al usuario, por ejemplo, hallar documentos sobre *Alhucemas* habiendo

teclado *Alucemas*, o documentos con la palabra *Casasa* habiendo teclado *CAZAZA*. O incluso *CAZZAZZA*.

También se igualan letras como *N* y *Ñ*, o vocales acentuadas con tildes con vocales sin tildes, e incluso una consulta del tipo *médicos* ó *doctores* (en plural) que daría también como resultado positivo las apariciones en el tesoro en número singular, es decir, *medico* y/o *doctor*. La *Q* es también equivalente a la *K* y la *B* a la *V*, con lo cual logramos una enorme flexibilidad de consulta a nivel fonético.

En resumen, el buscador rápido logra un sistema de acceso a los documentos inmediato pero a la vez potente, teniendo siempre presente que este buscador únicamente rastrea el texto contenido en el índice-tesoro del fondo y nunca en el interior de los documentos contenidos en el mismo.

3.1.5. RASTREO PROFUNDO (investigación).

Lo primero que debemos saber es que el rastreo o búsqueda profunda no es un servicio disponible en línea para el usuario del fondo y solo puede llevarse a cabo desde el equipo local y por parte del administrador, aunque en realidad, dicho servicio podría replicarse por parte de cualquier usuario sobre los archivos ya descargados en su disco duro.

El rastreo profundo consiste básicamente en la búsqueda de palabras en el interior de los documentos del fondo, pudiendo escrutarse desde una única carpeta hasta el total del fondo documental.

RASTREO PROFUNDO (investigación)

· Pensada en la más rigurosa investigación, la búsqueda profunda es para un historiador algo similar a lo que un acelerador de partículas podría suponer para un físico teórico.

La búsqueda booleana de palabras en el interior de todos los documentos es una herramienta a tener en consideración en el ámbito investigativo.

Debido a los altos requerimiento de computación este tipo de búsqueda solo puede efectuarse de forma local y no en línea.

Esta opción de búsqueda local puede realizarse mediante herramientas software como *Adobe Acrobat Reader*® u otras soluciones comerciales y/o de libre distribución, que

permitan lanzar búsquedas de texto lógicas o semánticas a múltiples archivos y carpetas sobre unidades de disco local o mapeadas por el sistema operativo. Estas búsquedas, al menos en el caso de la herramienta de *Adobe*, generan reportes completamente funcionales y archivables como documentos *PDF*.



En la sección de ANEXOS se muestra un reporte de ejemplo

3.1.6. Sistema Experto IA

Si bien se trata de una sistema de búsqueda opcional, es cierto que, dado el carácter abierto de todas las tecnologías empleadas por el ecosistema **MARAFd**, podremos aprovechar esta característica para combinar de forma sencilla el contenido de nuestro fondo documental con herramientas muy poderosas como son los *LMM* (*Large Language Model* ó *modelos de lenguaje grande*).

Sistema Experto IA (Beta)

· [**EN FASE EXPERIMENTAL**] Trabaje y experimente ya con la última versión del poderoso Sistema Experto **MyGPT®*** **MARAFd** entrenado especialmente para ello.

[**IMPORTANTE**] Se requiere para su uso cuenta GRATUITA de suscripción a ChatGPT® pudiendo usar para ello su cuenta Gmail®, Yahoo® ó Hotmail® para suscribirse. Este es un servicio externo y podría no estar operativo por causas ajenas, en ese caso pruebe más tarde.

* Debido a la posibilidad de alucinaciones propias de la IA, se recomienda verificar cualquier información obtenida a través de este servicio.

Estos agentes de inteligencia artificial, una vez entrenados y revisados de forma adecuada, pueden llevar el análisis e interpretación de la información contenida en el fondo a un nuevo nivel.


Por ello, aprovechando la accesibilidad en línea del 100% del contenido del fondo así como el almacenamiento de toda la información del mismo en formatos y estándares internacionales abiertos como puede ser el omnipresente *PDF*, la inteligencia artificial permite al usuario recabar información simplemente interactuando con la IA usando el lenguaje natural. Esta versatilidad de uso exclusiva de los nuevos modelos de IA abiertos, la convierte sin duda en una herramienta con un potencial muy interesante en el ámbito educativo y académico, sin desdeñar en absoluto el alcance que puede llegar a imprimir en las labores de investigación avanzada.

En nuestro caso de referencia implementamos un sistema experto basado en un *MyGPT®* (herramienta ofrecida por la empresa *OpenAI®*) entrenado especialmente para atacar el contenido de nuestro fondo documental.

Para obtener resultados satisfactorios, es fundamental que dicho entrenamiento se realice minuciosamente por el administrador mediante una configuración específica y el reforzamiento continuo del modelo a través de la interacción desde el *prompt* (caja de texto de entrada en la que tecleamos los mensajes al modelo).

En resumen, y aunque es importante conocer y manejar ciertas limitaciones de la IA para lograr un sistema de consulta eficiente y veraz, podemos afirmar que la inteligencia artificial nos ofrece una oportunidad única en el ámbito educativo y muy considerable en todo el espectro de la investigación, abriendo de pleno un nuevo y vasto horizonte aún por descubrir.



 En la sección de ANEXOS – *Implemetando IA para consultar nuestro fondo*, se muestra, con carácter orientativo, el contenido completo de la configuración* aplicada a nuestro *MyGPT®* entrenado como sistema experto en el acceso al *fondo documental digital MELILLA-RIFF* y que nos servirá de referencia para el diseño de cualquier otro *MyGPT®*.

***sistema_experto-entrenado_config_MyGPT_IA.txt**

MARAFd

3.1.7. Búsquedas con tecnología Google®


Con todo el contenido de nuestro fondo en línea y accesible desde Internet, podemos también aprovecharnos de las tecnologías clásicas de búsqueda, y en nuestro caso hemos optado por Google®.

Algo que debemos tener en consideración en este apartado es que, para que se pueda utilizar esta opción de búsqueda con un mínimo de garantías de éxito, es imprescindible que el buscador Google® haya indizado previamente el contenido del fondo y esto puede requerir un plazo indefinido de tiempo que no depende de nosotros. Por ello, no podemos obviar que aunque las búsquedas realizadas con el presente sistema puedan producir resultados positivos incluso sobre contenidos profundos (localizados en el interior de algunos documentos), éstos podrían no alcanzar a las últimas incorporaciones al fondo, por lo que es recomendable contrastar o ampliar las mismas con los otros métodos propuestos.

No obstante, y con el objeto de facilitar el proceso de indexado al robot de cualquier buscador, también es importante que preparemos y subamos al servidor un adecuado archivo **sitemap.xml** y un adecuado fichero **robots.txt**

Buscar con tecnología Google®

· Realiza búsquedas profundas en todos los documentos del fondo usando la tecnología del buscador Google®.

 En la sección de ANEXOS se muestra, con carácter orientativo, el contenido completo tanto del fichero **robots.txt** como del fichero **sitemap.xml**

4. Gestión de documentos

4.1. Tratamiento y carga de documentos: Cómo añadir nuevos documentos al fondo.

El proceso de carga o subida de documentos al servidor se realiza mediante un cliente *FTP* y de forma semiautomática gracias a la funcionalidad de sincronización de carpetas, la cual nos mantiene siempre sincronizado el contenido del equipo local con el servidor.

Sin embargo, antes de subir uno o varios documentos, es fundamental tener en consideración ciertos aspectos técnicos para lograr un fondo documental de máxima calidad y funcionalidad. Esto nos lleva irremediablemente a lo que algunos denominan “curación de datos”.

4.1.1. OCR (reconocimiento óptico de caracteres)

La curación de datos es básicamente todo lo que afecta a la adaptación de los documentos que vamos a indizar. En esta tarea de adaptación deberemos trabajar necesariamente con alguna herramienta de software como *Acrobat Reader*®, *PDF24*®, *PDFcreator*®, u otras similares.

Si el archivo ya está digitalizado y se trata de un documento, lo primero que debemos comprobar es si el texto del mismo, en caso de que sea un documento textual o combinado (texto e imágenes), es reconocible como tal para poder realizar búsquedas por palabras en su interior. De no ser así, deberemos pasar un *OCR* al mismo para convertir en texto puro todo el contenido del documento.

Aunque esta no es una tarea obligada para la preservación digital de documentos, de hecho existen grandes repositorios que obvian directamente este apartado, es muy importante realizarla siempre que sea posible porque mejorará drásticamente el nivel de accesibilidad a la información del fondo y el uso del mismo en tareas de investigación avanzadas.

A veces, tras aplicar un proceso de *OCR* a un documento *PDF*, el tamaño de este puede engrosarse de forma descabellada, por lo que posteriormente deberemos repasar el documento final con alguna herramienta de compresión que permita reducir su tamaño sin una pérdida significativa de calidad.

4.1.2. Digitalización de archivos documentales

La digitalización de documentos es una tarea ardua, lenta y laboriosa, y, aunque muchos puedan pensar lo contrario, requiere de cierta experiencia para lograr resultados aceptables. El escaneo de documentos originales, ya sean texto o imagen, debe realizarse, como mínimo, a una resolución de 300 ppp (píxeles por pulgadas) y en color real. Esta resolución, mantendrá un peso (en memoria) equilibrado con la calidad suficiente en su versión digital y garantizará que podamos acceder a los detalles, casi como si tuviéramos el original delante.

Es también importante prestar atención en el escaneo de documentos especialmente sensibles a este respecto como pueden ser material cartográfico, portulanos o cartas de navegación, manuscritos, etc.

4.1.3. Renombrado de archivos y carpetas

Es importante tomar en consideración ciertas normas básicas mínimas a la hora de establecer el nombre de los ficheros, pues algunos caracteres (como acentos, ñ o espacios) en los nombres de archivos o rutas pueden no ser gestionados de forma correcta por ciertos servidores, y esto puede desembocar en fallos de acceso.

El renombrado es, como la digitalización, otra labor fundamental a la hora de mantener la calidad del fondo pues forma parte del tesoro que constituye, como hemos dicho, el auténtico eje vertebral de éste en lo que a estructuración y búsquedas se refiere.

En este sentido, más importante aun puede llegar a ser la clasificación básica de carpetas del fondo en las que iremos introduciendo nuestros archivos, pues el nombre de estas carpetas y su ruta (*path*), forma parte del tesoro y por ello también será rastreado por el motor de búsqueda rápida y/o por el sistema experto que podamos implementar en nuestro fondo documental para la investigación.

4.2. Organización de documentos: Uso de carpetas, etiquetas y categorías para una mejor organización.

Los criterios de clasificación deben revisarse de forma rigurosa y exhaustiva para mantener un buen nivel de estructuración del contenido, no en vano, esta clasificación de archivos en carpetas forma parte directa del índice-tesoro que vertebra todo el contenido del fondo documental.

4.3. Búsqueda y filtrado: herramientas diversas para localizar documentos de manera eficiente.

Tal y como se ha visto en apartado precedente, el fondo documental digital Melilla-Rif ofrece diversos mecanismos de búsqueda en función de nuestras necesidades y que podemos resumir en:

- Navegación por carpetas: acceso al fondo documental completo mediante clic! De forma intuitiva a través de directorios.
- Búsqueda rápida en el tesoro: consistente en un sencillo y robusto código *PHP* ejecutado en el servidor que rastrea el índice-tesoro y genera enlaces dinámicos a los documentos.
- Consultas al Sistema Experto: *MyGPT® MARAFd-Sistema Experto MELILLA-RIF*. Utilizando un modelo largo de lenguaje *GPT-4o®* u otro similar. Si bien esta opción de búsqueda e investigación es opcional, puede ofrecer resultados muy interesantes y es bastante simple de implementar (*Ver sección ANEXOS – Implementando IA para consultar el fondo*)
- Tecnología de búsqueda Google® ó Bing® mediante búsquedas restrictivas al dominio del fondo a través de comandos como `site:calentamientogloballacelerado.net/ebiblio_melilla_riff/`
- Búsqueda/rastreo profundo. Requiere el uso de soluciones de software ajenas a la tecnología *MARAFd* tales como *Adobe Acrobat Reader®* u otras, y permite el acceso a búsquedas booleanas de palabras en el interior de los documentos de una carpeta o incluso de todas las carpetas del fondo documental, eso sí, con una alta carga computacional y un tiempo de procesamiento incluso de varias horas. Estas herramientas de búsqueda por palabras ofrecen la posibilidad de uso de los operadores lógicos *AND*, *OR*, etc.

4.4. Versionado: Gestión de diferentes versiones de un mismo documento.

El tratamiento del versionado puede resolverse fácilmente añadiendo un número o sufijo secuencial (*_01*, *_002*, *_bis*, etc) en el propio nombre del fichero de forma que se diferencien claramente las distintas versiones.

4.5. Seguridad y permisos

En este apartado mostramos como es posible restringir el acceso a un fondo documental en línea mediante clave de paso.

En nuestro caso estudiaremos el ejemplo del *Fondo Documental Digital de Protohistoria e Historia Antigua del Mediterráneo Occidental*, gestionado también mediante tecnología **MARAFd** y accesible en la dirección:

https://calentamientoglobalacelerado.net/ebiblio_historia_antigua_01/index.php



ACCESO RESTRINGIDO

AREA EXCLUSIVA PARA INVESTIGADORES [HISTORIA-01]

Ingrese clave de paso:

ADVERTENCIA: está tratando de acceder a un sector restringido con un gran volumen de material documental. Parte de este material podría encontrarse protegido por derechos de autor lo cual prohíbe expresamente su difusión a través de otros canales y/o medios. Utilice este material bajo su responsabilidad y únicamente con fines investigativos. Gracias

AVERTISSEMENT: vous êtes sur le point d'accéder à une zone restreinte contenant un grand volumen de matériel documentaire. Une partie de ce matériel podría être protégée por des droits d'auteur, ce qui interdit expressément sa diffusion por d'autres canaux y/o medios. Utilisez ce matériel sous votre propre responsabilité y uniquement à des fins de recherche. Merci

WARNING: You are about to access a restricted area with a large volume of documentary material. Some of this material may be protected by copyright, which expressly prohibits its dissemination through other channels and/or means. Use this material under your own responsibility and for research purposes only. Thank you

Ingrese aquí la clave

• Fecha actual en servidor: 26-03-2025

• Contenido alojado en servidor securizado:
<https://calentamientoglobalacelerado.net>

• Consulte posibilidades de acceso al administrador:
eurocamsuite@yahoo.es

• Tecnología **MARAFd** 2022©2025 de Código Abierto:
<https://calentamientoglobalacelerado.net/MARAFd/>

MARAFd

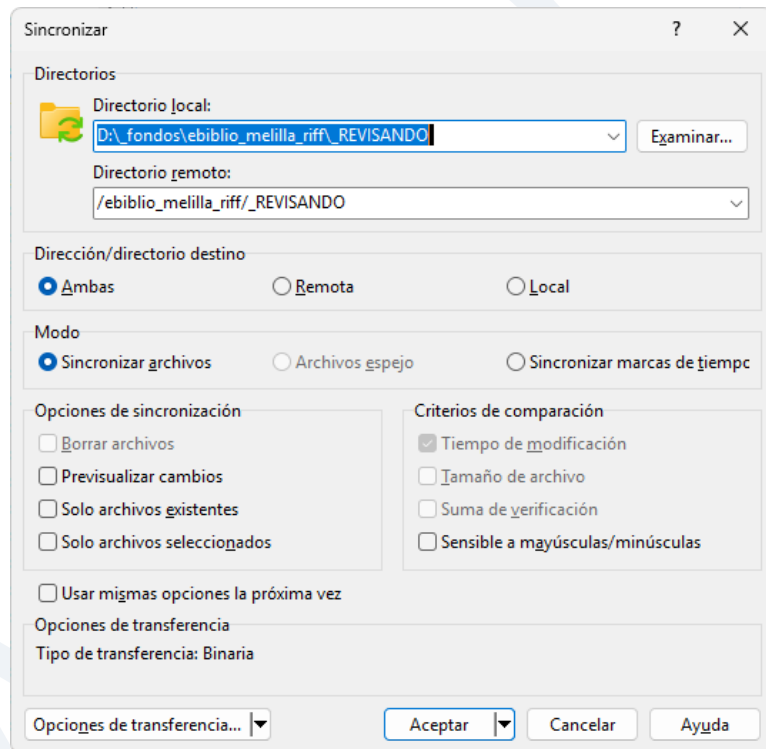
5. Flujos de trabajo

Revisión y normalización ajustada a lo dispuesto en el apartado anterior titulado *GESTIÓN DE DOCUMENTOS*

5.1. Definición de flujo operativo. Procesos manuales y automatizados en la gestión del fondo.

1. Recuperación y recopilación de archivos digitales de fuentes online o mediante digitalización (fotografía, escaneo, etc.)
2. Comprobación, revisión y adaptación (OCR, optimización del tamaño, etc.) del contenido digital para su inclusión en el fondo (derechos de autor, procedencia, datación, título, autor, etc.)
3. Renombrado descriptivo y selección de carpeta en función del tipo de contenido y respetando la limitación teórica de 2 KB para el *path* completo del archivo, comprendiendo éste la ruta, nombre y extensión del mismo.
4. Subida del fichero al servidor con cliente FTP (*Filezilla, WinSCP*, etc). Aunque ésta es una tarea manual, es muy recomendable realizar la subida de archivos desde el equipo local al servidor mediante el sincronizado de carpetas para mantener siempre el *mirror* exacto o copia espejo del fondo documental en el servidor y en el equipo local. Para llevar a cabo el sincronizado automático de carpetas, debemos configurar antes nuestro software cliente *FTP* realizando algunas pruebas para ello.

Si bien la sincronización de carpetas no es una tarea baladí, pues puede devenir en accidentes graves de pérdida de información si no se hace correctamente, personalmente utilizo *WinSCP* para esta labor con resultados satisfactorios hasta la fecha.



5. Refresco, generación y adaptación del índice-tesauro y subida automática de estos índices al servidor. Este proceso está automatizado por completo con *Batch-Scripts* y llamada a pequeñas herramientas ejecutables muy sencillas, las cuales se encargan de adaptar (*TIA.bas* > *TIA.exe*) y compactar (*compactTIA.bas* > *compactTIA.exe*) el índice-tesauro original (generado en el fichero **index_fondo_Melilla_Rif.asc** mediante el *scriptbatch* principal **indizar.cmd**) para un mejor procesamiento del mismo por parte del agente de inteligencia artificial. Ambos programitas, llamados directamente desde el *batchscript* **indizar.cmd** como digo, no son más que pequeñas rutinas encargadas de depurar y adaptar el índice-tesauro contenido en el fichero *index_fondo_Melilla_Rif.asc* y generar nuevos índices a partir de éste (*indice-tesauro.txt*, *indice-tesauro.csv* en el caso de *TIA.exe* ; *indice-tesauro_compacto.txt* e *indice-tesauro_compacto2.txt* en el caso de la rutina *CompactTIA.exe*) para lograr así un manejo más eficiente de ellos por parte de herramientas externas como son la herramienta de '**Búsqueda rápida**' en tesauro (código PHP ejecutado en el lado servidor) y otras como el propio '*Sistema Experto MARAFd-Melilla-Rif*' basado en un *MyGPT*® de *OpenAI* meticulosamente entrenado y el cual procesa por completo el tesauro para poder explorar, a partir de éste, el contenido completo del fondo documental. Este compactado del tesauro es fundamental. Piense que en el caso del *Fondo Documental Digital Melilla-Riff* que nos

ocupa aquí como ejemplo de referencia, el compactado del tesoro consigue reducir el tamaño de éste a los 500 KB, un peso aun manejable para el modelo *MyGPT*® que empieza a fallar al superar dicha cantidad (utilizando el motor *GPT-4o* y *GPT-4.5*).

6. Ante la necesidad de uso de nuestro índice-tesoro por parte de software de terceros u otros agentes de IA que así pudieran requerirlo, la *toolkit* **tesoro_json.py** ha sido escrita al objeto de convertir nuestro fichero **indice-tesoro_compacto2.txt** en un fichero estándar **.json** , tomando como dirección base el dominio o dirección de alojamiento de nuestro fondo documental para formar hiperenlaces o *urls* absolutas y funcionales.

5.2. Distribución de tareas en el flujo de trabajo.

Si bien los cometidos señalados en el apartado precedente como 1, 2 y 3, podrían llevarse a cabo perfectamente por cualquier usuario, las tareas reseñadas como 4 y 5 deben ser encomendadas en todo caso al administrador/es del fondo documental.

5.3. Posibilidad de coadministración o cooperación entre varios administradores.

Aunque en principio la figura del administrador del fondo puede recaer en una única persona, resulta viable la posibilidad de compartir dicha labor entre varias personas, ya sea sobre un mismo equipo local o incluso desde distintos equipos, siempre que sean acordadas las verificaciones oportunas sobre el contenido del fondo por parte de los administradores al objeto de evitar duplicidades o borrado accidental de archivos.

Este escenario no ha sido experimentado por el momento.

5.4. Seguimiento y control: Herramientas para supervisar el progreso de los flujos establecidos.

Aunque es el administrador quien debe practicar las verificaciones oportunas sobre el funcionamiento del fondo y los sistema de búsqueda, las pruebas de accesibilidad sobre los ficheros puede ser también efectuada y/o reportada por cualquier usuario conectado en línea, pudiendo contemplarse para ello un canal adecuado que facilite este tipo de reportes (email, formulario de contacto, *WhatsApp*, teléfono, etc.)

6. Integraciones con software alternativo de gestión documental

La filosofía *Open Source* y el espíritu de **MARAFd** garantiza, como puede verse en el *Fondo Documental Digital Melilla-Riff* usado aquí como referencia, una sencilla, rápida y transparente integración con herramientas de software y otros servicios externos tales como, buscadores, agentes de IA, hojas de cálculo, etc.

Por ello, y a pesar de que dicho escenario no ha sido experimentado, teóricamente, **MARAFd** no debería interferir en absoluto con el uso simultáneo de otro software externo dedicado a la gestión documental, al menos en principio.

7. Mantenimiento y soporte.

El mantenimiento y soporte de un fondo documental gestionado por **MARAFd** es una tarea sencilla que puede ser gestionado por una única persona y no debería suponer ningún problema. No obstante, toda vez configurado el fondo documental y plenamente funcional, el administrador deberá tener cierta experiencia en relación con el alojamiento web y el manejo de conexiones FTP ya que será el encargado de hospedar los archivos en el servidor y de ejecutar las tareas de actualización del índice-tesauro.

En caso de no contar con dicha cualificación, ésta puede lograrse a través de un periodo formativo de pocos días para garantizar un correcto y continuo funcionamiento del fondo.

Por otro lado, el permanente y completo soporte on-line ofrecido por el autor a través del site oficial <https://calentamientoglobalacelerado.net/MARAFd/> debería ayudar suficientemente a personas con cierto grado de conocimientos para poder implementar un sistema **MARAFd** completo desde cero.

7.1. Actualizaciones del sistema: Proceso para mantener el software actualizado.

Al tratarse de un ecosistema opensource, el propio administrador podrá ir actualizando y mejorando de forma continua las distintas herramientas que conforman el ecosistema y que alcanza a tecnologías diversas como pueden ser:

- la versión de *PHP* que correrá el servidor (esto debe hacerse normalmente a través del panel de control que ofrece el proveedor de servicios del hostingweb)
- la versión de *Python* instaladas en el sistema local para la ejecución de toolkits
- la versión del software cliente FTP utilizado para las conexiones al servidor (en nuestro caso de referencia *WinSCP*)
- la versión del editor ASCII empleado para la edición del código, en nuestro caso *NotePad++*
- la versión de librerías de *bootstrap* contenidas en las carpetas */css* y */js* alojadas en el servidor
- la versión del compilador *GCC* empleado en la compilación de los ficheros de código fuente *.bas*

7.2. Resolución de problemas comunes: Guía para solucionar incidencias frecuentes.

asdasd

7.3. Contacto y soporte técnico: Información de contacto para recibir asistencia adicional.

El canal de contacto oficial para el soporte técnico de **MARAFd** se ofrece a través del site oficial accesible en la dirección: <https://calentamientoglobalacelerado.net/MARAFd/>

8. Buenas prácticas.

8.1. Recomendaciones de uso: Consejos para aprovechar al máximo las funcionalidades de **MARAFd**.

Consulte con el autor en el site oficial:

<https://calentamientoglobalacelerado.net/MARAFd/>

8.2. Casos de uso: Ejemplos prácticos de cómo aplicar **MARAFd** en diferentes escenarios con resultados satisfactorios.

Escenarios/proyectos actuales con tecnología **MARAFd** implementada:

- *Fondo Documental Digital MELILLA-RIF* : Con más de 9 mil ficheros multimedia ya indizados y un volumen total de datos por encima de los 100 GB de información digital accesible en línea, este fondo se ha convertido en un referente para la investigación y la divulgación sobre la historia y la cultura de Melilla y el Rif.

https://calentamientoglobalacelerado.net/fondo_documental_Melilla_Riff/

- *Fondo Documental Digital sobre PROTOHISTORIA E HISTORIA ANTIGUA DEL MEDITERRANEO OCCIDENTAL* : Aplicando la misma tecnología de indizado y búsqueda, este fondo de ACCESO RESTRINGIDO ofrece un vasto repertorio de documentos relativos a las primeras civilizaciones y culturas del Mediterráneo Occidental, facilitando la investigación académica y la consulta especializada de una manera ágil, sencilla y exhaustiva.

https://calentamientoglobalacelerado.net/ebiblio_historia_antigua_01/

ANEXOS

ANEXO 1. GLOSARIO DE TÉRMINOS

TÉRMINO	Definición (...y licencia)	USO en MARAFd
Batch Scripting	<p>Batch Scripting para entornos Microsoft Windows® (en la guía puede aparecer como script-batch o batch-script indistintamente) es un lenguaje de procesamiento por lotes para automatizar tareas en terminales de uso común en automatización, backups, tareas repetitivas y gestión de sistemas.</p> <p>En entornos Windows, el ecosistema Batch es cerrado en tanto en cuanto depende de cmd.exe, que no es open-source. No obstante los scripts creados, como es el caso de MARAFd, pueden ser libres aunque su ejecución dependerá siempre del entorno propietario o abierto donde se ejecuten.</p>	<p>Se ejecuta en entornos Windows® 10/11 y es utilizado en la generación del índice tesoro del fondo documental completo mediante inidizar.cmd así como en el proceso de backup incremental del fondo (mediante llamada a comando robocopy.exe)</p>
Bootstrap	<p><i>Bootstrap</i> es un framework front-end open-source para diseñar sitios web responsivos y móviles. Incluye componentes CSS/JS predefinidos (grids, formularios, modales). Usado en desarrollo rápido de interfaces, prototipado, aplicaciones web y proyectos que requieren diseño consistente en distintos dispositivos.</p>	<p>Usado para proporcionar un aspecto más dinámico al frontend mediante el uso de efectos.</p>

TÉRMINO	Definición (...y licencia)	USO en MARAFd
CSS	Estándar del W3C y libre de derechos , <i>CSS</i> (Cascading Style Sheets) es un lenguaje para estilizar páginas web. Define diseño, colores, fuentes y responsividad. Separa contenido (HTML) de presentación. Usado en personalización de interfaces, diseño responsive, animaciones, frameworks (Bootstrap), y adaptación a distintos dispositivos o navegadores.	Usado para proporcionar un aspecto más dinámico al frontend mediante el uso de efectos.
GCC	<i>GCC</i> (GNU Compiler Collection) es un conjunto de compiladores open-source (licencia GPL) para lenguajes como C, C++, <i>FreeBasic</i> y Fortran. Permite optimizar y generar código ejecutable en múltiples plataformas. Usado en desarrollo de sistemas operativos (Linux), software embebido, aplicaciones críticas en rendimiento y entornos académicos o empresariales.	Usado para compilar el código fuente de las toolkits TIA.bas y compactTIA.bas escrito en <i>FreeBASIC</i> y generar código ejecutable altamente eficiente. Puede producir código ejecutables multiplataforma.
HTML5	Estándar del W3C y libre de derechos , <i>HTML5</i> es la última versión del lenguaje HTML, enfocada en semántica y soporte multimedia. Introduce elementos como <video>, <canvas>, y APIs (geolocalización, almacenamiento offline). Usado en desarrollo web/móvil, aplicaciones interactivas, juegos, sitios responsivos y PWA (Progressive Web Apps).	Usado para proporcionar un aspecto más dinámico al frontend mediante el uso de efectos.
JavaScript	<i>JavaScript</i> es un lenguaje de programación versátil y open-source , usado principalmente en desarrollo web para interactividad en navegadores. Con Node.js, también server-side. Común en front-end (React, Angular), back-end, apps móviles (React Native), APIs, juegos y aplicaciones dinámicas. Esencial en SPAs y desarrollo full-stack (frontend + backend)	Usado conjuntamente con CSS para proporcionar un aspecto más dinámico al frontend mediante el uso de efectos (bootstrap).

TÉRMINO	Definición (...y licencia)	USO en MARAFd
<p>PHP</p>	<p><i>PHP</i> es un lenguaje de scripting server-side, open-source, diseñado para desarrollo web. Dinámico, se integra con HTML. Usado en creación de páginas web, manejo de formularios, conexión con bases de datos, CMS (como WordPress), e-commerce y aplicaciones mediante frameworks (Laravel, Symfony).</p>	<p>Buscador rápido y generación de página HTML con contenido completo del Fondo Documental.</p>
<p>Python</p>	<p><i>Python</i> es un lenguaje de programación interpretado, open-source, multiplataforma y de sintaxis clara. Versátil, se usa en desarrollo web (Django, Flask), ciencia de datos (Pandas), IA (TensorFlow), automatización, scripting y educación. Ideal para prototipado rápido y proyectos con enfoque en legibilidad y productividad. Su licencia, <i>Python Software Foundation License</i> (compatible con GPL), permite:</p> <ul style="list-style-type: none"> • Uso, modificación y distribución libre, incluso en proyectos comerciales. • Herramientas asociadas (Cython, bibliotecas como Django) también suelen ser Open Source (MIT, BSD, etc.). • Promueve la colaboración y transparencia, clave en su adopción masiva. Ejemplo: Puedes modificar el intérprete de Python o crear derivados sin restricciones de licencia propietaria. 	<p>Usado en los toolkits localT.py (verificación del índice-tesauro a nivel local) y tesauro_json.py que adapta el tesauro al estándar JSON.</p>

**ANEXO 2. TABLA detallada de componentes del ECOSISTEMA
MARAFd aplicado al caso de referencia
Fondo Documental Digital Melilla-Riff**

Componente	Ubicación	Descripción	Ficheros resultantes
compactIA.exe (compactIA.bas)	Local (carpeta raíz contenedora del fondo documental en local)	Toolkit que reduce y compacta el fichero indice-tesauro.txt basado en config leída del fichero compactIA.asc El ejecutable .exe se obtiene compilando la primera vez el archivo fuente .bas mediante el compilador <i>GCC</i> incluido en MARAFd.zip	indice-tesauro_compacto.txt indice-tesauro_compacto2.txt cTIA_tesauro_report.txt
indizar.cmd	Local (carpeta raíz contenedora del fondo documental en local)	Script-batch de Microsoft Windows para generar el tesauro y llamada a programas externos TIA.exe y compactIA.exe (para adaptar y compactar el índice) y a winscp.exe para subida a servidor por FTP (leyendo config desde upload_script.txt	index_fondo_Melilla_Rif.asc numregs.dat
localt.py	Local (carpeta raíz contenedora del fondo documental en local)	Toolkit escrita en <i>Python</i> para verificar el índice-tesauro original en modo local. Lee el fichero <code>index_fondo_Melilla_Rif.asc</code> y permite verificar el enlace local abriendo el fichero con la aplicación asociada del sistema operativo. Incorpora sencillo motor de búsqueda por palabra.	No genera ficheros. Solo permite verificar enlaces locales a documentos del fondo.

Componente	Ubicación	Descripción	Ficheros resultantes
tesauro_json.py	Local (carpeta raíz contenedora del fondo documental en local)	Toolkit escrita en <i>Python</i> para convertir el indice-tesauro_compacto2.txt al formato estándar .json añadiendo el dominio base del repositorio del fondo para poder generar hiperenlaces funcionales. Lee el fichero indice-tesauro_compacto2.txt y escribe indice-tesauro_compacto2.json	indice-tesauro_compacto2.json
TIA.exe (TIA.bas)	Local (carpeta raíz contenedora del fondo documental en local)	Toolkit de análisis adaptación y normalización del fichero <code>index_fondo_Melilla_Rif.asc</code> creado por indizar.cmd y escritura de nuevos ficheros resultados y reporte. El archivo de reporte nos ayuda a detectar irregularidades en nombres de carpetas y/o archivos del fondo. El ejecutable .exe se obtiene compilando la primera vez el archivo fuente .bas mediante el compilador <i>GCC</i> incluido en MARAFd.zip	indice-tesauro.txt indice-tesauro.csv TIA_tesauro_report.txt
upload_script.txt	Local (carpeta raíz contenedora del fondo documental en local)	Archivo de configuración y comandos para conexión a servidor por FTP y subida de los índices.	No produce ficheros, solo pasa comandos FTP al software winscp.exe para subir ficheros indice al servidor.

Componente	Ubicación	Descripción	Ficheros resultantes
<p>../css</p> <p>../js</p>	<p>Servidor (carpeta del fondo documental o la elegida para albergar el frontend)</p>	<p>Carpetas con librerías <i>css</i> y rutinas <i>js</i> utilizadas por el estándar bootstrap necesarias para <i>frontend</i></p> <p>Cuando creamos código mediante IA generativa solo debemos informar a la IA la existencia de dichas carpetas para que ésta utilice <i>bootstrap</i> en sus creaciones.</p>	<p>No genera ficheros</p>
<p>index.php</p>	<p>Servidor (carpeta del fondo documental)</p>	<p>Código recursivo ejecutado desde el servidor web y que genera la página <i>HTML</i> con todas las carpetas y archivos del fondo documental. La página generada ofrece hiperenlaces a todos los documentos.</p> <p>Si ubicamos el <i>frontend</i> en la misma carpeta raíz del servidor en la que se alberga el fondo, deberíamos renombrar este fichero para evitar conflicto con index.html</p>	<p>lista_archivos.html</p>
<p>finder3.php</p>	<p>Servidor (carpeta raíz del fondo o carpeta elegida para contener el frontend)</p>	<p>Buscador rápido sobre índice-tesauro que genera reportes</p>	<p>No genera ficheros. Solo rastrea el tesauro y reporta resultado con hiperenlaces.</p>