main ⌄   Branches   Tags

Go to file   <> Code ⌄   ...

.claude		
.github		
crawl4ai		
deploy/docker		
docs		
prompts		
tests		
.env.txt		
.gitattributes		
.gitignore		
CHANGELOG.md		
CODE_OF_CONDUCT.md		
CONTRIBUTORS.md		
Dockerfile		
JOURNAL.md		
LICENSE		
MANIFEST.in		
MISSION.md		
PROGRESSIVE_CRAWLI...		
README-first.md		
README.md		
ROADMAP.md		
SPONSORS.md		
cliff.toml		
docker-compose.yml		
mkdocs.yml		
pyproject.toml		
requirements.txt		
setup.cfg		

About

🚀 🤖 Crawl4AI: Open-source LLM Friendly Web Crawler & Scraper. Don't be shy, join here: https://discord.gg/jP8KfhDhyN

🔗 crawl4ai.com

📖 Readme
⚖️ Apache-2.0 license
💠 Code of conduct
〰️ Activity
☆ 56.1k stars
👁 307 watching
⑂ 5.6k forks

Report repository

Releases 11

🏷 Release v0.7.7 Latest
5 days ago

+ 10 releases

Sponsor this project

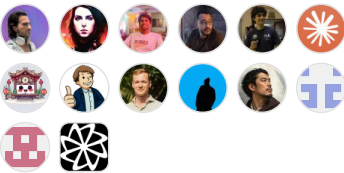👤 unclecode UncleCode

♡ Sponsor

Learn more about GitHub Sponsors

Packages

No packages published

Used by 2.7k

+ 2,665

Contributors 59

+ 45 contributors

Languages

● **Python** 99.2%   ● **Other** 0.8%

📖 README   ⚖️ Code of conduct   ⚖️ Apache-2.0 license

🚀🤖 Crawl4AI: Open-source LLM Friendly Web Crawler & Scraper.

Crawl4AI turns the web into clean, LLM ready Markdown for RAG, agents, and data pipelines. Fast, controllable, battle tested by a 50k+ star community.

✨ [Check out latest update v0.7.7](#)

✨ **New in v0.7.7**: Complete Self-Hosting Platform with Real-time Monitoring! Enterprise-grade monitoring dashboard, comprehensive REST API, WebSocket streaming, smart browser pool management, and production-ready observability. Full visibility and control over your crawling infrastructure. [Release notes →](#)

✨ Recent v0.7.6: Complete Webhook Infrastructure for Docker Job Queue API! Real-time notifications for both `/crawl/job` and `/llm/job` endpoints with exponential backoff retry, custom headers, and flexible delivery modes. No more polling! [Release notes →](#)

✨ Previous v0.7.5: Docker Hooks System with function-based API for pipeline customization, Enhanced LLM Integration with custom providers, HTTPS Preservation, and multiple community-reported bug fixes. [Release notes →](#)

▶ 🤓 My Personal Story

▶ Why developers pick Crawl4AI

🚀 Quick Start

1. Install Crawl4AI:

```
# Install the package
pip install -U crawl4ai

# For pre release versions
pip install crawl4ai --pre

# Run post-installation setup
crawl4ai-setup

# Verify your installation
crawl4ai-doctor
```

If you encounter any browser-related issues, you can install them manually:

```
python -m playwright install --with-deps chromium
```

2. Run a simple web crawl with Python:

```python
import asyncio
from crawl4ai import *

async def main():
    async with AsyncWebCrawler() as crawler:
        result = await crawler.arun(
            url="https://www.nbcnews.com/business",
        )
        print(result.markdown)

if __name__ == "__main__":
    asyncio.run(main())
```

3. Or use the new command-line interface:

```
# Basic crawl with markdown output
crwl https://www.nbcnews.com/business -o markdown

# Deep crawl with BFS strategy, max 10 pages
crwl https://docs.crawl4ai.com --deep-crawl bfs --max-pages 10

# Use LLM extraction with a specific question
crwl https://www.example.com/products -q "Extract all product prices"
```

💖 Support Crawl4AI

> 🎉 **Sponsorship Program Now Open!** After powering 51K+ developers and 1 year of growth, Crawl4AI is launching dedicated support for **startups** and **enterprises**. Be among the first 50 **Founding Sponsors** for permanent recognition in our Hall of Fame.

Crawl4AI is the #1 trending open-source web crawler on GitHub. Your support keeps it independent, innovative, and free for the community — while giving you direct access to premium benefits.

♡ **BECOME A SPONSOR**

○ **CURRENT SPONSORS** **11**

🤝 Sponsorship Tiers

- 🌱 **Believer ($5/mo)** — Join the movement for data democratization
- 🚀 **Builder ($50/mo)** — Priority support & early access to features
- 💼 **Growing Team ($500/mo)** — Bi-weekly syncs & optimization help
- 🗄️ **Data Infrastructure Partner ($2000/mo)** — Full partnership with dedicated support
  Custom arrangements available - see SPONSORS.md for details & contact

Why sponsor?
No rate-limited APIs. No lock-in. Build and own your data pipeline with direct guidance from the creator of Crawl4AI.

See All Tiers & Benefits →

✨ Features

▶ 📝 **Markdown Generation**

▶ 📊 **Structured Data Extraction**

▶ 🌐 **Browser Integration**

▶ 🔍 **Crawling & Scraping**

▶ 🚀 **Deployment**

▶ 🎯 Additional Features

Try it Now!

✨ Play around with this [CO Open in Colab]

✨ Visit our [Documentation Website](#)

Installation 🛠️

Crawl4AI offers flexible installation options to suit various use cases. You can install it as a Python package or use Docker.

▶ 🐍 Using pip

▶ 🐳 Docker Deployment

🔬 Advanced Usage Examples 🔬

You can check the project structure in the directory [docs/examples](#). Over there, you can find a variety of examples; here, some popular examples are shared.

▶ 📝 Heuristic Markdown Generation with Clean and Fit Markdown

▶ 🖥️ Executing JavaScript & Extract Structured Data without LLMs

▶ 📚 Extracting Structured Data with LLMs

▶ 🧁 Using Your own Browser with Custom User Profile

> 💡 **Tip:** Some websites may use **CAPTCHA** based verification mechanisms to prevent automated access. If your workflow encounters such challenges, you may optionally integrate a third-party CAPTCHA-handling service such as [CapSolver](#). They support reCAPTCHA v2/v3, Cloudflare Turnstile, Challenge, AWS WAF, and more. Please ensure that your usage complies with the target website's terms of service and applicable laws.

✨ Recent Updates

▶ **Version 0.7.7 Release Highlights - The Self-Hosting & Monitoring Update**

▶ **Version 0.7.5 Release Highlights - The Docker Hooks & Security Update**

▶ **Version 0.7.4 Release Highlights - The Intelligent Table Extraction & Performance Update**

▶ **Version 0.7.3 Release Highlights - The Multi-Config Intelligence Update**

▶ **Version 0.7.0 Release Highlights - The Adaptive Intelligence Update**

Version Numbering in Crawl4AI

Crawl4AI follows standard Python version numbering conventions (PEP 440) to help users understand the stability and features of each release.

▶ 📈 Version Numbers Explained

📖 Documentation & Roadmap

> 🚨 **Documentation Update Alert**: We're undertaking a major documentation overhaul next week to reflect recent updates and improvements. Stay tuned for a more comprehensive and up-to-date guide!

For current documentation, including installation instructions, advanced features, and API reference, visit our [Documentation Website](#).

To check our development plans and upcoming features, visit our [Roadmap](#).

▶ 📈 **Development TODOs**

🤝 Contributing

We welcome contributions from the open-source community. Check out our [contribution guidelines](#) for more information.

I'll help modify the license section with badges. For the halftone effect, here's a version with it:

Here's the updated license section:

📄 License & Attribution

This project is licensed under the Apache License 2.0, attribution is recommended via the badges below. See the [Apache 2.0 License](#) file for details.

Attribution Requirements

When using Crawl4AI, you must include one of the following attribution methods:

▶ 📈 **1. Badge Attribution (Recommended)**

▶ 📖 **2. Text Attribution**

📚 Citation

If you use Crawl4AI in your research or project, please cite:

```
@software{crawl4ai2024,
  author = {UncleCode},
  title = {Crawl4AI: Open-source LLM Friendly Web Crawler & Scraper},
  year = {2024},
  publisher = {GitHub},
  journal = {GitHub Repository},
  howpublished = {\url{https://github.com/unclecode/crawl4ai}},
  commit = {Please use the commit hash you're working with}
}
```

Text citation format:

```
UncleCode. (2024). Crawl4AI: Open-source LLM Friendly Web Crawler & Scraper [Computer software].
GitHub. https://github.com/unclecode/crawl4ai
```

📧 Contact

For questions, suggestions, or feedback, feel free to reach out:

- GitHub: [unclecode](#)
- Twitter: [@unclecode](#)
- Website: [crawl4ai.com](#)

Happy Crawling! 🕸️ 🚀

🗾 Mission

Our mission is to unlock the value of personal and enterprise data by transforming digital footprints into structured, tradeable assets. Crawl4AI empowers individuals and organizations with open-source tools to extract and structure data, fostering a shared data economy.

We envision a future where AI is powered by real human knowledge, ensuring data creators directly benefit from their contributions. By democratizing data and enabling ethical sharing, we are laying the foundation for authentic AI advancement.

🌟 Current Sponsors

🏢 Enterprise Sponsors & Partners

Our enterprise sponsors and technology partners help scale Crawl4AI to power production-grade data pipelines.

Company	About	Sponsorship Tier
scrapeless	Scrapeless is the best full-stack web scraping toolkit offering Scraping API, Scraping Browser, Web Unlocker, Captcha Solver, and Proxies, designed to handle all your data collection needs.	🥈 Silver
CAPSOLVER	AI-powered Captcha solving service. Supports all major Captcha types, including reCAPTCHA, Cloudflare, and more	🥉 Bronze
KIPO AI	Helps engineers and buyers find, compare, and source electronic & industrial parts in seconds, with specs, pricing, lead times & alternatives.	🥇 Gold
kidoCODE [KidoCode](#)	Kidocode is a hybrid technology and entrepreneurship school for kids aged 5–18, offering both online and on-campus education.	🥇 Gold
aleph null	Singapore-based Aleph Null is Asia's leading edtech hub, dedicated to student-centric, AI-driven education—empowering learners with the tools to thrive in a fast-changing world.	🥇 Gold

👩‍🤝 Individual Sponsors

A heartfelt thanks to our individual supporters! Every contribution helps us keep our opensource mission alive and thriving!

Want to join them? [Sponsor Crawl4AI →](#)

Star History

Star History

unclecode/crawl4ai