

AlzoLab

Laboratorio de lingüística de corpus en el navegador, sin escribir código

EN PRODUCCIÓN

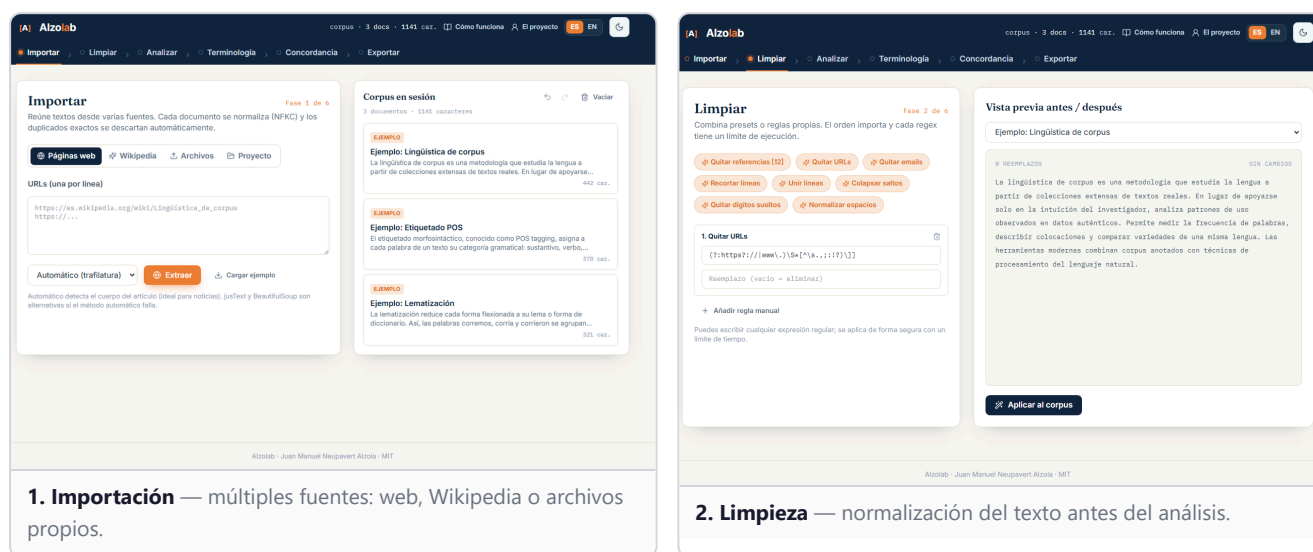
Python · Streamlit · spaCy

QUÉ HACE

Convierte un conjunto de textos en un corpus analizable sin escribir una línea de código: importación desde varias fuentes (web, Wikipedia, archivos propios), limpieza y normalización, análisis morfosintáctico con spaCy, extracción automática de términos relevantes (C-value) y concordancias KWIC para ver cada término en su contexto real de uso.

Todo el flujo — desde subir los textos hasta exportar los resultados — ocurre dentro de la misma interfaz, pensado para alguien sin experiencia de programación.

CÓMO SE VE EN USO



The image displays two side-by-side screenshots of the AlzoLab web application interface, illustrating the initial steps of the corpus analysis workflow.

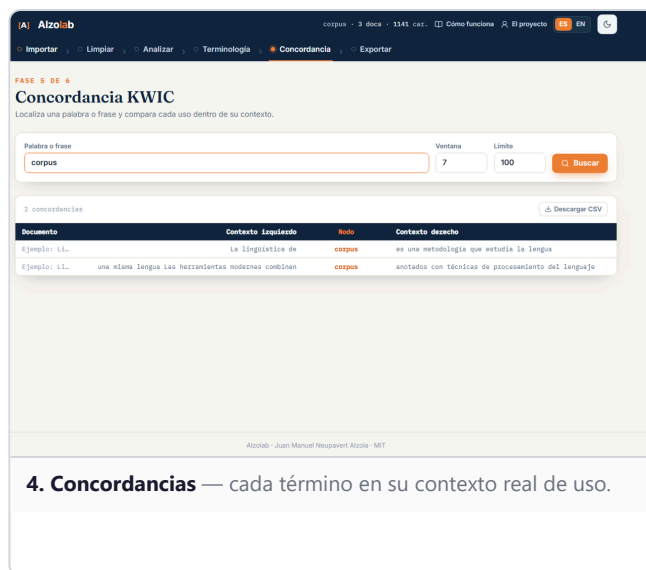
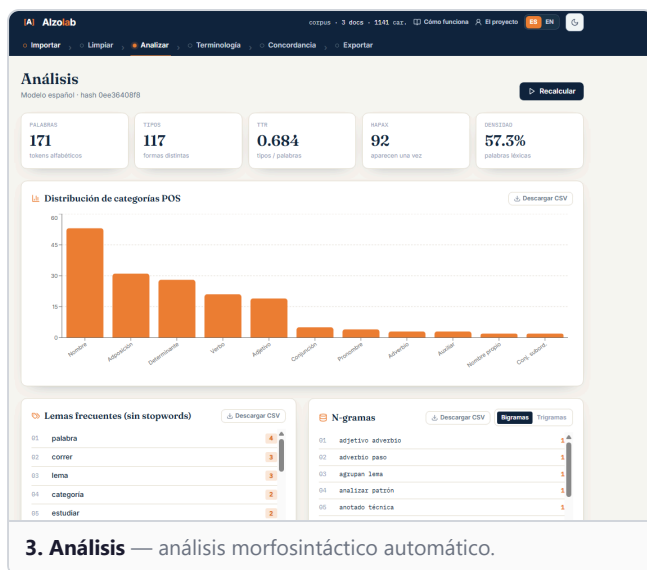
Left Screenshot (Importar): This panel shows the 'Importar' (Import) step, labeled 'Fase 1 de 6'. It includes a navigation bar with tabs: Importar, Limpiar, Analizar, Terminología, Concordancia, and Exportar. The main content area is divided into two sections. The left section, 'Importar', describes importing text from various sources (web, Wikipedia, archives, project) and normalizing (NFKC) and deduplicating. It features a 'URLs (una por línea)' input field with a text area containing a Wikipedia URL. Below this is a dropdown menu set to 'Automático (trafilitura)' and an 'Extraer' button. The right section, 'Corpus en sesión', shows a summary of the current corpus: '3 documentos · 1541 caracteres'. It includes three example cards: 'Ejemplo: Lingüística de corpus', 'Ejemplo: Etiquetado POS', and 'Ejemplo: Lematización', each with a brief description and a 'Ver' link.

Right Screenshot (Limpiar): This panel shows the 'Limpiar' (Clean) step, labeled 'Fase 2 de 6'. It includes the same navigation bar. The main content area is divided into two sections. The left section, 'Limpiar', describes combining presets or custom rules to clean the text. It features a '1. Quitar URLs' section with a text input field containing a regex pattern and a 'Reemplazo' field with the value 'vacío = a línea nueva'. Below this is a '+ Añadir regla manual' button. The right section, 'Vista previa antes / después', shows a preview of the text before and after cleaning. It includes a dropdown menu set to 'Ejemplo: Lingüística de corpus' and a 'Aplicar al corpus' button.

Footer: Both screenshots have a footer that reads 'AlzoLab · Juan Manuel Neupavert Alzola · MIT'.

1. Importación — múltiples fuentes: web, Wikipedia o archivos propios.

2. Limpieza — normalización del texto antes del análisis.



SERVICIOS QUE PUEDO OFRECER CON ESTO

- **Extracción de terminología especializada** de un fondo documental (técnico, jurídico, médico) como base para un glosario o una memoria de traducción.
- **Análisis de corpus a medida** para traductores, editoriales o equipos de contenido que trabajan con grandes volúmenes de texto.