SOPHIA · Caso documentado de diseño de procesos

De la limpieza del texto a la reconstrucción

Cómo conectar operaciones, preservar información y mantener la continuidad entre fases.

Módulos conectados y recorridos flexibles

SOPHIA nació de un flujo de traducción en el que necesitaba moverme constantemente entre aplicaciones, documentos, carpetas y recursos externos, mientras mantenía el seguimiento de las tareas y de los archivos que iba generando.

Las primeras automatizaciones resolvían necesidades concretas. A medida que el proyecto crecía, surgió otro reto: conseguir que las distintas operaciones funcionaran como partes de un mismo proceso, sin perder la flexibilidad necesaria para abordar trabajos diferentes.

No se trataba únicamente de reunir scripts en una aplicación. Había que identificar qué necesitaba cada fase, qué información convenía preservar, cómo preparar las salidas para las operaciones posteriores y dónde mantener la intervención humana.

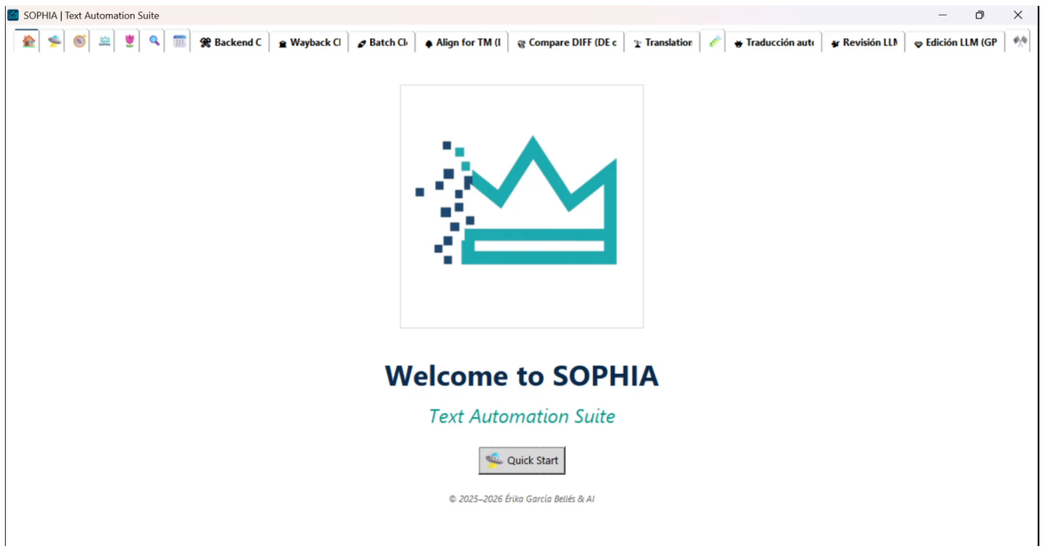
Un entorno común, distintos recorridos

Home proporciona un punto de partida común, mientras que SOPHIA Navigator organiza los módulos por fases y permite acceder directamente a ellos. La arquitectura modular facilita tanto el uso independiente de determinadas herramientas como su combinación en recorridos más amplios.

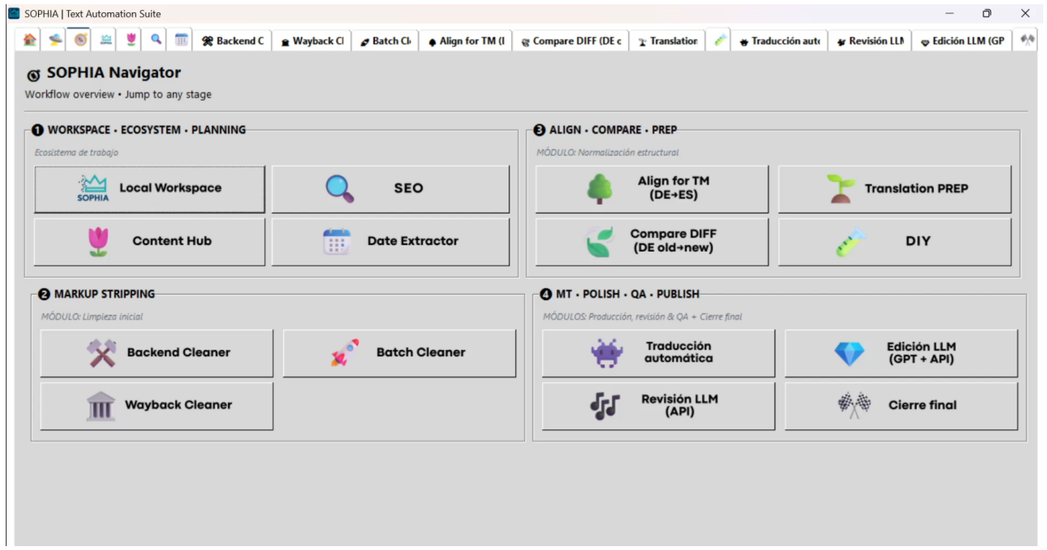
Esta integración no implica que todos los módulos transfieran automáticamente sus resultados: en diferentes momentos del proceso intervienen la selección de archivos, la revisión y las decisiones de la persona que utiliza el sistema.

Principio de diseño

Conectar las operaciones sin imponer un único recorrido ni eliminar los puntos en los que es necesaria la intervención humana.



Home · Un punto de partida común



SOPHIA Navigator · Organización por fases y acceso directo

Conservar hoy lo que necesitaremos después

La limpieza y la segmentación transforman el contenido para facilitar su procesamiento. Sin embargo, el texto normalizado y la tabla de segmentos no contienen necesariamente todas las referencias que necesitaremos al final del recorrido.

Por eso, SOPHIA genera archivos auxiliares que conservan información en el momento en que todavía está disponible y permiten recuperarla en fases posteriores.

Paragraph Map

Origen: segmentación en DIY

Registra los párrafos, su texto, los IDs de los segmentos y cuántos segmentos corresponden a cada párrafo.

Se genera durante la segmentación y permite recuperar posteriormente la distribución del contenido en párrafos.

Clean Structure

Origen: Backend Cleaner

Guarda referencias del documento original: bloques, extractos de texto, información estructural y elementos de formato, como enlaces, cursivas, shortcodes o anclas.

Se genera para las entradas procesadas por Backend Cleaner cuando se han podido extraer bloques.



Paragraph Map (izquierda) y Clean Structure (derecha) · Referencias complementarias

Dos funciones complementarias

Los dos JSON conservan información diferente. Paragraph Map permite reconstruir los párrafos a partir de los segmentos; Clean Structure proporciona referencias adicionales del contenido original cuando necesitamos preparar una reconstrucción anotada.

Cómo se aprovechan en Cierre final

Cierre final utiliza Paragraph Map para reconstruir los párrafos y, cuando dispone de Clean Structure, incorpora referencias del formato y la estructura originales.

De este modo, la información preservada en las primeras fases puede volver a utilizarse varias operaciones después, sin depender exclusivamente del contenido del archivo segmentado.

La decisión de diseño

Conservar las referencias necesarias antes de que las transformaciones del texto dejen de hacerlas accesibles.

Archivos sucesivos para un mismo recorrido

La segmentación genera un primer XLSX con el texto dividido en segmentos identificados. Ese archivo es el punto de partida de las operaciones posteriores, pero no constituye un único libro que se modifica indefinidamente.

A medida que avanza el recorrido, SOPHIA prepara nuevos archivos de trabajo a partir de los resultados anteriores. Cada uno responde a las necesidades de la fase correspondiente y deja preparada la información que aprovecharán las siguientes.

Un recorrido operativo posible

1. DIY

Segmentación

Genera un XLSX con la hoja Segmentos, los identificadores, las columnas de contenido y las notas.

2. OPUS-MT

Preparación y pretraducción

Prepara un nuevo archivo mediante una plantilla que incorpora columnas de trabajo y espacios para referencias de estilo. Después genera una primera versión traducida mediante traducción automática neuronal.

3. Triple Check

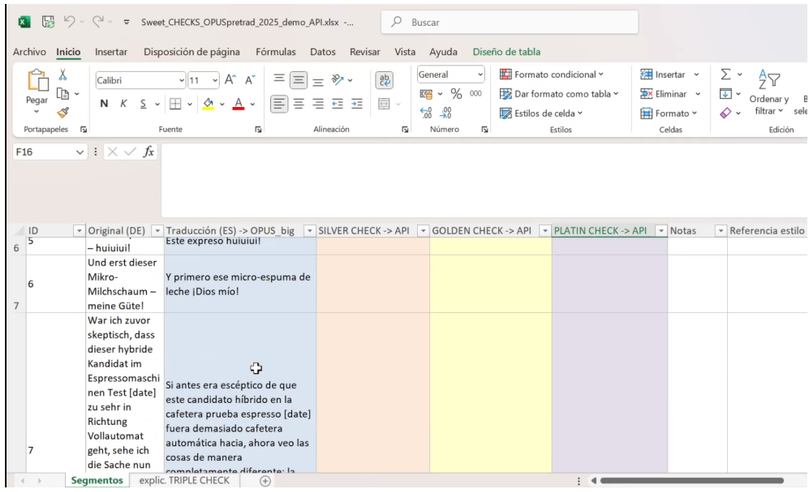
Revisión por capas

Trabaja sobre una copia __FINAL.xlsx y realiza tres pasadas sucesivas de revisión: SILVER, GOLDEN y PLATIN.

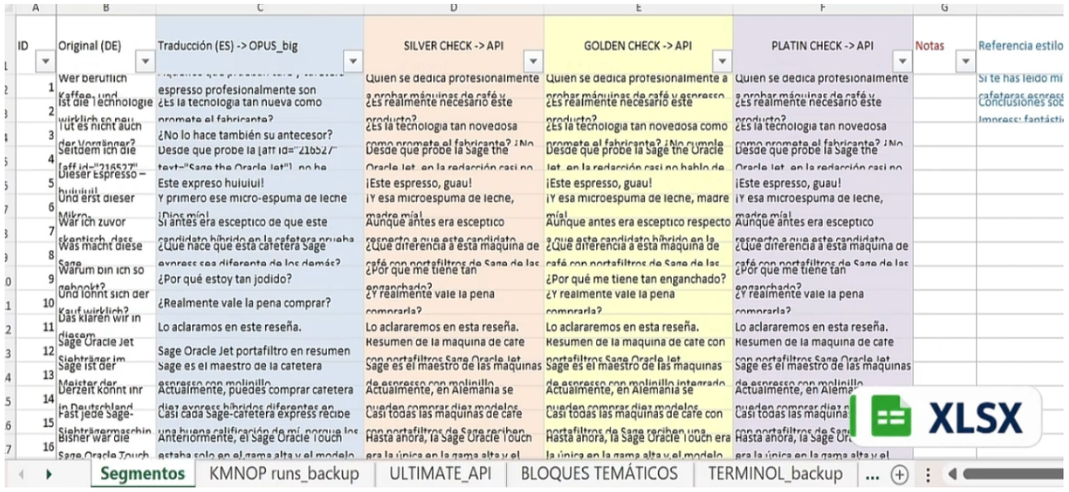
4. Edición LLM (GPT + API)

Edición asistida

Prepara un nuevo archivo _GPT.xlsx para continuar con la edición. Tras las operaciones de revisión y validación correspondientes, permite obtener el TXT final.



OPUS-MT · Pretraducción y columnas preparadas para la revisión



Triple Check · SILVER, GOLDEN y PLATIN cumplimentadas

Esta sucesión de archivos permite que cada módulo aproveche la estructura y los recursos preparados anteriormente, sin tener que reconstruirlos manualmente en cada etapa.

Al mismo tiempo, mantiene diferenciadas las operaciones automatizadas, las propuestas generadas mediante IA y las decisiones editoriales que requieren revisión humana.

La decisión de diseño Preparar cada output teniendo en cuenta los requisitos de las fases posteriores, sin tratar los archivos como resultados aislados.

Cerrar el recorrido recuperando la información del principio

En Cierre final, SOPHIA utiliza Paragraph Map para reconstruir los párrafos a partir del contenido trabajado. Cuando también dispone de Clean Structure, reúne ambas fuentes para generar una vista HTML anotada con referencias del formato y la estructura originales.

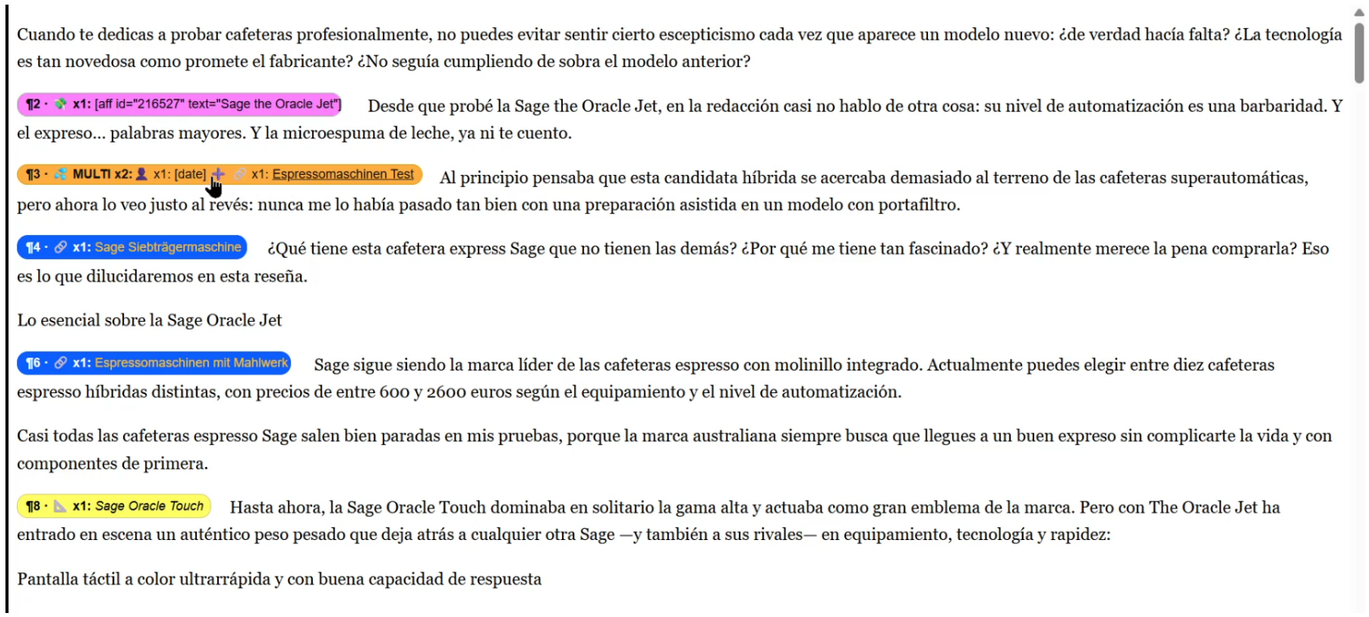
Así, la información preservada durante la limpieza y la segmentación vuelve a estar disponible en una fase muy posterior, cuando el texto ya ha pasado por distintas operaciones de traducción, revisión y edición.

Una vista que facilita la reconstrucción

El HTML anotado presenta los párrafos reconstruidos e incorpora pistas diferenciadas por colores para identificar elementos del documento original, como cursivas, enlaces o shortcodes.

No pretende restaurar automáticamente una copia íntegra del HTML de origen. Su función es ofrecer referencias que permitan revisar y completar el traslado del contenido a WordPress.

La interfaz también permite trabajar por bloques, copiar contenido sin arrastrar las anotaciones visuales y marcar de forma reversible lo que ya se ha trasladado. Estas ayudas complementan el recorrido operativo, aunque su diseño y sus beneficios cognitivos se desarrollarán con mayor profundidad en la tercera muestra de trabajo.



Cuando te dedicas a probar cafeteras profesionalmente, no puedes evitar sentir cierto escepticismo cada vez que aparece un modelo nuevo: ¿de verdad hacía falta? ¿La tecnología es tan novedosa como promete el fabricante? ¿No seguía cumpliendo de sobra el modelo anterior?

¶2 · x1: [aff id="216527" text="Sage the Oracle Jet"] Desde que probé la Sage the Oracle Jet, en la redacción casi no hablo de otra cosa: su nivel de automatización es una barbaridad. Y el expreso... palabras mayores. Y la microespuma de leche, ya ni te cuento.

¶3 · MULTI x2: x1: [date] x1: Espressomaschinen Test Al principio pensaba que esta candidata híbrida se acercaba demasiado al terreno de las cafeteras superautomáticas, pero ahora lo veo justo al revés: nunca me lo había pasado tan bien con una preparación asistida en un modelo con portafiltro.

¶4 · x1: Sage Siebträgermaschine ¿Qué tiene esta cafetera express Sage que no tienen las demás? ¿Por qué me tiene tan fascinado? ¿Y realmente merece la pena comprarla? Eso es lo que dilucidaremos en esta reseña.

Lo esencial sobre la Sage Oracle Jet

¶6 · x1: Espressomaschinen mit Mahlwerk Sage sigue siendo la marca líder de las cafeteras espresso con molinillo integrado. Actualmente puedes elegir entre diez cafeteras espresso híbridas distintas, con precios de entre 600 y 2600 euros según el equipamiento y el nivel de automatización.

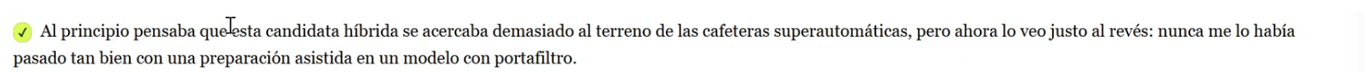
Casi todas las cafeteras espresso Sage salen bien paradas en mis pruebas, porque la marca australiana siempre busca que llegues a un buen expreso sin complicarte la vida y con componentes de primera.

¶8 · x1: *Sage Oracle Touch* Hasta ahora, la Sage Oracle Touch dominaba en solitario la gama alta y actuaba como gran emblema de la marca. Pero con The Oracle Jet ha entrado en escena un auténtico peso pesado que deja atrás a cualquier otra Sage —y también a sus rivales— en equipamiento, tecnología y rapidez:

Pantalla táctil a color ultrarrápida y con buena capacidad de respuesta

Vista HTML anotada: guía de reconstrucción y traslado a WordPress

Indicador reversible de traslado



✓ Al principio pensaba que esta candidata híbrida se acercaba demasiado al terreno de las cafeteras superautomáticas, pero ahora lo veo justo al revés: nunca me lo había pasado tan bien con una preparación asistida en un modelo con portafiltro.

Detalle del indicador tras marcar un bloque como trasladado

Lo que me enseñó este recorrido

Al desarrollar SOPHIA, fui descubriendo que conectar herramientas no consiste únicamente en automatizar las transiciones entre ellas. También implica comprender las dependencias del proceso y anticipar qué información necesitará cada fase, incluso cuando todavía quedan varias operaciones por delante.

Preservar esas relaciones permite mantener la continuidad del trabajo sin renunciar a la flexibilidad de los recorridos ni a la intervención humana.