Saturn Studio – Sitefetch



itefetch es una herramienta diseñada para descargar un sitio web completo y guardarlo como archivo de texto, lo que facilita su uso en modelos de inteligencia artificial. Permite especificar páginas concretas a descargar y ofrece opciones de concurrencia para optimizar el proceso. Además, proporciona una API para desarrolladores y utiliza la biblioteca mozilla/readability para extraer eficientemente el contenido web.

¿Qué es este módulo?

Este módulo permite conectar Saturn Studio con **Sitefetch**, una herramienta diseñada para descargar sitios web completos y guardarlos como archivos de texto. Esto facilita su uso en modelos de inteligencia artificial, extracción de contenido y análisis de datos web. También ofrece opciones de concurrencia, formato Markdown y uso de selectores CSS para precisión en el scraping.

Descripción de los comandos

Fetch a Site

Parámetro	Descripción	Ejemplo
URL del sitio	Página web desde la cual se extraerá el contenido.	https://rocketbot.com/es/
Concurrencia	Número de solicitudes simultáneas permitidas.	10
Selector de contenido	Selector CSS para extraer el contenido deseado.	div
Límite	Número máximo de páginas a procesar.	10
Markdown	Si está activado, el contenido se descargará en formato Markdown.	true
Asignar resultado a variable	Variable donde se almacenará el resultado.	{ “filename”: “sitefetch_1743…”, “id”: “…”, “size”: 36559, “provider”: “sitefetch” }

Descarga el contenido de un sitio web completo o parcial, con opciones de

formato y concurrencia, y lo guarda como archivo de texto o Markdown.

Ejemplos de uso

- **Fetch a Site:** Descargar contenido desde https://rocketbot.com/es/ con un selector CSS específico y guardar el resultado como archivo Markdown para análisis posterior.
- **Limit + Concurrency:** Descargar solo las primeras 10 páginas del sitio, procesadas en paralelo para mayor velocidad.

Requisitos

1. Tener acceso a la URL pública del sitio web deseado.
2. Usar un selector CSS válido (por ejemplo: div, article, main) para extraer contenido relevante.
3. Contar con una credencial y configuración correcta de Sitefetch si se accede mediante su API privada.
4. Verificar que el sitio no tenga restricciones técnicas (como captchas o bloqueos de scrapers).