Instalação Firecrawl docker

Link: https://github.com/mendableai/firecrawl/tree/main?tab=readme-ov-file

Follow on X Follow on LinkedIn Join our Discord

Firecrawl

Empower your AI apps with clean data from any website. Featuring advanced scraping, crawling, and data extraction capabilities.

This repository is in development, and we’re still integrating custom modules into the mono repo. It's not fully ready for self-hosted deployment yet, but you can run it locally.

What is Firecrawl?

Firecrawl is an API service that takes a URL, crawls it, and converts it into clean markdown or structured data. We crawl all accessible subpages and give you clean data for each. No sitemap required. Check out our documentation.

Pst. hey, you, join our stargazers :)

GitHub stars

How to use it?

We provide an easy to use API with our hosted version. You can find the playground and documentation here. You can also self host the backend if you'd like.

Check out the following resources to get started:

- ☑ API
- ☑ Python SDK
- ☑ Node SDK
- ☑ Go SDK
- ☑ Rust SDK
- ☑ Langchain Integration □□□□
- ☑ Langchain JS Integration □□□□
- ☑ Llama Index Integration □□
- ☑ Dify Integration
- ☑ Langflow Integration
- ☑ Crew.ai Integration
- ☑ Flowise AI Integration
- ☑ Composio Integration
- ☑ PraisonAI Integration
- ☑ Zapier Integration
- ☑ Cargo Integration
- ☑ Pipedream Integration
- ☑ Pabbly Connect Integration
- ☐ Want an SDK or Integration? Let us know by opening an issue.

To run locally, refer to guide here.

API Key

To use the API, you need to sign up on Firecrawl and get an API key.

Crawling

Used to crawl a URL and all accessible subpages. This submits a crawl job and returns a job ID to check the status of the crawl.

```
curl -X POST https://api.firecrawl.dev/v1/crawl \
    -H 'Content-Type: application/json' \
    -H 'Authorization: Bearer fc-YOUR_API_KEY' \
    -d '{
      "url": "https://docs.firecrawl.dev",
      "limit": 100,
```

```
      "scrapeOptions": {
        "formats": ["markdown", "html"]
      }
    }'
```

Returns a crawl job id and the url to check the status of the crawl.

```
{
  "success": true,
  "id": "123-456-789",
  "url": "https://api.firecrawl.dev/v1/crawl/123-456-789"
}
```

Check Crawl Job

```
curl -X GET https://api.firecrawl.dev/v1/crawl/123-456-789 \
  -H 'Content-Type: application/json' \
  -H 'Authorization: Bearer YOUR_API_KEY'
```

```
{
  "status": "completed",
  "total": 36,
  "creditsUsed": 36,
  "expiresAt": "2024-00-00T00:00:00.000Z",
  "data": [
    {
      "markdown": "[Firecrawl Docs home page![light logo](https://mintlify.s3-us-west-1.amazonav
      "html": "<!DOCTYPE html><html lang=\"en\" class=\"js-focus-visible lg:[--scroll-mt:9.5rem]
      "metadata": {
        "title": "Build a 'Chat with website' using Groq Llama 3 | Firecrawl",
        "language": "en",
        "sourceURL": "https://docs.firecrawl.dev/learn/rag-llama3",
        "description": "Learn how to use Firecrawl, Groq Llama 3, and Langchain to build a 'Chat
        "ogLocaleAlternate": [],
        "statusCode": 200
      }
    }
  ]
}
```

Scraping

Used to scrape a URL and get its content in the specified formats.

```
curl -X POST https://api.firecrawl.dev/v1/scrape \
    -H 'Content-Type: application/json' \
```

```
    -H 'Authorization: Bearer YOUR_API_KEY' \
    -d '{
      "url": "https://docs.firecrawl.dev",
      "formats" : ["markdown", "html"]
    }'
```

Response:

Revision #1
Created 2024-09-29 02:46:47 UTC by Administrador
Updated 2024-09-29 02:57:51 UTC by Administrador