



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 avanzados y administradores

Revisión: **0.99s** – 07 /09/2025

Website oficial: <http://MARAFd.net>

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



*“El mundo, y muy especialmente las administraciones públicas, necesitan urgentemente una solución real de digitalización masiva y **MARAFd**, como gestor de fondos digitales de segundo nivel situado directamente entre el usuario y el servicio básico de hosting-web y sin ninguna capa de software intermedia, puede ser esa solución.”*

Dominio oficial de la tecnología **MARAFd**
www.MARAFd.net

Resumen: **MARAFd** puede definirse como un gestor de fondos digitales de segundo nivel situado entre el sistema operativo del usuario y el servicio básico de *hosting-web* sin ninguna capa software intermedia. Sin necesidad de la hiperestructuración propia de los sistemas de bases de datos documentales, priorizando la preservación documental y una accesibilidad en línea permanente e inmediata al cien por cien del contenido indizado, **MARAFd** respeta los principios de procedencia (origen) y de orden (cronológico) como pilares fundamentales de la archivística.

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

Junto a la posibilidad de coexistir sin conflictos con otro tipo de soluciones software alternativas para la gestión documental digital, su carácter abierto y *Open Source* garantizan una rápida y sencilla integración con todo tipo de 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 acceder al contenido del fondo interactuando con un modelo de inteligencia artificial entrenado **MyGPT®**.

La presente guía, dirigida a administradores y usuarios avanzados que necesiten exprimir todo el potencial de esta tecnología, analiza en profundidad todos sus aspectos técnicos, además de sus ventajas propias y casos de aplicación real, así como la relevancia que su adopción puede suponer en la preservación del patrimonio digital y la investigación en cualquier ámbito de aplicación.

Palabras clave: *Gestión inteligente fondos documentales digitales, digitalización, patrimonio documental, archivos, código abierto, tesauro automatizado, transición digital, escalabilidad, preservación digital.*

Abstract: **MARAFd** can be defined as a second-tier manager of digital collections positioned between the user's operating system and the basic web-hosting service, with no intermediate software layer. Without the hyper-structuring typical of document-database systems, and prioritizing documentary preservation and permanent, immediate online access to one hundred percent of the indexed content, **MARAFd** adheres to the archival principles of provenance (origin) and order (chronological) as fundamental pillars.

Technically, **MARAFd** is an open-source ecosystem technology designed to manage digital archival collections through an automated thesaurus index that primarily leverages operating-system resources for file indexing. Its architecture—based on open standards such as *PHP*, *BatchScripting*, *Python*, *SFTP*, *JavaScript*, and *HTML5*—eliminates any dependency on complex and costly traditional database systems, prioritizing and guaranteeing, as noted, documentary preservation as well as accessibility, scalability, and, in particular, the maximum reduction of implementation and maintenance costs. The thesaurus index, generated from the folder structure and file names themselves, enables instant searches and nimbly handles information volumes on the order of several terabytes or millions of objects, with no limit other than that imposed by the capacity of the contracted web server (web hosting).

Alongside the ability to coexist without conflicts with other alternative software solutions for digital document management, its open and open-source nature ensures quick and simple integration with all kinds of tools in the field of artificial intelligence. This is demonstrated by already operational, real-world cases such as the documentary collection on *Melilla and the Rif*, where users can access the collection's content by interacting with a trained **MyGPT®** AI model.

This guide, aimed at administrators and advanced users who need to get the most out of this technology, provides an in-depth analysis of all its technical aspects, its inherent advantages, and real application cases, as well as the significance that its adoption can have for the preservation of digital heritage and research in any field of application.

Keywords: *intelligent management of digital archival collections, digitization, documentary heritage, archives, 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 principalmente a administradores potenciales que requieran de una herramienta rápida, accesible y gratuita para la gestión de fondos digitales y grandes repositorios de archivos en línea, así como usuarios avanzados que necesiten exprimir todo el potencial que esta herramienta puede ofrecer en el área de la investigación y/o académica.

1.2. ¿Qué es **MARAFd** ?

La transición digital ha exacerbado la necesidad de soluciones innovadoras en el ámbito de la digitalización capaces de gestionar un crecimiento exponencial e imparable de datos que puede acabar colapsando a muchas organizaciones. No podemos obviar que, en este aumento imparable de información digital convergen, por un lado los objetos nacidos digitales (creados ya en soporte digital) y por otro los productos de un proceso controlado de digitalización, lo cual nos conduce a una situación de colapso aún más complicada si cabe. En este delicado contexto y ante este descomunal desafío, **MARAFd** emerge como una respuesta disruptiva que pretende situar al fondo documental digital temático como un centro de información integral, capaz de abrazar prácticamente cualquier área del conocimiento humano y científico y con posibilidades reales de abrir la puerta a ese poder infinito que es el acceso pleno y libre a la información.

En este sentido, no es en absoluto descabellado afirmar que **MARAFd** rompe literalmente los esquemas conceptuales de la archivística digital tradicional al combinar una simplicidad técnica inusitada con una eficacia contundente en la organización y accesibilidad a la información.

En un plano meramente práctico podemos afirmar que **MARAFd** permite:

- Trabajar de forma segura y ágil, sin intermediación de software y bajo un control permanente del usuario, con miles o millones de documentos y una estructura comprensible de carpetas y archivos siempre accesibles en línea.
- Garantizar la preservación constante de repositorios completos de forma nativa, sencilla y automática al trabajarse siempre con una copia (mirror local) exacta del repositorio completo en el ordenador local y otra en línea, además de posibles *backups* automatizados complementarios implementados por el propio usuario.
- Automatización de procesos como la generación del índice-tesauro (columna vertebral del fondo) y otros flujos de trabajo mediante sencillos y archivos editables *BatchScript* (ficheros **.cmd** ó **.bat**),
- Mecanismos de búsqueda rápida y/o avanzada con acceso inmediato a todos los archivos indizados mediante sencillas rutinas *PHP* que lanzan búsquedas en el índice-tesauro alojado en el servidor y generan automáticamente hiperenlaces a los objetos digitales, así como otras soluciones software para el rastreo local profundo dirigidas a la investigación tales como *Adobe Acrobat Reader*, *Filelocator*, etc.
- Costes de instalación y mantenimiento asumibles por cualquier entidad así como una dificultad de implementación accesibles para cualquier usuario sin ninguna formación técnica específica.
- Integración y conexión de nuestro fondo documental con herramientas abiertas de inteligencia artificial que faciliten el análisis y la interpretación de datos y con accesibilidad al contenido del fondo mediante modelos abiertos de inteligencia artificial entrenados como *MyGPT®*.
- 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 un volcado del repositorio local sobre el nuevo servidor web.

Y algo no menos importante es que logra todo ello:

- Sin recurrir a complejos sistemas de software de base de datos ni a una engorrosa superestructura de metadatos.
- Sin pagar ni una sola licencia privativa.
- Sin necesidad de contratar personal técnico cualificado para su instalación ni externalizar el servicio de mantenimiento o crear dependencias con terceros.
- Sin costes estructurales fijos de soporte ni de mantenimiento, más allá de los derivados del servicio de alojamiento web (*hosting*).

Para ello **MARAFd** utiliza básicamente recursos nativos del propio sistema operativo (ya sea *Microsoft Windows®* ó *Linux*) que automatizan el indizado de ficheros y su puesta en línea.

Mediante un mismo **índice-tesauro** generado en local y compartido automáticamente con el servidor web, así como el uso de tecnologías abiertas como *HTML*, *JavaScript* para crear un *frontend* personalizado y dinámico, y sencillos guiones *PHP* (*backend*) ejecutados en el servidor que dotan al fondo de mecanismos de búsqueda y 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 en el nombrado de ficheros y la transferencia (que puede hacerse mediante sincronización de carpetas y archivos desde el software cliente *FTP*) entre el equipo local y el servidor.

Este tesauro es realmente un fichero de texto plano (*UTF8*) creado de forma automática en local mediante la ejecución de un sencillo *BatchScript* codificado en un fichero tipo **.cmd** y dada la análoga estructura de archivos y carpetas del fondo tanto en la copia local como en la remota (alojada en el servidor web), el tesauro es compartido y siempre idéntico en ambos lados (local-servidor). Podemos decir que ésta es, grosso modo, la idea principal sobre la que pivota el funcionamiento básico de un fondo gestionado mediante tecnología **MARAFd**.

También es fundamental comprender que, al prescindirse de cualquier sistema de base de datos en el funcionamiento del fondo, el **índice-tesauro** (y la propia estructura de carpetas) actúa como único y auténtico eje vertebrador del contenido. Sobre él se realizarán las búsquedas rápidas y a partir de él se generarán (de forma dinámica

mediante *PHP*) los hiperenlaces a todos los objetos digitales indizados. Dicho sea de paso, los objetos digitales indizados pueden ser de cualquier formato multimedia, no limitándose éste únicamente al formato documental (*PDF, DOC, JPG, etc.*) y permitiéndose siempre el indizado de cualquier formato de archivo imaginable. La accesibilidad a dichos ficheros multimedia desde el propio navegador de internet con el que accedamos al fondo dependerá siempre de la extensiones que este software permita abrir. En este sentido, archivos como por ejemplo *PDF, JPG, MP3, MP4*, podrán abrirse normalmente desde el navegador *Chrome, Mozilla*, etc. sin problemas.

Por otro lado, y redundando en todo lo anteriormente descrito, también podemos decir que el enfoque básico de la tecnología **MARAFd** se centra en un *mirror local-servidor* que comparten un mismo tesoro, a partir del cual podremos realizar las búsquedas y acceder a todos los objetos generando de forma dinámica los hiperenlaces a éstos.

Es de vital importancia, eso sí, constatar que las tradicionales fichas bibliográficas y los sistemas convencionales de bases de datos brillan aquí por su ausencia. Tampoco se recurre a un complejo sistema de metadatos, como suele darse en los sistemas convencionales de repositorios digitales, ya que todos los metadatos o información necesaria para localizar cualquier fichero del fondo se encuentran codificada en la propia cadena de texto que contiene el *path* completo del fichero, esto es, la ruta de carpetas más el nombre mismo del fichero, con una longitud máxima permitida para cada objeto indizado de unos 2 mil caracteres.

MARAFd prescinde así de cualquier tipo de dependencia software sirviéndose únicamente de los recursos que el propio sistema operativo ofrece de forma nativa para mantener perfectamente ordenado y bajo control absoluto cualquier repositorio sin otro límite que la capacidad de almacenamiento disponible en local y en *hosting web*.

Esta simplificación extrema en cuanto al **índice-tesoro** se refiere, lo convierte en una suerte de índice “ampliado”, en el sentido de que contiene metadatos o información codificada embebida en el propio nombre del archivo y la ruta de carpetas, ofreciendo así una mayor legibilidad y una accesibilidad inmediata a los ficheros ya sea buscando en las cadenas de texto del *path* de todos los objetos (contenidas en el **índice-tesoro**) o navegando directamente por la estructura de directorios del fondo (ver apartado “*Acceso al fondo completo mediante navegación por carpetas*” en esta misma guía).

Dicho **índice-tesauro**, como ya se apuntó anteriormente, es generado localmente y compartido por el cliente y el servidor de forma automatizada e inmediata mediante *BatchScript*, lo cual resuelve de facto cualquier tipo de incidencia en el proceso de indizado del fondo que pudiera afectar a la coherencia de la información.

Para poder comprender mejor este concepto basta pensar en el siguiente nombre como ejemplo:

fondo_04/publicaciones/ayuntamiento/2025_III_informe_catastral_definitivo_urbanismo_Girona_CA.pdf

Estas son las líneas que conformarán el tesauro de cualquier fondo. Fíjese que en realidad no necesita más información que la ofrecida por el propio nombre (*path* completo más nombre del archivo) para localizar un documento de forma inmediata a través de un simple motor de búsqueda de cadenas de texto. La consignación del año de elaboración del documento a modo de prefijo en todos los objetos a indizar, produce una sólida estratificación cronológica en el contenido indizado y realza el principio de orden (cronológico) propio de la ciencia archivística como uno de los pilares básicos en la catalogación de archivos.

Este modelo de catalogación, aunque pueda parecer banal e incluso limitado en algunos aspectos, supone un cambio drástico de paradigma que omite la hiperestructuración de datos y la vinculación del fichero a referencias externas en una base de datos, constituyendo realmente una solución disruptiva frente a los sistemas convencionales de gestión documental y ofreciendo ventajas considerables especialmente en cuanto simplicidad técnica y costos de infraestructura se refiere. En este sentido y por citar un ejemplo real, el fondo documental digital *MELILLA-RIFF* no es sino una prueba palpable de ello.

1.3. Marco conceptual y teórico de la tecnología y breve comparativa en su ámbito de aplicación

Como venimos apuntado, este sencillo y productivo modelo en el que un único **índice-tesauro** “ampliado” (con nombres de carpetas y archivos descriptivos) vertebrata y gestiona a la vez ambos repositorios (local y remoto, este último accesible en línea) redefine conceptualmente la estructuración y explotación de la información tal y como la conocemos escapando a la complejidad y costes inherentes de los tradicionales sistemas

dedicados a la gestión de fondos documentales digitales y liberando el patrimonio digital de dependencias externas que puedan acabar lastrando su preservación.

La preservación intergeneracional de repositorios digitales de patrimonio documental basados en software de gestión de base de datos documentales, puede suponer a menudo un proceso de enorme complejidad y costes elevados, factores ambos que, unidos, suelen acabar llevando al traste proyectos completos de digitalización.

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

Además, dados sus costes y simplicidad técnica extrema, **MARAFd** puede incluso erigirse como una tecnología puente o de transición en procesos de digitalización urgentes y ante futuras herramientas que probablemente lleguen de la mano de la inteligencia artificial u otras tecnologías emergentes. Tenemos que empezar a abordar el problema y tenemos que hacerlo con carácter urgente.

Cuadro 1: Comparación de MARAFd y sistemas convencionales

Característica	MARAFd	Sistemas Convencionales
Dependencia de base de datos	Ninguna; indizado nativo a partir de archivos y carpetas	Requieren motores SQL o NoSQL
Capas de software	Única capa entre usuario y hosting-web	Múltiples capas (middleware, servidores de BD, etc.)
Coste de licencias	0 €, código abierto CC BY-NC 4.0	Altos costes de licencias y mantenimiento
Escalabilidad	Millones de archivos sin hardware especial	Limitada por la arquitectura de la base de datos
Integración con IA	Soporta módulos MyGPT-MARAFd y futuros agentes de IA	Costosa de integrar

Figura 1: Tabla comparativa de tecnologías generada por modelo ia COPILOT® Think-Deeper

Tampoco podemos obviar que, valorando su equilibrio entre escalabilidad, preservación, accesibilidad y costes, su adopción puede ser viable tanto en el ámbito de fondos documentales de archivos públicos como privados, pasando por PYMEs, instituciones públicas, equipos científicos y multidisciplinares, etc.

En cuanto a la restricción de acceso público a la información, en caso necesario, podría también abordarse de forma sencilla mediante procesos de encriptación total o parcial de archivos y/o carpetas, además de un posible control de acceso al fondo mediante clave y control de usuarios, tal y como se ha hecho en el caso del *Fondo Documental de Protohistoria e Historia Antigua del Mediterráneo Occidental* cuyo acceso se halla restringido mediante *password*.

1.4. Uso de tecnologías estándares subyacentes abiertas.

El marcado carácter *Open Source* de **MARAFd** garantiza futuras mejoras y una fácil integración con nuevas herramientas (ver *toolkits*) y/o tecnologías disruptivas como *ChatGPT®*, etc., que nos permiten ampliar sus funcionalidades sin límites y, sobre todo, sin comprometer su esencia minimalista.

El principio *Open Source* de **MARAFd** nos lleva irremediablemente al uso obligado de herramientas libres y nos invita a una coexistencia transparente, dinámica y eficiente entre diversas tecnologías abiertas que acaban integrándose a la perfección en una suerte de ecosistema.

En esta línea, el ecosistema **MARAFd** hace un uso exclusivo de herramientas de desarrollo abiertas y software libre 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* pueden abrirse y modificarse con cualquier editor *ASCII* como *NotePad++*, etc.
- **BatchScripting:** Con un potencial a veces subestimado y otras desconocido, este recurso nativo de los sistemas operativos es capaz, a través de los llamados ficheros de procesamiento por lotes, de automatizar por completo la generación y sincronización del tesoro entre el equipo local y remoto, además de otros muchos procesos. Los ficheros *BatchScripts* , a igual que los *.PHP*, son editables con cualquier editor *ASCII* como *NotePad++*, etc.
- **HTML5/JavaScript:** la combinación de ambas tecnologías nos ofrece la posibilidad de implementar una web dinámica personalizada (*frontend*), adaptable y multiplataforma que posibilite el acceso total al fondo mediante cualquier navegador estándar y en

cualquier tipo de dispositivo (móvil o escritorio) y plataforma (*Android, Linux, Windows* y *Mac*) conectado a internet. Los ficheros con código *HTML/JavaScript*, al igual que en los casos anteriores, son editables con cualquier editor *ASCII* como *NotePad++*, etc.

- **Cliente FTP:** La transferencia de archivos y/o sincronización automática de carpetas para mantener ambas copias del fondo (*mirror*) puede encomendarse a cualquier software cliente *FTP* de libre distribución tales como *FileZilla*, *WinSCP* (recomendado), etc. Resumiendo un poco, el cliente *FTP* básicamente nos permite subir físicamente los ficheros desde el equipo local hasta el servidor. La sincronización de carpetas entre el equipo local y el *hosting web* facilita la integridad de las copias para que todas las carpetas contengan el mismo contenido exacto de ficheros.

- **Python:** utilizado para el desarrollo de sencillas aplicaciones (denominadas en nuestro caso como *toolkits*) ejecutadas en local, éstas pueden ayudarnos en tareas como la verificación del tesoro desde el equipo local (**localT.py**) y/o la adaptación del índice-tesoro desde un formato **.txt** al estándar **.json** (mediante **tesauro_json.py**) para un posible uso o tratamiento por terceras aplicaciones. Los ficheros con código *Python* son también editables con cualquier editor *ASCII* como *NotePad++*, etc.

- **FreeBASIC (v1.08.1)** con compilador **GCC-9.3.0**: Popular y potente lenguaje de programación utilizado para escribir y compilar las *toolkits* que, si bien serán analizadas en mayor detalle en apartado posterior “**Herramientas complementarias (Toolkits)**” , pasamos a citar aquí brevemente:

- **TIA.bas (TIA.exe)**: normaliza el tesoro original que tiene extensión **.asc** en formato **.txt** y **.csv** y además revisa el tesoro en busca de fallos o caracteres no válidos, generando un reporte con los fallos a revisar en el **ficheroTIA_tesoro_report.txt**
- **CompactTIA.bas (CompactTIA.exe)**: compacta los tesoros para reducir su peso, basándose en las premisas establecidas en el fichero **compactTIA.asc** en el cual se indican las extensiones de los archivos del fondo que deberán omitirse en el tesoro, y también la url base que será común a todos los objetos que conforman el fondo. Estas omisiones nos permiten reducir el tamaño del índice tesoro de forma significativa, acelerando los tiempos de subida (FTP) y el manejo del mismo por los buscadores u otras herramientas externas como por ejemplo el sistema experto *MyGPT®*.

Por tanto, aunque no puedan considerarse como esenciales, la finalidad de estas utilidades no es otra que estandarizar y compactar el índice-tesauro original (generado localmente en un fichero de extensión **.asc**) para facilitar su utilización por parte del modelo largo del lenguaje *MyGPT®*, u otras herramientas externas que deseen implementarse.

En nuestro caso hemos optado por este lenguaje y compilador por su alta velocidad de ejecución y sencillez de uso, entre otros motivos. Aunque como siempre es posible emplear interfaces *IDE* específicas para el desarrollo, los ficheros con código *FreeBASIC* son editables con cualquier editor *ASCII* como *NotePad++*, etc. y compilables fácilmente desde la línea de comandos.

- **ChatGPT®-MyGPT®**: El entrenamiento y ajuste de sistemas expertos basados en modelos largos de lenguaje como *MyGPT®* nos ofrece también un importante potencial a explotar de cara a la investigación académica y/o científica. En el caso de **MARAFd** podemos realizar, siempre limitado por la capacidad del modelo, un procesamiento inteligente del índice-tesauro del fondo para lograr así 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 y en la carpeta **docs** del fichero empaquetado **ecosistema-marafd-completo.zip** encontrará el archivo **sistema_experto-entrenado_config_MyGPT_IA.txt** con todas las instrucciones necesarias para entrenar un *MyGPT®* y especializarlo en consultas a cualquier fondo documental digital.

1.5. Estructura básica del ecosistema **MARAFd**.

Pasamos a ver ahora los distintos elementos que conforman el ecosistema de nuestra tecnología para tratar así de comprender mejor su funcionamiento.

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

De forma muy 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 y en el

servidor web. Esta característica intrínseca 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, la 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 *BatchScripting* 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* estamos codificando ya metadatos como el año, título, autor, idioma y tipo del documento, una información que nos permitirá localizar archivos en línea mediante un sencillo motor de búsqueda de cadenas implementado en *PHP* ejecutado en el lado del servidor.

Otra ventaja nada desdeñable que nos ofrece el tesauro, es que a partir de éste (generado en local y subido al servidor mediante un escueto *batchscripting*) y una sencilla rutina *PHP* ejecutada en el servidor web podemos generar una página dinámica *HTML* con la totalidad del contenido del fondo, clasificada por carpetas y con acceso directo a todos los ficheros indizados. Una web dinámica que permite una cómoda navegación por los directorios y ofrece acceso rápido al total de los archivos del fondo, todo ello en tiempo real. Esta opción se corresponde en la web principal del Fondo Documental Digital MELILA-RIFF con el menú **ACCESO A FONDO COMPLETO**.

Por otro lado, entre las restricciones técnicas a considerar del índice-tesauro podemos destacar las siguientes:

- Un límite máximo de **2000 caracteres** por nombre de archivo y ruta completa.
- La exclusión obligada en el nombre de los ficheros (y directorios) de símbolos y/o **caracteres especiales** o **acentuados** (con tilde) para lograr cierta homogeneidad en la información y mayor eficacia en las búsquedas.

1.5.3. Rutinas *PHP* (servidor) y *toolkits* locales.

Es importante entender que **MARAFd** funciona gracias a un conjunto de pequeño guiones y sencillas utilidades de código abierto engranadas de forma completamente transparente en lo que se podría denominar, por similitud, un ecosistema de minúsculas y sencillas rutinas de código abierto que logran su objetivo con eficacia.

En este sentido, además del fichero principal *BatchScripting* **indizar.cmd** existen en **MARAFd** otras pocas rutinas encaminadas a realizar diversas tareas. Este script de indizado se encarga, además de generar el índice-tesauro, ejecutar estas distintas herramientas (a las que denominaremos *toolkits*) que serán estudiadas en mayor detalle en posteriores apartados, y que logran aumentar notablemente el potencial y la conectividad de nuestra tecnología.

Estas *toolkits* principales, ejecutadas en local, son:

- **TIA.exe** (compilado de **TIA.bas**) : valida la integridad del tesauro y genera reporte de *warnings* para revisar,
- **compactTIA.exe** (compilado de **compactTIA.bas**) : encargada del compactado y generación de tesauros más ligeros (a partir del tesauro original) para poder ser procesados por agentes inteligentes, en nuestro caso **MyGPT®**,
- **tesauro_json.py** : adapta la estructura del índice-tesauro al formato *JSON*
- **localT.py** : de ejecución opcional, permite la verificación de enlaces del índice-tesauro en modo local.

Ejecutadas ya en el lado del servidor encontramos también pequeñas rutinas como **finder3.php** ó **index.php** , encargadas de facilitar las búsquedas en el índice-tesauro o generar la página *HTML* que permite la navegación por directorios a través del fondo completo.

Por otro lado, también encontraremos una *toolkit* como.

1.6. Flujo básico de trabajo para el añadido de archivos.

De forma muy resumida, podemos afirmar que el proceso básico necesario que el administrador debe ejecutar con cada subida de archivos al fondo es sencillo, seguro y rápido, dado que se encuentra prácticamente automatizado en su totalidad.

Dicho proceso consiste básicamente en:

- **Renombrado** de archivo/s según lógica descriptiva siguiendo criterios preestablecidos y copia de estos archivos nuevos en carpeta local.
- **Sincronización** con el servidor desde cliente *FTP* (mirror). Las herramientas para conexiones por *FTP* desde cliente incorporan funciones de sincronización de carpetas entre local y servidor y en nuestro caso estamos trabajando con el software gratuito *WinSCP*
- **Indizado** automatizado mediante ejecución de fichero *BatchScripting* local (**indizar.cmd**) que genera y actualiza el tesauro en local y en servidor además de otras tareas que se estudiarán en mayor detalle.

Estos tres pasos son los que deberán repetirse con cada nueva incorporación de archivos al fondo, que podrá realizarse archivo a archivo o de forma masiva, es decir, en bloques.

1.7. Algunos beneficios destacables derivados de usar **MARAFd** frente a otros sistemas de gestión documental digital

Aunque a priori el uso de **MARAFd** no tiene por qué ser incompatible con el uso simultáneo de otras soluciones especializadas en la gestión documental de fondos digitales y en teoría es posible que **MARAFd** coexista con otras tecnologías, es importante destacar en ella ciertas ventajas relevantes que la diferencian del resto, 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.

1.8. 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.

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 para evitar malos entendidos.

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 **.bas** podremos usar el compilador citado que encontramos en la carpeta **compiladores_toolkits** del paquete comprimido **ecosistema-marafd-completo.zip**

Por otro lado, en lo referente a transferencias al servidor y sincronización de carpetas, éstas pueden ser encomendadas con garantía a cualquier software libre de los muchos existentes, entre otros:

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

Por otro lado, en el uso de algunas herramientas complementarias que se han desarrollado para tareas específicas como es el caso de **localT.py** destinada a la carga, visualización y verificación del índice-tesauro desde el equipo local, ó **tesauro_json.py** que adapta el contenido del tesauro al formato estándar **.json** por si fuera necesario para un posible uso por parte de herramientas externas, se requiere para su ejecución la instalación en nuestro sistema del intérprete *Python* incluido en la carpeta **compiladores_toolkits** del paquete comprimido **ecosistema-marafd-completo.zip**

El *toolkit* **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 los *toolkits* **localT.py** y **tesauro_json.py** y al objeto de evitar posibles infecciones, incompatibilidades o problemas de otra índole, la descarga del instalador deberá hacerse siempre desde el servidor oficial:

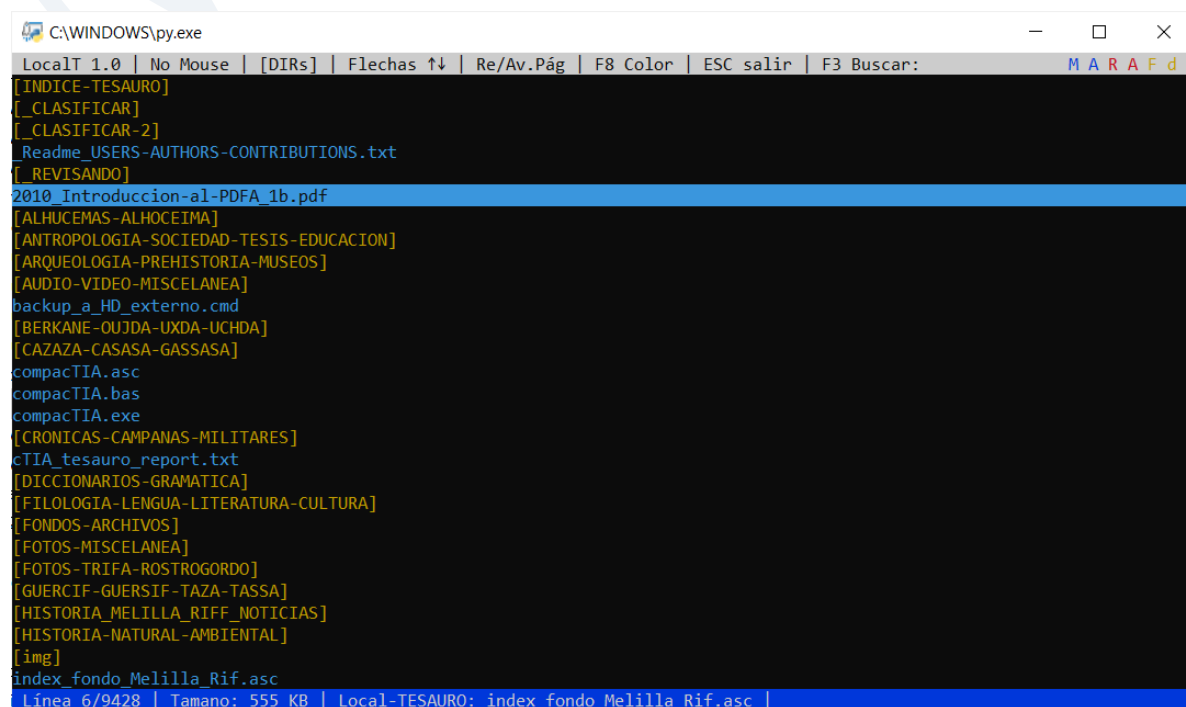
- Versión 3.12.2 del intérprete *Python*, disponible en la dirección:

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

- O bien recurrir a la incluida en la carpeta **compiladores_toolkits** del paquete comprimido **ecosistema-marafd-completo.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:
[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 | Tamano: 555 KB | Local-TESAURO: index_fondo_Melilla Rif.asc
```

2.1.3. Servicio *hosting-web*.

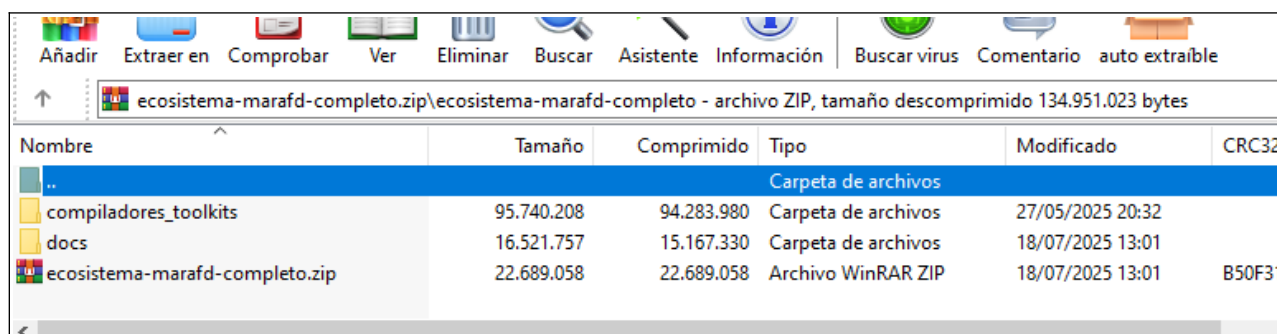
Para mantener un fondo documental digital en línea mediante la tecnología **MARAFd** resulta imprescindible tener contratado un proveedor de servicios de internet el cual nos ofrezca, junto al servicio de hospedaje web, la posibilidad de ejecutar código PHP directamente sobre el servidor.

El espacio de alojamiento necesario en servidor deberá ser suficiente para albergar todos los ficheros del fondo documental. Además, por un coste muy reducido, normalmente podremos contratar con el mismo proveedor la compra y el mantenimiento de un nombre de dominio para nuestro fondo del tipo www.fondodocumental.net o similar, logrando así una mejor visibilidad y accesibilidad a nuestro contenido, así como un indizado mucho más óptimo para los buscadores.

El coste total de estos servicios, incluido el nombre de dominio, suele oscilar entre 100 y 200 euros anuales para una capacidad de alojamiento en servidor de unos 200 GBytes, lo cual, al no requerirse ningún tipo de licencia privativa para la implementación de **MARAFd** ni de mantenimiento, supone un coste reducido y fácilmente asumible para cualquier entidad.

2.2. Instalación del ecosistema **MARAFd**

Una vez confirmadas las exigencias citadas en el apartado anterior, debemos descargar desde la web oficial del proyecto www.MARAFd.net el archivo empaquetado **ecosistema-marafd-completo.zip** que descomprimiremos en una carpeta de nuestro disco duro. Al descomprimir el mismo encontramos el contenido siguiente:



Nombre	Tamaño	Comprimido	Tipo	Modificado	CRC32
..			Carpeta de archivos		
compiladores_toolkits	95.740.208	94.283.980	Carpeta de archivos	27/05/2025 20:32	
docs	16.521.757	15.167.330	Carpeta de archivos	18/07/2025 13:01	
ecosistema-marafd-completo.zip	22.689.058	22.689.058	Archivo WinRAR ZIP	18/07/2025 13:01	B50F31

Este fichero contiene archivos que vamos a necesitar replicar en nuestras carpetas locales y también subir al servidor, y son los ficheros que debemos editar y adaptar a nuestro

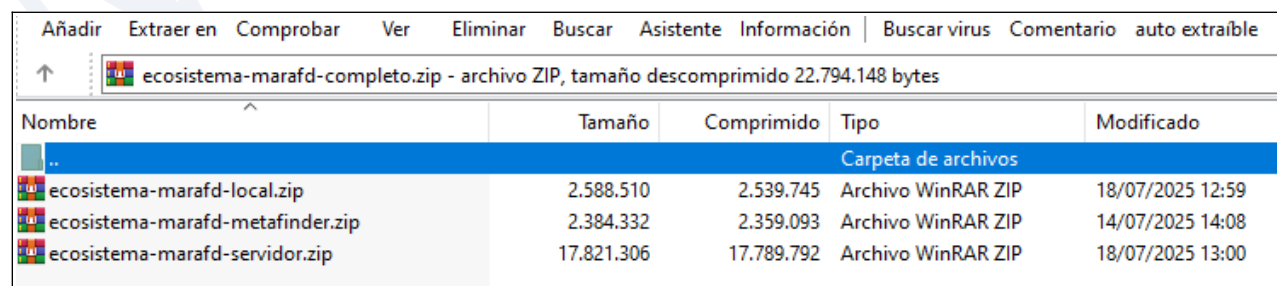
antojo (tanto en diseño como en funcionalidad) y a las rutas (*path*) absolutas y relativas (consignadas en el índice-tesauro) de nuestro fondo documental para que todo el sistema funcione. No desespere si no entiende algo pues lo vamos a explicar absolutamente todo y además, siempre puede lanzar su consulta en el foro público de esta tecnología abierto a tal fin y gestionado por los autores, accesible desde:

https://calentamientoglobalacelerado.net/foro_01/

La carpeta **compiladores_toolkits** contiene el compilador *FreeBASIC-GCC* así como la versión del intérprete *Python* utilizada para poder correr los programas-*toolkits* **.py**, ambos necesarios para poder adaptar y compilar los programas/rutinas **.bas** y ejecutar los **.py**

En la carpeta **docs** encontraremos abundante documentación acerca de la tecnología **MARAFd** incluida también la presente guía técnica, además de otros contenidos de interés sobre la misma que pretenden ayudarnos a comprender mejor como funciona todo el sistema.

El fichero comprimido **ecosistema-marafd.zip** contiene a su vez otros 3 archivos comprimidos que deberemos igualmente descomprimir en esa misma ubicación.



Nombre	Tamaño	Comprimido	Tipo	Modificado
..			Carpeta de archivos	
ecosistema-marafd-local.zip	2.588.510	2.539.745	Archivo WinRAR ZIP	18/07/2025 12:59
ecosistema-marafd-metafinder.zip	2.384.332	2.359.093	Archivo WinRAR ZIP	14/07/2025 14:08
ecosistema-marafd-servidor.zip	17.821.306	17.789.792	Archivo WinRAR ZIP	18/07/2025 13:00

Una vez descomprimido **ecosistema-marafd.zip** encontramos estos tres ficheros, y podemos decir que ahora sí contamos con todos los ficheros necesarios que debemos editar y adaptar para montar nuestro propio fondo, así como algunos ficheros de ejemplo que nos ayudarán a comprender mejor cualquier parte del proceso.

Resumiendo un poco:

- **ecosistema-marafd-local.zip** : Contiene los ficheros que deben albergarse, una vez editados y adaptados a las rutas de nuestro fondo documental, en el equipo local.

- **ecosistema-marafd-servidor.zip** : Contiene los ficheros que deberán adaptarse igualmente a las rutas de nuestro fondo documental y subirse al servidor, a la carpeta donde se ubicará el frontend-backend (web principal y buscador) de nuestro fondo.
- **ecosistema-marafd-metadinder.zip** : Contiene los pocos ficheros que conforman el metabuscador o metafinder, que permite la conexión de múltiples fondos documentales para poder rastrear y buscar documentos en todos ellos a la vez.

Antes de comenzar a trabajar, es importante comprender como se estructura el fondo y qué archivos vamos a necesitar modificar, tanto a nivel local (en nuestro equipo) como en el servidor. Piense que los ficheros que serán alojados en el servidor también serán editados en modo local y subidos después mediante el software cliente *FTP*.

Como veremos, son pocos los ficheros que deberemos configurar para adaptar y hacer funcionar nuestro fondo documental. Para ello he tomado como ejemplo un fondo ya plenamente funcional y que es el *Fondo Documental Digital MELILLA-RIFF* el cual ya se encuentra en pleno funcionamiento y es accesible en línea, de esta forma evitaré afirmaciones que podrían estar equivocadas o no probadas.

Con un número de archivos superior a los 9000 ficheros y un volumen total de más de 110 *GBytes* de información en el momento de elaboración de este manual, la estructura de este fondo documental es plenamente adaptable a nuestro proyecto y eso es lo que vamos a tratar de explicar con detalle a continuación.

2.2.1. Estructura básica de carpetas del paquete.

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

ecosistema-marafd-completo.zip (paquete descargado)

- └ **compiladores** (*FreeBasic GCC-9.3.0, Python 3.12.2*)
- └ **docs** (biblioteca **MARAFd**, manuales y otros recursos de ayuda)

ecosistema-marafd.zip

- └ **ecosistema-marafd-local** (ficheros a alojar en equipo local)
 - └ **ecosistema-marafd-servidor** (ficheros del *frontend-backend* a alojar en servidor)
 - └ **ecosistema-marafd-metafinder** (metabuscador a alojar en servidor. Opcional)
-

Este paquete, ya descomprimido, es recomendable que quede siempre en nuestro disco duro sin modificaciones, pues se trata de los ficheros maestros que serán editados y adaptados para hacer funcionar nuestro fondo, por lo que trabajaremos siempre sobre copias de los mismos y no sobre los originales. Aun así, en caso de pérdida o modificaciones fallidas o accidentales, siempre podemos recuperar el fichero original desde el archivo empaquetado **ecosistema-marafd-completo.zip**

Lo siguiente será montar la estructura de carpetas y archivos del fondo en nuestro equipo local, si no la tenemos ya montada, para después transferirlo todo directamente al servidor desde el software cliente *FTP*. Esta transferencia puede hacerse una vez que hayamos editado y actualizado las rutas y enlaces establecidos en los ficheros *BatchScripting* y comprobado que los índices-tesauros del fondo se generan correctamente en nuestro equipo local.

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. Lo importante es que siempre debemos mantener copia-mirrored exacto de las carpetas en el sistema local y en el servidor.

Aunque es perfectamente factible alojar todos los ficheros *HTML* y *PHP* que conforman el *front-back-end* en el mismo directorio que contiene la estructura completa y los archivos del fondo, en este caso he optado, y así lo recomiendo, por albergar los archivos *HTML*, *PHP* y las carpetas con las librerías de *bootstrap* (*/css* y */js*) del *frontend* en la carpeta

fondo_documental_Melilla_Riff distinta a la que alberga el propio fondo completo (entendido como todos los documentos que lo conforman) y que sería **ebiblio_melilla_riff** de manera que, a la hora de trabajar y adaptar el *frontend-backend* no mezclamos los ficheros del fondo en sí con los archivos de HTML y otros de nuestro *frontend-backend*. Si bien esta modularización (división) es opcional, parece bastante recomendable dado que separa y simplifica aún más la estructura del fondo.

Algo a tener siempre muy en cuenta serán las rutas (*path*) indicadas tanto en los ficheros **.php** como en los **.cmd** (guiones ejecutados localmente para el indizado del fondo y actualización de tesauros locales y también remotos) para que los hiperenlaces generados apunten correctamente a cada uno de los ficheros del fondo.

También debemos considerar los ficheros de imagen u otros elementos que componen el *frontend-backend* y que deberán ser alojados en la carpeta correspondiente de nuestro servidor para una correcta visualización del *frontend*, por ejemplo **/img/**

En cualquier caso, antes de iniciar el proceso de implementación del nuevo fondo documental mediante la edición y adaptación de los archivos ejecutables (guiones *batchscript* y rutinas *PHP*) es imprescindible que contemos ya con una copia exacta del fondo documental completo con una estructura de carpetas (aunque se puede modificar en cualquier momento), tanto en el servidor como en nuestro equipo local, y que ambas copias mantengan la misma estructura de nombres de carpetas, es decir, que los fondos (local y remoto) estén alojados en idénticas rutas relativas. Esto significa que la carpeta principal o directorio padre que alberga los subdirectorios y archivos del fondo, tanto en el disco duro local de nuestro ordenador como en el servidor remoto (servidor web) tengan exactamente el mismo nombre, solo de esta forma podremos luego generar los tesauros localmente mediante un simple *batchscript* (**indizar.cmd**), subirlos al servidor de forma automatizada y poder generar los hiperenlaces funcionales a los documentos a partir del archivo de texto **indice-tesauro**.

La siguiente imagen nos sirve de orientación o ejemplo a la hora de visualizar la estructura básica de las carpetas y ficheros del fondo, tanto a nivel local como del *hostingweb* (servidor web).

Carpetas y archivos en disco duro local (archivos de la web ¹ y fondo documental ²)	Carpetas y archivos en servidor web (archivos de la web ¹ y fondo documental ²)
<p>Réplica de la web (frontend-backend) del fondo en equipo local con librería de archivos <i>bootstrap</i> (<i>js</i>, <i>css</i>) en carpeta distinta a la del propio 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) ├── finder3.php (motor rastreo tesoro) ├── carrusel.html (web de carrusel visual) ├── ── ./js ├── ── ./css └── * * * * *</pre> <p>Réplica espejo del fondo documental completo en equipo local con nombre coincidente en el servidor para la correcta generación del tesoro 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 (carpetas_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 └── ── carpetas+archivos_fondo </pre>	<p>Réplica de la web (frontend-backend) accesible en línea del fondo en equipo remoto con librería de archivos <i>bootstrap</i> (<i>js</i>, <i>css</i>) en carpeta distinta a la del propio 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) ├── finder3.php (motor rastreo tesoro) ├── carrusel.html (web de carrusel visual) ├── ── ./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) ├── lista_archivos.html (generada por index.php para navegación por carpetas) ├── indice-tesoro.csv (creado en local) ├── indice-tesoro.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 └── ── 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 <i>js</i> y <i>css</i> 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 tesoro que se genera localmente.</p> <p>Los índice-tesoro, generado localmente y transferido al servidor mediante procesamiento por lotes automatizado (BatchScripting), 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.

```

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-OBRA-S-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 compactTIA.asc
26/03/2025 12:43 64.000 compactTIA.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-compactTIA.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 nuestro objetivo con solvencia y

velocidad suficiente, recurrimos al uso del comando **robocopy** integrado *Microsoft Windows 10/11®* y accesible 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 destino 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 proceso principal *batchscript indizar.cmd*. De esta forma, con cada nuevo indizado en el fondo documental estaremos realizando un proceso de respaldo automático.

2.4.2. Copiado de archivos en equipo local y sincronización con el servidor *hostingweb* mediante software cliente *FTP*.

La tarea de aportaciones de archivos al fondo es prácticamente el único proceso que debe efectuarse de forma manual por el administrador, que se encargará de alojar los ficheros en ambos lados, el equipo local y el servidor.

Este proceso consiste básicamente en, toda vez realizado el necesario e imprescindible proceso de renombrado descriptivo del/los fichero/s, copiar 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 *hosting-web* y situaremos ambos lados (cliente y servidor) en el directorio (carpeta) que vamos a sincronizar, efectuando dicha sincronización de forma automática o manual pero asegurándonos que ambas carpetas, la local y la remota, sean copias exactas tras el proceso de sincronizado.

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*). Esto garantiza la integridad del *mirror* (copia espejo de las carpetas/archivos) en todo momento. En cualquier caso, recomiendo familiarizarnos antes con la aplicación *FTP* que utilicemos para evitar borrados accidentales que pueden darse al sincronizar en la dirección equivocada.

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.

La generación y actualización del índice-tesauro es un proceso completamente automatizado mediante un guión *batch* (también se denominan *batchscript* o *scriptbatch*) que recae en el archivo por lotes **indizar.cmd**, dicho archivo de procesamiento por lotes recoge y ejecuta todos los procesos necesarios para el indizado y actualización del tesauro en local y remoto, además de realizar otras funciones auxiliares como revisiones y adaptaciones del tesauro y que citaremos más adelante en este manual.

El fichero que se muestra a continuación es un ejemplo que puede adaptarse a la ruta base (dominio o *url*) 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://calentamientoglobacelerado.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 **compactTIA.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-tesauro.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://calentamientogloballacelerado.net/ebiblio_melilla_riff/

2.6. Ejecución de *PHP* en lado del servidor.

El ecosistema **MARAFd** hace uso también de la tecnología abierta *PHP*, la cual se ejecuta en el lado del servidor.

A continuación veremos 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**) ubicado en el directorio raíz del fondo documental, 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.

El motor es muy sencillo, pero aún así está pulido para ignorar en cada búsqueda mayúsculas y minúsculas, acentos y otros símbolos ortográficos, además de otras sutiles cualidades en el plano fonético que permiten encontrar indistintamente términos escritos, por ejemplo, con Q y K, B y V, C y Z, etc., y pese a permitir como máximo dos palabras por búsqueda, dichas características dotan de gran flexibilidad al mecanismo de rastreo.

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

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.

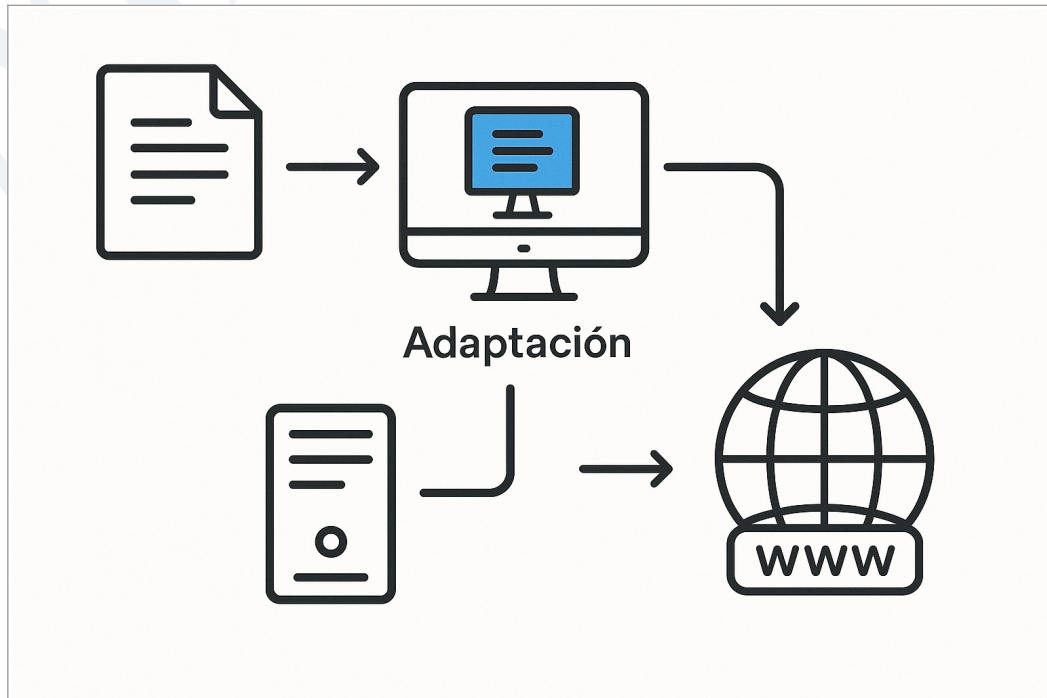
ENVIAR DOCUMENTO/s al fondo

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.

En esta sección podemos ofrecer meramente un enlace a texto informativo acerca del procedimiento básico a seguir para aportar archivos al fondo documental y como debería de enviarlos al administrador.

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.

Acceso a FONDO COMPLETO

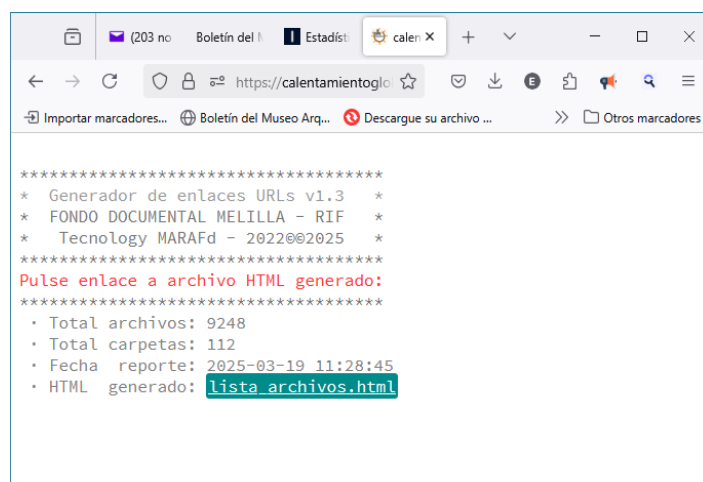
• Con el generador automático de enlaces en tiempo real podrá navegar a través de las carpetas del fondo documental completo, clasificadas* por temática y accediendo a sus más de **NUEVE MIL** documentos y archivos a golpe de Clic!

**LA ESTRUCTURA ACTUAL DE CARPETAS ES PROVISIONAL*

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 URLs* (**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. En realidad se trata de una página *html* generada de forma dinámica por el código *PHP* contenido en **index.php** y en el caso del *Fondo Documental Melilla-Rif*, dado el tamaño de este, es bastante extensa, por lo que a veces puede requerir un par de segundos de proceso antes de que podamos verla en la ventana de nuestro navegador.

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 en la página de forma inmediata, resaltando todos los resultados coincidentes.

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 (en tesauro)** 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!

El sistema de BÚSQUEDA RÁPIDA se encuentra optimizado permitiendo búsquedas flexibles en el tesauro del fondo

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 eurocansuite@aroba@yahoo.es o utilice el SISTEMA EXPERTO desde la web principal del fondo.

ej ANUAL (2 palab./word MAXIMO)

Buscar

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

✓ CRONICAS-CAMPANAS-MILITARES/ANNUAL-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 tecleado *Alucemas*, o documentos con la palabra *Casasa* habiendo tecleado *CAZAZA*, ó 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* (en plural) daría también como resultado positivo las apariciones en el tesoro en número singular, es decir, *medico*. Además, la letra *Q* se ha asimilado también 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-tesauro del fondo y nunca en el interior de los documentos contenidos en el mismo.

3.1.5. RASTREO PROFUNDO (investigación).

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.

Debemos saber 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 siempre puede 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.

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 (en fase experimental).

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. Resulta obvio que, en caso de optar por la implementación de esta opción, cada fondo documental requerirá un **MyGPT®** personalizado, pero podremos aprovechar la configuración ya establecida para el caso del *Fondo Documental Melilla-Rif* y que encontrará en el paquete descargado. Esta configuración, con las adaptaciones que procedan en su caso, y que deberá suministrar a su **MyGPT®** está contenida en el fichero **sistema_experto-entrenado_config_MyGPT_IA.txt** contenido en la carpeta \local del paquete

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 es el *PDF*, el *Sistema Experto MARAFd MyGPT®* en esta ocasión permite al usuario recabar información simplemente interactuando con la IA a través del *prompt* (caja de texto de entrada en la que tecleamos los mensajes al modelo) 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*.

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**

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®.

Lanzar



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 del fondo.

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 podemos entender que esta no es una tarea obligada para la preservación digital de documentos y de hecho existen grandes repositorios que parecen obviar 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.

Hemos de tener muy en cuenta que, a veces, tras aplicar un proceso de *OCR* a un documento *PDF*, el tamaño de éste puede engrosar de forma considerable, 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.

Igualmente, en algunos casos, el procesamiento *OCR* de un archivo *PDF* podría alterar la calidad de algunas imágenes con la consecuente pérdida de información, por tanto, es un proceso delicado al que siempre debemos prestar la debida atención.

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:calentamientoglobalacelerado.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 de forma sencilla 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



5. Flujos de trabajo.

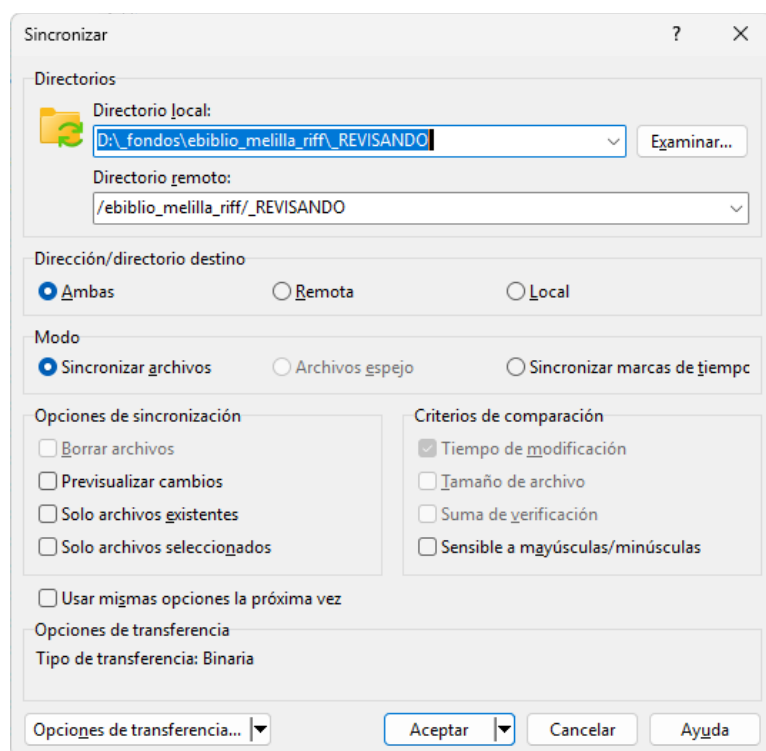
Revisión y normalización ajustada a lo dispuesto en el apartado anterior titulado *GESTIÓN DE DOCUMENTOS*

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 *BatchScriptings* 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 tesauro 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-tesauro por parte de software de terceros u otros agentes de IA que así pudieran requerirlo, la *toolkit* **tesauro_json.py** ha sido escrita al objeto de convertir nuestro fichero **indice-tesauro_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://calentamientoglobacelerado.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 corre el servidor (esto debe hacerse normalmente a través del panel de control que ofrece el proveedor de servicios del *hostingweb*)
- la versión de *Python* instalada 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 el popular *NotePad++*
- la versión de librerías de *bootstrap* contenidas en las carpetas **/css** y **/js** y que debemos situar en la carpeta que aloje el *frontend* en el servidor

- la versión del compilador GCC empleado en la compilación de los ficheros de código fuente editables y escritos en lenguaje *FreeBASIC* y con extensión **.bas**

7.2. Resolución de problemas comunes. Cómo solucionar incidencias frecuentes.

Puede utilizar para ello el [Foro público moderado para usuarios MARAFd](#) en el que los propios autores le ayudarán a resolver cualquier duda o problema que pudiera surgirle durante el proceso de implementación o tras él.

- https://calentamientoglobalacelerado.net/foro_01/index.php

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. METABÚSQUEDAS: Posibilidad de búsquedas simultáneas en diversos fondos documentales.

La posibilidad de búsquedas simultáneas en distintos fondos es una capacidad inherente de la tecnología **MARAFd** que incrementa de forma infinita su escalabilidad. Dicha función de metabúsqueda es un componente esencial que permite consultas simultáneas en múltiples fondos documentales digitales mediante la integración de sus índices-tesauro de una forma transparente, sencilla y muy rápida, sin escribir ni una sola línea de código. Esta funcionalidad transforma la gestión documental al unificar el acceso a repositorios distribuidos en diferentes instituciones, sin necesidad de migrar o duplicar datos. De forma completamente transparente y sin escribir una sola línea de código,

podemos integrar varios los fondos deseados para rastrear todos los tesauros a la vez en una sola búsqueda.

8.1. REQUISITOS PREVIOS PARA LA OPCIÓN DE METABÚSQEDAS.

Antes de nada necesitará disponer de un fondo documental digital operativo (y accesible desde internet) funcionando con la tecnología **MARAFd**. Si aún no es su caso o tiene cualquier duda al respecto, consulte con el autor por cualquiera de los canales disponibles en la web oficial www.MARAFd.net

8.2. DATOS NECESARIOS PARA AGREGAR CUALQUIER FONDO DOCUMENTAL.

Por cada fondo que quiera integrar deberá enviar un email al administrador de MARAFd indicando los siguientes datos (estos son de ejemplo):

- **Nombre del archivo tesauo:** 'index_fondo_historia_01.asc'
- **URL base donde se encuentra el fondo digital:** 'https://calentamientoglobalacelerado.net/ebiblio_historia_antigua_01/'
- **Nombre del Fondo Documental:** 'FONDO HISTORIA-01'
- **Nombre abreviado del Fondo Documental:** 'Historia Antigua 01'

El fichero del índice-tesauo del fondo a indizar debe ser siempre accesible en línea y hospedado en la URL base con el nombre indicado, solo así se podrá integrar el mismo en el metabuscador.

Si desea usted montar un metabuscador propio deberá respetar la siguiente estructura en el fichero 'fondos_agregados.csv' que se aloja en la carpeta '/metafinder' de su servidor.

```
[  
'file' => 'index_fondo_Melilla_Rif.asc',  
'base_url' => 'https://calentamientoglobalacelerado.net/ebiblio_melilla_riff/',  
'source_name' => 'Melilla-Rif',  
'title' => 'FONDO MELILLA-RIF'
```

```
]
,
[
'file' => 'index_fondo_historia_01.asc',
'base_url' => 'https://calentamientoglobacelerado.net/ebiblio_historia_antigua_01/',
'source_name' => 'Historia Antigua 01',
'title' => 'FONDO HISTORIA-01'
]
```

Estructura real del fichero 'fondos_agregados.csv' (con datos de ejemplo) que debe hospedarse en la misma carpeta '/metafinder' del servidor web en la que se encuentra ubicado el programa metafinder.php

Con la información almacenada en el el fichero 'fondos_adgregados.csv' manteniendo siempre la estructura aquí mostrada, el código PHP implementado en **metafinder.php** logra la búsqueda simultánea en los dos fondos reseñados. Añadir nuevos fondos documentales es, como ven, bastante sencillo aportando únicamente la dirección URL completa del índice-tesauro y la URL base de dicho fondo.

Supongamos por ejemplo que varios ayuntamientos de una provincia han implementado ya su fondo documental digital y ahora desean integrar un servicio de búsqueda que rastree todos los fondos documentales de ayuntamientos de esa provincia. Para ello solo deben editar el buscador **metafinder.php** e incluir en él los fondos que quieran.



8.3. VENTAJAS DE AGREGAR UN FONDO DOCUMENTAL AL METABUSCADOR **MARAFd**.

- Proceso libre, claro y gratuito sin cuotas ni licencias.
- Máxima visibilidad del patrimonio documental digital.

8.4. AÑADIR UN METABUSCADOR PROPIO EN SU SERVIDOR WEB.

- La arquitectura completamente abierta y basada en tesauro permite crear metafondos temáticos, además, lo primero que debe saber es que, gracias a la factorización/modularización del sistema, usted no necesitará escribir ni una sola

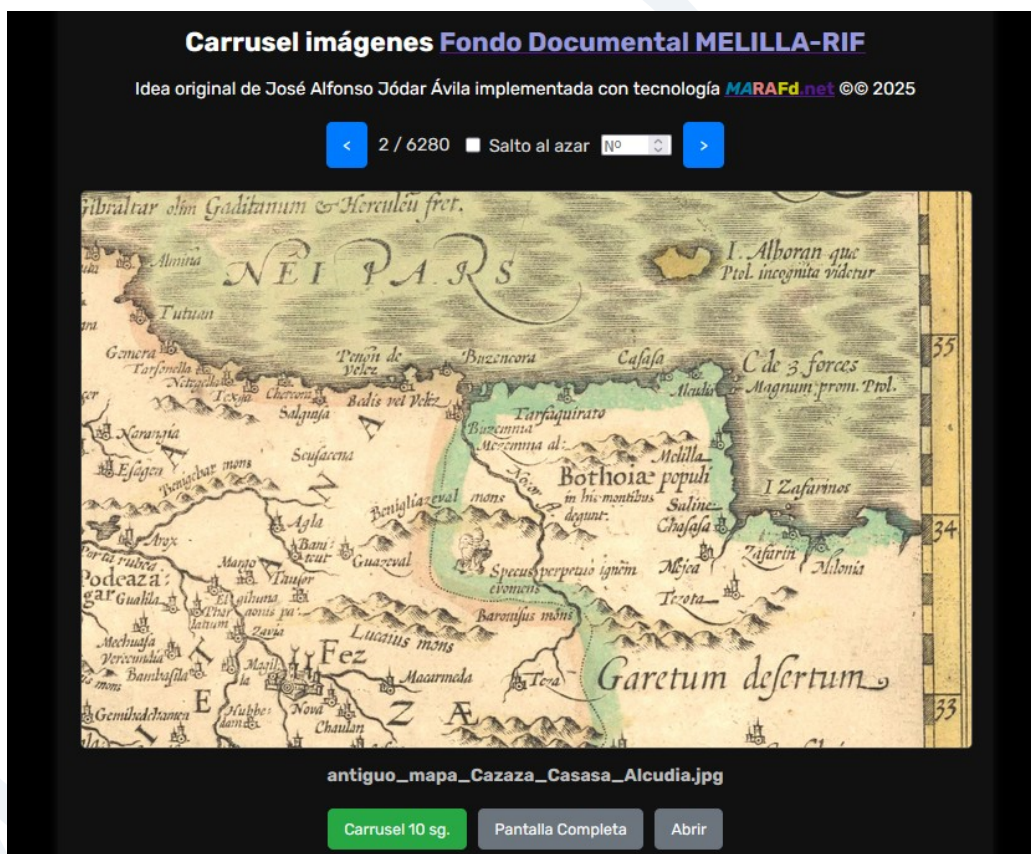
línea de código para agregar fondos documentales al metabuscador. De forma libre y gratuita usted puede editar el fichero 'fondos_agregados.csv' hospedado en la carpeta '/metafinder/' de su servidor y añadir los índices-tesauros de los fondos que desee agregar.

- También puede consultar la última edición del presente MANUAL TÉCNICO en la web oficial www.MARAFd.net para saber como adaptar rápidamente el fichero 'metafinder.php' a sus necesidades o consultar con el autor en la web oficial de esta tecnología abierta.
- Mediante esta tecnología de código abierto, cualquier administrador puede crear su propia colección de fondos uniendo fondos de entidades como por ejemplo, ayuntamientos de una provincia, comunidad autónoma, o un país, colegios, universidades, museos, etc.

9. Carrusel de imágenes. Un recorrido visual por el fondo.

Implementada a instancia de un usuario, se trata de una sencilla y potente funcionalidad a través de la cual podemos recorrer visualmente todo el contenido del fondo, y también, utilizarlo como salva-pantallas en nuestro equipo. Esta función es encomendada a los ficheros **carrusel.html** y **carrusel.js** que deberán editarse para adaptarlos a nuestras necesidades antes de subirlos al servidor. Además, el fichero **carrusel_exc.dat** nos permite excluir ciertas carpetas de archivos cuyas imágenes no deseamos mostrar en el carrusel.

Si bien puede resultar una herramienta banal, es crucial comprender como la arquitectura abierta de **MARAFd** y la sencillez extrema de su diseño nos ofrece un enorme potencial en el ámbito de la gestión inteligente de fondos documentales digitales. Con unas pocas líneas de código *HTML* y un sencillo guión *JavaScript* logramos implementar y ofrecer esta funcionalidad, poco frecuente en otros repositorios, que creemos aporta un considerable y atractivo valor añadido para el usuario.



10. Buenas prácticas.

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

Consulte con el autor en el site oficial:

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

10.2. 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://calentamientoglobacelerado.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://calentamientoglobacelerado.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 BatchScripting 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 , CSS (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	GCC (GNU Compiler Collection) es un conjunto de compiladores open-source (licencia GPL) para lenguajes como C, C++, FreeBasic 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 , HTML5 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
PHP	<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>
Python	<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 editables/no editables del ECOSISTEMA **MARAFd aplicado al caso de referencia del *Fondo Documental Digital Melilla-Riff***

Componente	Ubicación	Descripción	Ficheros resultantes
compactTIA.exe (compactTIA.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 compactTIA.asc El ejecutable .exe se obtiene compilando el archivo fuente .bas mediante el compilador <i>FreeBasic-GCC</i> incluido en ecosistema-MARAFd-completo.zip	indice-tesauro_compacto.txt indice-tesauro_compacto2.txt cTIA_tesauro_report.txt (estos ficheros son generados a partir del tesauro index_fondo_Melilla_Rif.asc)
carrusel.html carrusel.js	Servidor (carpeta raíz contenedora del fondo documental)	Estos ficheros proporcionan la funcionalidad del carrusel de imágenes y deberán editarse para su adaptación a nuestras necesidades.	
indizar.cmd	Local (carpeta raíz contenedora del fondo documental en local)	ScriptBatch de Microsoft Windows para generar el tesauro y llamada a programas externos TIA.exe y compactTIA.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 (este guion llama a compactTIA.exe que actualiza los 3 tesauros citados en el apartado anterior)

Componente	Ubicación	Descripción	Ficheros resultantes
localt.py	Local (carpeta raíz contenedora del fondo documental en local)	Toolkit escrita en <i>Python</i> para verificar integridad del los enlaces del índice-tesauro original en modo local. Lee el fichero index_fondo_Melilla_Rif.asc 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.
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 index_fondo_Melilla_Rif.asc 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

Componente	Ubicación	Descripción	Ficheros resultantes
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 índice al servidor.
../css ../js	Servidor (carpeta del fondo documental o la elegida para albergar el frontend)	Carpetas con librerías <i>css</i> y rutinas <i>js</i> utilizadas por el estándar bootstrap necesarias para <i>frontend-backend</i> 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.	No genera ficheros
index.php	Servidor (carpeta del fondo documental)	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. Si ubicamos el <i>frontend-backend</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	lista_archivos.html
finder3.php	Servidor (carpeta raíz del fondo o carpeta elegida para contener el frontend)	Buscador rápido sobre índice-tesauro que genera reportes	No genera ficheros. Solo rastrea el tesauro y reporta resultado con hiperenlaces.

MARAFd