

Manual del Operador

Ark of Noahledge · archivo de conocimiento sin conexión · Revisión 1.5 · 2026-09-27

SI USTED TIENE UNA IMPRESIÓN DE LA REVISIÓN 1.2, DOS COSAS ESTÁN MAL, Y UNA DE ELLAS LE CUESTA LA MITAD DE LA BÚSQUEDA.

El capítulo 8 no tiene ningún paso que instale los paquetes. Copiar el archivo no los instala, y el capítulo 8 de la revisión 1.2 pasa directamente de *instale Python* a *reconstruya los dos archivos*. Si lo sigue, termina con un archivo que funciona y **sin búsqueda híbrida**, y todas las comprobaciones de ese capítulo pasan igual. Aquí es el paso 3. Nada más en la 1.2 se lo va a decir.

El archivo de biblioteca se movió. El comando de respaldo del capítulo 3 y el capítulo 8 lo nombraban como `<ARCHIVE>/07-corpora-supplemental/library.xml`. **Ahora es**

`<ARCHIVE>/library.xml`, en la raíz del archivo, y cubre todos los estantes en vez de uno solo. Corregido en el texto el 2026-09-12 pero nunca impreso, así que una impresión de la 1.2 lo manda a un archivo que no existe. El síntoma **A** del capítulo 6 trae el segundo síntoma de la misma falla: el servidor arranca con normalidad y todos los enlaces a documentos fallan.

Qué cambió desde la Revisión 1.4 (2026-09-26). El asistente cambió su manera de responder, y los capítulos 1, 3, 4 y 9 cambiaron con él. **Ningún paso y ninguna etiqueta cambiaron.**

- **Capítulo 1: ahora responde con su propio conocimiento y contrasta la respuesta con el archivo.** Hasta la 1.4 respondía *usando únicamente los pasajes*. La respuesta sigue mostrando primero lo que el archivo respalda, con un `[1]` en cada afirmación, y después, en un bloque marcado, la respuesta completa del propio modelo. - **Capítulo 3: una página con todo lo que sirve el nodo**, en `http://localhost:8090/home/`, y un menú *Services* en la página de búsqueda. - **Capítulo 4: las respuestas son más largas.** Alrededor de un minuto la respuesta, unos minutos más la comprobación del segundo modelo. Qué son las dos partes de una respuesta, y la línea de la segunda parte que le indica detenerse y leer los pasajes. - **Capítulo 9, punto 1:** encuentra lo que dicen las fuentes *y agrega lo que sus modelos aprendieron*, y marca qué es cada cosa. - **El capítulo 5 no cambió.** Lea el pasaje fuente antes de actuar sobre cualquier número. Con respuestas más largas importa más, no menos.

Qué cambió desde la Revisión 1.3 (2026-09-17). Tres cifras, todas porque el índice creció de unos 7 millones de pasajes a 39 millones. Ningún paso cambió.

- **El ejemplo de `status` del capítulo 3** ahora dice `dense: pq over 39,073,563 vectors`. El número es cuántos pasajes cubre la búsqueda por significado. Lo que le dice que la híbrida funciona son las palabras `dense: pq over`, no el número. - **El capítulo 4 ya no dice que toda búsqueda híbrida después de la primera es rápida.** Después de la primera, cada una tarda unos diez segundos. Es normal, no una falla. - `index-provenance.py --rebuild` **tarda unos diez minutos, no uno** (síntoma **D** del capítulo 6, paso 4 del capítulo 8). Con una 1.3 impresa en la mano pensaría que se colgó. No se colgó.

Qué cambió desde la Revisión 1.2 (2026-09-08).

- **El capítulo 8 ganó el paso 3**, que instala los paquetes, y el paso 6 ahora pide `--mode hybrid` para que la comprobación final pueda fallar. Sin eso, el capítulo terminaba declarando completo un nodo recuperado a medias. - **El capítulo 3 ya no dice que la híbrida se decide al arrancar.** No es así. La mitad que busca por significado se carga al usarla y puede fallar después de que todo se vea bien, que es lo que pasó el 2026-09-16 durante tres horas. Hay una tabla de lo que `ark.py status` informa realmente. - **La tabla de etiquetas del capítulo 4 estaba mal y ahora está completa.** Traía dos de las cinco etiquetas, describía `uncovered` como algo que no es, y no tenía entrada para `model only - no citation`, la única de la lista que sí es una falla. La primera columna ahora cita la pantalla palabra por palabra. - **El capítulo 4 dice que `hybrid` ya viene elegido** cuando se abre la página.

Qué cambió desde la Revisión 1.1 (2026-09-07). El capítulo 3 está reescrito alrededor de un solo comando que arranca todo, y **arrancar los dos modelos ahora es tarea suya y no de quien construyó esto** — la Revisión 1.1 decía que no lo era, y eso hacía imposible aprobar dos pasos del simulacro anual usando solo estas páginas. El síntoma **E** del capítulo 6 cambió con ello, y el capítulo 8 ahora dice qué necesita una máquina reconstruida para cada modelo.

Si usted tiene una impresión de la Revisión 1.0 (2026-09-06), además le falta lo que añadió la 1.1: *Encontrar un lugar por su nombre* en el capítulo 3a, y que la razón por la que un lugar puede no aparecer es cómo etiqueta el mapa y no el zoom; una segunda pasada trimestral en el capítulo 7, porque la de la 1.0 lee solo el disco frío y el disco de trabajo necesita la suya; y la línea de abajo para el `python` que puede leer mapas.

Implementa la especificación v1.7 §8.10 · El rol se define en `OPERATOR-ROLE.md` Edición en inglés: `OPERATOR-MANUAL-EN.md` . Imprima una copia y guárdela con los discos.

Nota sobre el idioma. La pantalla del sistema está en inglés. En este manual los textos que usted verá aparecen **en inglés y sin traducir**, porque es lo que va a leer en la pantalla; la explicación está en español. Traducir una etiqueta que usted nunca verá sería peor que no traducir nada.

Antes de empezar: complete esto

Este manual no puede saber dónde está su archivo. Escríbalo aquí, una vez, a mano.

	SU RESPUESTA
Dónde está el archivo en esta máquina	
Nombre escrito en el disco frío	
Quién construyó esto	
A quién llamar	
Fecha del último simulacro	
El <code>python</code> que puede leer mapas (cap. 3a)	
La carpeta donde se descomprimió <code>llama-server</code> (cap. 3)	

En todo el manual, `<ARCHIVE>` significa la primera línea. Es la carpeta que contiene `10-index`, `07-corpora-supplemental` y `13-ark-node` .

Los comandos. Escríbalos exactamente. En Ubuntu use la Terminal. En Windows use **Git Bash**, no el Símbolo del sistema, porque los comandos están escritos para Git Bash.

Tres comandos de este manual no son `python`, y son justamente los que dependen de esa diferencia: el capítulo 3 inicia `kiwix-serve`, el capítulo 3a inicia `pmtiles serve`, y el capítulo 7 empieza con `bash`. Esta página antes afirmaba que todo lo que usted necesita es `python`, que es lo que le dice a un lector de Windows que puede ignorar la línea de Git Bash de arriba. El capítulo 7 es exactamente donde eso habría fallado, y el capítulo 7 no se puede omitir.

0. Si está leyendo esto

Algo pasó, o alguien está probando si estas páginas sirven. En cualquiera de los dos casos, esta es la página correcta.

Usted no puede dañar el archivo por leerlo. Nada en los capítulos 1 a 5 modifica un solo archivo. Si tiene dudas, haga esos y deténgase.

Lo que puede esperar. Todo lo de los capítulos 7 y 8. Son mantenimiento y reparación, no emergencias. Si la máquina enciende y la búsqueda funciona, nada es urgente.

Lo que no puede esperar. Si un disco hace un ruido que antes no hacía, desconéctelo. Un disco que está fallando empeora mientras funciona.

1. Qué es esto, en una página

Un computador sin conexión a internet que guarda una copia de una gran cantidad de material de referencia: medicina, reparación, agricultura, ingeniería, mapas, libros de texto. Alrededor de dos terabytes.

Encima de eso hay una búsqueda que encuentra pasajes y, cuando sus modelos están funcionando, un asistente que responde preguntas con su propio conocimiento, **contrasta la respuesta con esos pasajes**, y le muestra qué partes respalda el archivo y de dónde salió cada una.

De ahí se desprenden tres cosas, y son la razón completa de este manual.

1. **Solo puede comprobar lo que el archivo contiene.** Cuando el archivo no cubre una pregunta, la parte de la respuesta que cita pasajes dice `NOT IN ARCHIVE`, y todo lo que el modelo agregue de lo que aprendió queda marcado como no proveniente del archivo. Ese mensaje es el sistema funcionando, no fallando.
 2. **La respuesta es un resumen. El pasaje es la fuente.** Cualquier número sobre el que usted vaya a actuar se lee en el pasaje, no en la respuesta.
 3. **Funciona por etapas.** La búsqueda funciona solo con el archivo. El asistente necesita más, y el capítulo 3 le muestra cómo arrancarlo. Si el asistente no está disponible, la búsqueda sigue funcionando y la pantalla se lo dirá.
-

2. Los discos

Hay al menos dos copias de todo: la que está en la máquina, y uno o más **discos fríos** que permanecen desconectados.

Un disco frío que se deja conectado no es un respaldo. Comparte lo que le pase a la máquina, que es justamente lo que existe para sobrevivir. Conéctelo para el trabajo y desconéctelo al terminar.

Para conectarlo: enchúfelo a un puerto USB. Espere a que la máquina lo muestre. Anote la letra o el nombre con que aparece; lo va a necesitar.

Para desconectarlo: siempre expúlselo correctamente primero. En Windows, el ícono de la bandeja para quitar hardware de forma segura. En Ubuntu, el símbolo de expulsar al lado del disco. Después desenchufe. **No se salte este paso.** Halar un disco mientras escribe es la forma más común en que un archivo se daña por culpa de la persona que intentaba protegerlo.

3. Encenderlo y buscar

Un solo comando arranca todo.

```
python <ARCHIVE>/bin/ark.py up
```

Arranca cada parte que no esté ya funcionando e imprime una tabla que dice qué subió, qué no, y por qué. **Ejecutarlo dos veces no hace daño:** le pregunta a cada parte si ya está respondiendo antes de arrancar nada.

Vale la pena conocer dos más:

```
python <ARCHIVE>/bin/ark.py status  
python <ARCHIVE>/bin/ark.py down
```

status dice qué está funcionando en este momento y no arranca nada. **down** detiene lo que el lanzador arrancó.

Qué arranca, y qué pierde usted sin cada parte:

PARTE	SIN ELLA
archive	los documentos originales detrás de cada cita. Los enlaces dejan de abrir
node	la página de búsqueda misma. No funciona nada
tiles	los mapas, capítulo 3a. Lo demás no se ve afectado
primary	el panel de respuestas. La búsqueda y los pasajes siguen funcionando
crosscheck	la segunda opinión sobre respuestas que llevan una cifra

Las tres primeras son la búsqueda. Las dos últimas son el asistente, y son las dos que una máquina dada puede no ser capaz de correr.

Ahora abra un navegador en esta máquina y vaya a:

```
http://localhost:8090/
```

Debe aparecer una caja de búsqueda. **Eso es el sistema funcionando.** Si el navegador no muestra nada, capítulo 6, síntoma **B**.

Todo lo demás que sirve el nodo aparece en una sola página, con el estado de cada parte:

```
http://localhost:8090/home/
```

La página de búsqueda, la biblioteca con todos los libros del disco, el mapa, las páginas de chat de los dos modelos y la salud del nodo. El menú *Services*, arriba a la derecha en la página de búsqueda, abre la misma lista. **Las dos páginas de chat hablan directamente con un modelo: no se consulta nada y no se cita nada**, y la página lo dice. Solo abren en esta misma máquina.

Si una parte dice `down` y nombra algo que falta

El lanzador le dice qué hacer en esa misma línea, casi siempre un archivo que hay que descomprimir de `09-software`. Haga eso y vuelva a ejecutar `up`.

Si se niega a arrancar un modelo diciendo que la máquina no lo puede sostener, eso no es una falla. Es el lanzador negándose a fallar despacio, en vez de dejar que la carga corra un minuto y luego se derrumbe. Todo lo demás arranca igual.

Si el lanzador no arranca del todo

Es una comodidad y no es el sistema. Todo comando que ejecutaría se puede imprimir en su lugar, y luego escribirlo a mano:

```
python <ARCHIVE>/bin/ark.py commands
```

Eso no arranca nada. Y si ni eso funciona, estos dos son el piso. Cada uno va en su propia ventana de terminal, y las dos ventanas quedan abiertas.

Ventana 1, el servidor del archivo, que sirve los documentos detrás de cada cita:

```
<ARCHIVE>/09-software/kiwix-tools/bin/kiwix-serve --port 8080 --library <ARCHIVE>/li
```



Ventana 2, la página de búsqueda misma:

```
python <ARCHIVE>/13-ark-node/ark-api/serve.py
```

El segundo termina con una línea como `serving on http://0.0.0.0:8090/`. Cerrar cualquiera de las dos ventanas detiene esa parte. Si `kiwix-serve` dice que el archivo de biblioteca no existe, vaya al capítulo 6, síntoma A. Ese par le da búsqueda, pasajes y todas las citas, sin mapas y sin asistente.

En Windows cada parte abre su propia ventana. Cerrar una detiene esa parte, y cinco ventanas cerradas a la vez se ve exactamente igual que cinco cosas fallando a la vez. Si `status` dice que todo está abajo y usted no ejecutó `down`, las ventanas se cerraron. Ejecute `up` otra vez.

Que la búsqueda híbrida esté disponible NO se decide al arrancar, y el lanzador no se lo puede decir. La mitad que busca por significado se carga la primera vez que algo se la pide, así que todavía puede fallar después de que todo se vea bien. El 2026-09-16 este nodo funcionó tres horas con el lanzador informando una configuración sana y la búsqueda híbrida silenciosamente no disponible.

`up` le dice cuál Python eligió. Eso es todo lo que puede saber en ese momento. `status` es lo que informa la respuesta, en la línea `node`:

LO QUE DICE STATUS EN LA LÍNEA NODE	QUÉ SIGNIFICA
dense: pq over 39,073,563 vectors	la híbrida está funcionando, sea cual sea el número
dense not loaded yet - it is lazy, it loads on the first query	normal, no es una falla. Todavía nadie se la ha pedido. Haga una búsqueda y mire otra vez
DENSE FAILED TO LOAD - keyword only	la híbrida no está disponible. La razón se imprime debajo de la tabla
no keeper venv at C:\ark-env	los paquetes nunca se instalaron. El paso 3 del capítulo 8 los instala

Tanto una híbrida que funciona como solo-palabras son configuraciones que funcionan; el capítulo 4 explica la diferencia.

3a. Los mapas (opcional)

El archivo guarda **843 GB de mapas**: un mapa de calles de todo el planeta y un modelo de elevación de todo el planeta. Son cerca del 38% de todo lo que hay en el disco y necesitan una parte más funcionando.

`ark.py up` la arranca como `tiles`, junto con todo lo demás, así que si usted siguió el capítulo 3 ya está funcionando y puede saltar al navegador más abajo. `status` le dice si está arriba.

A mano, si no está usando el lanzador, es una tercera ventana:

```
<ARCHIVE>/09-software/pmtiles-cli/pmtiles serve <ARCHIVE>/08-maps --port 8081 --cors="**"
```

En Windows el programa es `pmtiles.exe` en esa misma carpeta. En Ubuntu, descomprima primero `go-pmtiles_1.31.2_Linux_x86_64.tar.gz` de esa carpeta.

La salida correcta es una sola línea que nombra el puerto:

```
Serving . on port 8081 and interface 0.0.0.0 with Access-Control-Allow-Origin: *
```

`--cors=""` **no es un adorno**. Sin eso el mapa no dibuja nada y el navegador reporta un error de seguridad en vez de un archivo faltante. Si el mapa sale en blanco, revise esa opción primero.

Después abra:

```
http://localhost:8090/map/
```

Un mapa del mundo, centrado en Caracas. Dos casillas: **hillshade** enciende el sombreado del relieve, **3D terrain** inclina la vista. Arrastre para moverse, use la rueda para acercar.

Qué esperar. Dibuja calles, agua, fronteras, nombres de lugares y cobertura del suelo hasta el nivel de calle, y elevación en todas partes. No tiene fotografía satelital, ni tráfico, ni caja de búsqueda, ni indicaciones de ruta. No le puede decir cómo llegar a un sitio; le puede mostrar qué hay ahí.

Desde un teléfono o una tableta en el mismo WiFi, use la dirección del nodo en vez de `localhost` — la misma dirección del capítulo 3, con `/map/` al final. No hay que editar nada para que esto funcione.

Encontrar un lugar por su nombre

El mapa no tiene caja de búsqueda. *Dónde queda Maracaibo* es otro comando, y no necesita ningún servidor, ni siquiera el de mosaicos de la ventana 3:

```
python <ARCHIVE>/bin/index-gazetteer.py --find maracaibo

1 place named 'maracaibo':
  Maracaibo          locality pop_rank=12    10.6498,  -71.6418  (z3)
```

Esos dos números son latitud y longitud, y son justamente lo que la consulta de clima siempre necesitó:

```
python <ARCHIVE>/bin/koppen-lookup.py 10.6498 -71.6418 Maracaibo
```

Si eso dice `rasterio is not installed`, **es el comando correcto y el** `python` **equivocado**. Leer los rásters del mapa necesita un entorno aparte, y su ruta es la última línea de la tabla al comienzo de este manual. En la máquina donde se construyó esto es `C:\ark-env\Scripts\python.exe`, así que el comando queda:

```
C:/ark-env/Scripts/python.exe <ARCHIVE>/bin/koppen-lookup.py 10.6498 -71.6418 Maracaibo
```

El nodo le imprime la forma que funciona: `/api/spatial` comprueba cuál intérprete puede importar rasterio antes de escribir el comando.

Muchos lugares comparten nombre, así que la respuesta es una lista. `--find franklin` devuelve trece, ordenados por rango de población, el mayor primero. El archivo no puede saber a cuál se refiere usted. Si toma la primera línea sin leer las coordenadas, puede terminar a mil cien kilómetros de donde creía estar.

Pregúntele en cualquiera de sus idiomas. 46.861 lugares bajo 271.848 nombres en 42 idiomas: `--find tokyo` y `--find 東京都` llegan al mismo punto, igual que `--find cairo` y `--find القاهرة`.

Si un lugar no aparece, hay dos razones y la segunda es la común. Se construyó con los niveles de zoom o a 8, que son pueblos y no aldeas. Y se construyó con la capa `places` del mapa, que etiqueta **solo pueblos, regiones y países:** un lago, una montaña, un parque nacional o unas ruinas se dibujan en el mapa y no tienen nombre buscable en ningún zoom. `Lake Titicaca` nunca aparecerá por nombre; el lago sí está en el mapa. Ubíquelo a la vista y apunte con el cursor: el panel ahora muestra la latitud y la longitud bajo el puntero, y el comando de clima exacto para copiar. En un teléfono, toque la pantalla; toque otra vez para soltar.

Si el mapa sale en blanco o sin nombres

LO QUE VE	QUÉ SIGNIFICA
pantalla gris, mensaje sobre teselas	la ventana 3 no está corriendo, o se omitió <code>--cors="*"</code>
el mapa dibuja pero sin nombres de lugares	la página y sus tipografías se están sirviendo desde lugares distintos. Use <code>http://localhost:8090/map/</code> , no el archivo en el disco
el mapa dibuja, las montañas se ven raras	el relieve se dibuja con datos de elevación, no con fotografías. El sombreado está exagerado a propósito
nada, página en blanco	el nodo mismo no está corriendo. Vuelva al capítulo 3, ventana 2

El mapa no puede dañar nada. Solo lee. Si se comporta mal, cierre la ventana y el resto del sistema sigue igual.

4. Hacerle una pregunta

Escriba una pregunta y presione `Search`. Los resultados aparecen a la izquierda.

El selector de modo, al lado de la caja de búsqueda:

OPCIÓN	QUÉ HACE
keyword	busca las palabras que usted escribió. Siempre disponible
hybrid	además busca por significado, así encuentra pasajes que usan otras palabras. Necesita más partes del sistema. La primera búsqueda hybrid tarda hasta un minuto , y cada una de las siguientes tarda unos diez segundos

hybrid ya viene elegido cuando se abre la página. Usted no tiene que seleccionarlo. Pasó a ser el modo por omisión porque una búsqueda solo por palabras de *qué hago con alguien en shock* devolvía, en el tercer lugar, un pasaje de física sobre ondas de choque. La híbrida no comete ese error.

Si elige **hybrid** y la pantalla dice que se ejecutó como **keyword**, eso no es una falla. Le está diciendo la verdad sobre lo que hizo, y el capítulo 3 explica cómo averiguar por qué.

A la derecha está la respuesta, cuando los modelos están funcionando. Si no lo están, lo dice claramente y los pasajes de la izquierda siguen siendo reales. Para arrancarlos, vea el capítulo 3; si el lanzador se niega, el síntoma E del capítulo 6 explica qué significa eso.

Una respuesta tiene dos partes, siempre en este orden.

1. **Lo que el archivo respalda**, con un [1] en cada afirmación. Cada número de esta parte está copiado de un pasaje.
2. **La respuesta propia del modelo**, en un bloque que empieza con **UNSOURCED - NOT FROM THE ARCHIVE**: una respuesta completa a partir de lo que el modelo aprendió, que nada en este disco ha comprobado. **Su primera línea dice si coincide con el archivo, si le agrega algo o si lo contradice. Si dice que contradice el archivo, deténgase y lea los pasajes** (capítulo 5).

Una respuesta tarda alrededor de un minuto. La comprobación del segundo modelo aparece debajo unos minutos después.

Qué significan las etiquetas. Son cinco, y la primera columna trae las palabras exactas de la pantalla — en inglés, como aparecen — para que la página y esta tabla se puedan comparar sin traducir entre las dos.

LO QUE DICE LA PANTALLA	QUÉ SIGNIFICA	QUÉ HACER
grounded in the archive	cada afirmación proviene de un pasaje mostrado	lea el pasaje antes de actuar sobre cualquier número. Capítulo 5
grounded in the archive + unsourced note	la respuesta está respaldada y el modelo además respondió con su propio conocimiento , en el bloque marcado al final. Desde la revisión 1.5 ese bloque es una respuesta completa, no una nota	la parte citada responde por sí misma. El bloque no. Trátelo como algo que le contaron, no como algo que leyó
Not in the passages found - model recall only	nada de esta respuesta salió del archivo. El modelo está diciendo lo que recuerda, y lo está admitiendo	compruébelo en otra parte antes de actuar. No es falso por definición, pero nada en este disco lo respalda
archive does not cover this	el archivo no contiene la respuesta, y el modelo lo dijo en vez de inventar una	es un buen resultado, no una falla. Pruebe con otras palabras; si se repite, el vacío es real. Capítulo 9
model only - no citation	una respuesta sin cita y sin admisión	la única de esta lista que sí es una falla. No actúe sobre ella, y avísele a quien mantiene el nodo

La revisión 1.2 tenía esta tabla mal, y vale la pena decir cómo, porque todavía hay copias impresas. Ponía `uncovered` como *parte de la respuesta no está respaldada por los pasajes*, que no es lo que significa esa etiqueta, y le daba el significado correcto a una fila aparte llamada `NOT IN ARCHIVE`. `NOT IN ARCHIVE` es la frase que escribe el modelo; `archive does not cover this` es lo que después muestra la pantalla. Una etiqueta, no dos. Tampoco traía ninguna de las tres etiquetas siguientes, así que quien tenga la 1.2 impresa no tiene entrada para la única de la lista que es una falla.

Dos cosas más en pantalla, que no son etiquetas de respaldo:

adjacent	un pasaje que el sistema trajo porque otro pasaje mencionaba una tabla que no contenía. Debajo explica por qué
un [1] en la respuesta	haga clic. Abre el pasaje del que salió esa afirmación

5. La única regla

Lea el pasaje original antes de actuar sobre cualquier número.

Dosis, presiones, voltajes, temperaturas, tolerancias, tiempos. El sistema los marca cuando los detecta y le indica que verifique. Esa advertencia no es un formalismo ni está ahí porque el sistema sea poco confiable en general.

Está ahí porque **la falla que produce este sistema no es un mensaje de error. Es una respuesta segura, bien escrita y correctamente citada, con algo equivocado adentro.** Eso ya pasó durante las pruebas, más de una vez, y cada vez la respuesta se veía exactamente igual a las correctas.

Haga clic en el [1]. Lea el pasaje. Confirme que el número está escrito ahí, con las mismas unidades y en el mismo contexto. Después actúe.

Y anótelos o déjelo en pantalla. No actúe sobre un número que escuchó y no leyó.

6. Cuando algo sale mal

Busque su síntoma. Haga únicamente lo que dice.

A. kiwix-serve dice que el archivo de la biblioteca no existe, o arranca y todos los enlaces a documentos dan una página de error. La biblioteca es específica de cada máquina y se reconstruye, no se restaura. **Los dos síntomas tienen la misma causa y la misma solución,** y el segundo importa más, porque el servidor va a decir que cargó todo bien. Ejecute:

```
python <ARCHIVE>/bin/kiwix-library.py
```

Escribe `<ARCHIVE>/library.xml` e imprime dos líneas que empiezan por `all`. Después repita el capítulo 3.

B. El navegador no muestra nada en `localhost:8090`. Mire la ventana 2. Si terminó con un error, anote las últimas tres líneas y vaya a **G**. Si sigue funcionando, pruebe con `http://127.0.0.1:8090/`.

C. La búsqueda no devuelve nada, para nada. Falta el índice o no se puede leer. Vaya al capítulo 8.

D. Un `[1]` abre el pasaje equivocado, o no abre nada. Deje de usar las respuestas. Significa que el índice y su mapa no coinciden. Ejecute:

```
python <ARCHIVE>/bin/index-provenance.py --rebuild
```

Espere `0 artifact(s) still pending`. Alrededor de diez minutos; no se ha colgado. Después reinicie la ventana 2.

E. La pantalla dice que el modelo no está funcionando. La búsqueda y los pasajes no se ven afectados en ningún caso. Para arrancarlo:

```
python <ARCHIVE>/bin/ark.py status
```

Si `primary` dice `down`, ejecute `python <ARCHIVE>/bin/ark.py up`. Cargar tarda alrededor de un minuto; `status` otra vez se lo mostrará.

Si el lanzador se niega, lea la línea que imprime. Hay dos clases y significan cosas distintas.

1. **Nombra un archivo que falta**, casi siempre algo que descomprimir de `09-software`. Hágalo y ejecute `up` otra vez.
2. **Dice que la máquina no puede sostener el modelo**, con dos números: lo que el modelo necesita y lo que la máquina tiene libre. **Eso no es una falla y no hay nada que reparar.** Esta máquina corre búsqueda, pasajes y citas, y no el asistente. El capítulo 8 dice qué necesita una máquina para cada modelo.

El segundo modelo, `crosscheck`, corre en la memoria normal del sistema y no en una tarjeta gráfica, así que una máquina que no puede correr `primary` todavía puede correr ese.

F. Un disco suena, vibra raro, o se quedó callado cuando antes no. Desconéctelo. No ejecute nada contra él. Anote cuál era.

G. Cualquier cosa que no esté en esta lista. Eso es un defecto de este manual, no un error suyo. Anote: qué estaba haciendo, el comando exacto, y las últimas tres líneas de la pantalla. Esa nota es lo más valioso que usted puede producir; es lo que hace que se agregue una página aquí.

7. Cada tres meses: comprobar que los discos siguen buenos

Cada tres meses. Toma una hora o más. No es urgente, y no se puede omitir, porque la falla que detecta es silenciosa.

Los archivos se degradan en el disco sin que nada lo reporte. El archivo conserva su nombre, su tamaño y su fecha, y parte de su contenido ya no es lo que se escribió. La única forma de encontrarlo es leer cada byte y compararlo contra una huella tomada cuando el archivo estaba sano.

Conecte el disco frío (capítulo 2) y ejecute:

```
bash <ARCHIVE>/bin/backup-cold.sh --verify
```

Va a parecer que se congeló. No es así.

Este comando lee cada byte del disco, más de dos terabytes, y no imprime absolutamente nada mientras lo hace. Una hora sin salida en pantalla es normal y esperado. No lo detenga. No desconecte nada. Vaya a hacer otra cosa.

La línea que hay que esperar es:

```
cold copy verified clean
```

Cualquier otra cosa, incluida cualquier línea que contenga **FAILED**, significa que debe anotar exactamente lo que decía y detenerse. Después expulse el disco correctamente (capítulo 2) y desconéctelo.

En Windows este comando necesita Git Bash. Si no está instalado, esa es una tarea del constructor, no suya.

Esto revisa la copia FRÍA. El disco de trabajo necesita su propia pasada

El comando de arriba lee el disco frío. Esa es la copia desde la que usted restauraría, así que es la que más importa, y no es todo el trabajo.

La degradación de bits no cambia el tamaño ni la fecha de un archivo, así que el espejo que actualiza la copia fría se salta el archivo dañado y no copia nada. Los bytes viejos y buenos de la copia fría entonces verifican limpios, y el disco de trabajo es el que carga el daño. El párrafo de arriba dice que la falla es silenciosa; así es como sigue siendo silenciosa incluso después de una verificación limpia.

La pasada que lee el disco de trabajo:

```
bash <ARCHIVE>/bin/verify.sh
```

Misma forma, misma paciencia, misma regla sobre **FAILED**. Corra las dos, en cualquier orden, una vez por trimestre. Si solo corre una, está revisando la copia que no usa.

8. Reconstruir en otra máquina

El capítulo largo. Necesita el disco frío y un computador que funcione.

Lea esto antes de empezar

Lo que recupera depende de la máquina, y el archivo no puede darle todo en ninguna de las dos.

Si la máquina ya tiene Windows: todo lo de este manual funciona, incluida la búsqueda `hybrid`. Python y todos sus paquetes están en el disco — y **el paso 3 de más abajo es el que los instala**. Copiar el archivo no los instala.

Si la máquina está vacía: el disco puede instalarle Ubuntu (`09-software/os-ubuntu`), y entonces usted tendrá **búsqueda, pasajes y citas, pero no** `hybrid` — los paquetes que necesita esa mitad no están en el disco para Linux. Es una carencia conocida, registrada en la especificación como §12 punto 7.

El asistente es una tercera pregunta, aparte de las dos anteriores. Los dos modelos no quedan instalados por copiar el archivo: `llama-server` se descomprime de `09-software/llamacpp-bin` en una carpeta que usted elija, y el lanzador nombra el archivo exacto y la carpeta cuando falta. Entonces:

- `primary` quiere una tarjeta gráfica con unos **15 GB libres**. En esta construcción ocupa 14.801 MiB de una tarjeta de 16.050 MiB, así que una tarjeta más pequeña será rechazada en vez de cargarse a medias. - `crosscheck` corre **enteramente en la memoria del sistema** y quiere unos **21 GB de RAM** y ninguna tarjeta gráfica. Una máquina con bastante memoria y sin GPU utilizable puede correr ese y no el otro.

Una búsqueda por palabras con una cita para cada respuesta, sobre dos terabytes de material de referencia, es un archivo que funciona. Es el piso al que este sistema fue diseñado para caer, no una recuperación fallida.

1. Copie el archivo del disco frío a la máquina nueva. Use el copiador de archivos del propio sistema operativo. Copie la carpeta `civbackup` completa. Es grande; espere horas. Verifique comparando el tamaño de la carpeta en ambos lados antes de continuar.

2. Instale Python, si la máquina no lo tiene. Windows: ejecute `09-software/python-3.12.9-amd64.exe` desde la copia y marque **Add Python to PATH**. Ubuntu: ya viene instalado.

Compruebe que quedó:

```
python --version
```

3. Instale los paquetes. SOLO EN WINDOWS. Copiar el archivo no los instala. Sin ellos el nodo funciona **solo por palabras**: la búsqueda, los pasajes y las citas funcionan, y `hybrid` no. Nada se lo

advierde con palabras que usted reconocería como una advertencia, y por eso este paso está numerado y no mencionado de paso.

```
python -m venv C:\ark-env  
C:\ark-env\Scripts\python.exe -m pip install --no-index --find-links=<ARCHIVE>\09-software
```

No usa red ni la necesita. Tarda varios minutos e instala 221 paquetes desde el disco. El segundo comando es largo; escríbalo en una sola línea.

Compruebe que quedó:

```
C:\ark-env\Scripts\python.exe -c "import numpy, faiss, sentence_transformers; print('ok')"
```

Tiene que imprimir `ok`. Si imprime cualquier otra cosa, los dos comandos de arriba no terminaron: ejecútelos otra vez y lea lo que dicen.

Hay un script, y no es para esto. `bin/rebuild-venv.sh` construye el mismo entorno con todas las comprobaciones incluidas, pero **se niega a construir en** `C:\ark-env` — en cualquier forma de escribirlo — porque su trabajo es demostrar que una reconstrucción funciona *al lado* de un entorno que ya está corriendo, sin destruirlo. Es una herramienta de verificación. **Los dos comandos de arriba son el procedimiento de recuperación**, y están escritos aquí en vez de referenciados para que un script que no arranca nunca sea la razón por la que usted quedó atascado.

En Ubuntu, omite este paso. Los paquetes de este disco están compilados para Windows. Tendrá búsqueda, pasajes y citas, y no `hybrid` — la carencia §12 punto 7 nombrada al principio de este capítulo.

4. Reconstruya los dos archivos que son específicos de esta máquina. Ninguno se restaura desde el respaldo, porque los dos describen dónde están las cosas, y las cosas se movieron.

```
python <ARCHIVE>/bin/index-provenance.py --rebuild  
python <ARCHIVE>/bin/kiwix-library.py
```

El primero termina con `0 artifact(s) still pending`, después de unos diez minutos. El segundo escribe `<ARCHIVE>/library.xml` e imprime dos líneas que empiezan por `all`.

No omita el segundo porque el archivo ya esté ahí. Un archivo restaurado trae su propio `library.xml`, y las rutas que contiene nombran el disco donde se construyó. La verificación de integridad no cubre ese archivo, así que nada le avisará: `kiwix-serve` lo cargará, dirá que todo salió bien, mostrará la lista de libros, y todos los enlaces a documentos fallarán.

5. Enciéndalo exactamente como en el capítulo 3:

```
python <ARCHIVE>/bin/ark.py up
```

Lea la tabla que imprime. Cualquier parte que diga `down` nombra en esa misma línea lo que necesita. **Y lea la línea `node` aunque diga `UP`** — la tabla del capítulo 3 dice qué le puede informar ahí. Si dice `no keeper venv`, el paso 3 no quedó.

6. Confirme que la cadena está completa. En Windows:

```
C:\ark-env\Scripts\python.exe <ARCHIVE>/13-ark-node/ark-api/serve.py --selftest --mode hybrid
```

En Ubuntu, donde no hay mitad híbrida que probar:

```
python <ARCHIVE>/13-ark-node/ark-api/serve.py --selftest --mode keyword
```

Hace preguntas reales y **falla si algún resultado no se puede convertir en una cita**, que es la única propiedad que vale la pena comprobar.

`--mode` **no es opcional aquí, y omitirlo es como este capítulo terminaba con un resultado falso.** La autopruueba usa `keyword` por omisión, que no necesita ningún paquete — así que en Windows pasaría en una máquina donde el paso 3 nunca se hizo, y el capítulo le diría que la recuperación está completa sobre un nodo sin búsqueda híbrida. Pedir `hybrid` es lo que hace que la comprobación pueda fallar.

Después, a mano: busque algo y haga clic en una cita.

La recuperación está completa cuando las dos cosas son ciertas: la cita abre el documento original, y `ark.py status` informa una línea `dense:` que funciona en la línea `node` — o, en Ubuntu, dice solo-palabras, que es el piso que este capítulo describió al empezar y es un archivo que funciona.

9. Lo que este sistema no puede hacer

Dicho sin rodeos, porque una herramienta cuyos límites se desconocen termina usada donde no debe.

1. **No es un médico, ni un ingeniero, ni un abogado.** Encuentra lo que dicen fuentes publicadas y agrega lo que sus modelos aprendieron, y marca qué es cada cosa. Juzgar si alguna de las dos aplica a la situación que usted tiene enfrente es tarea suya.
2. **No sabe nada de lo que pasó después de que fue construido.** Ni noticias, ni actualizaciones, ni correcciones publicadas desde entonces.
3. **Su material médico en español es más escaso y más viejo que el de inglés.** Aproximadamente la cuarta parte del tamaño, y varias ediciones atrás en algunos títulos. **Cuando una fuente en español y una en inglés no coincidan en una cifra clínica, la de inglés tiene más probabilidad de estar vigente.**
4. **Puede equivocarse con toda seguridad.** Vea el capítulo 5. Ese es el modo de falla.
5. **No puede repararse solo.** Todo lo de aquí funciona o reporta que no funcionó.
6. **No todas las máquinas corren todas sus partes.** El asistente necesita hardware que la búsqueda no necesita, y el archivo no puede suministrar hardware. El capítulo 8 dice qué necesita cada modelo; el síntoma E del capítulo 6 dice qué hacer cuando una máquina no lo alcanza, que es nada.

10. A quién preguntar, y qué hacer si no hay a quién

Complete el recuadro del comienzo de este manual. Si esa persona no está disponible:

1. **Aquí no hay nada secreto.** El archivo es material abierto y estos procedimientos están publicados. Cualquier persona con soltura frente a un computador puede leer este manual.
 2. **El sistema no se deteriora si se deja quieto.** Se deteriora si se opera mal. Ante la duda, no haga nada y deje las dos copias desconectadas.
 3. **El capítulo 8 es el sistema completo.** Si una máquina muere, el disco más ese capítulo son una recuperación completa. Para eso existe el disco.
-

Fin del manual. Cada paso que no se haya podido completar únicamente con estas páginas es un defecto de este documento. Anótelo y se corregirá aquí.