Python scrapers authentication

Hackathon

Jaime Jimenez - Atish Joottun
Thibault Meunier - Antonin Vlcek
IETF 123, July 2025, Madrid

Goals

Scraping libraries in Python

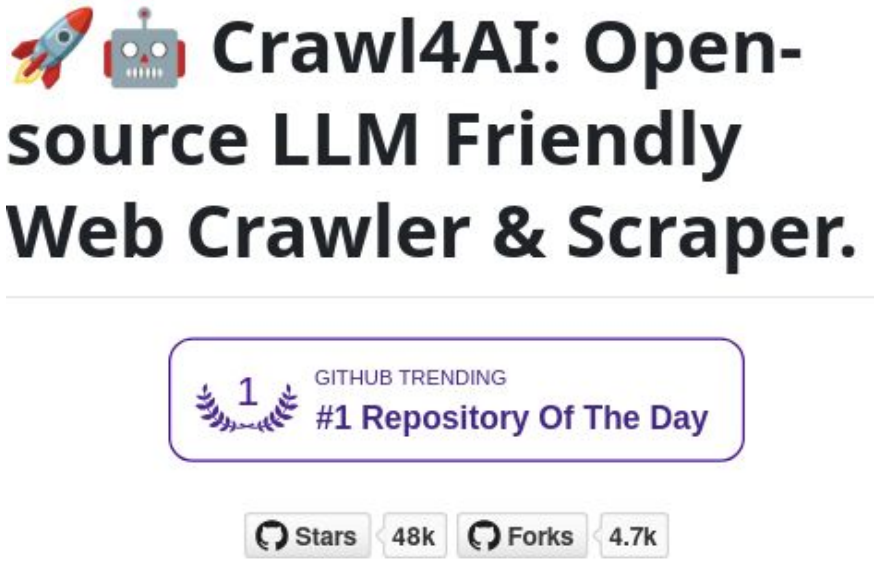
1. crawl4ai
2. scrapy

What we do

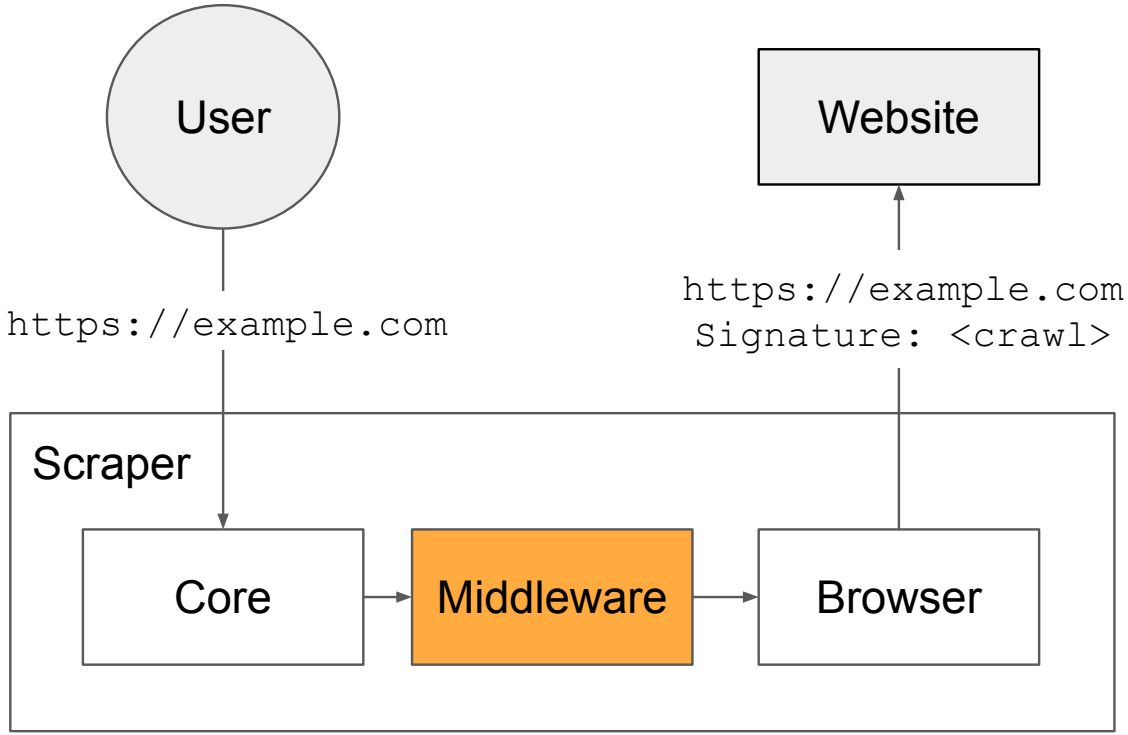
1. Integrate with webbotauth
2. Interop with Typescript

Non-goal

1. LLM integration



Inspection into scrapers



Inspection into scrapers

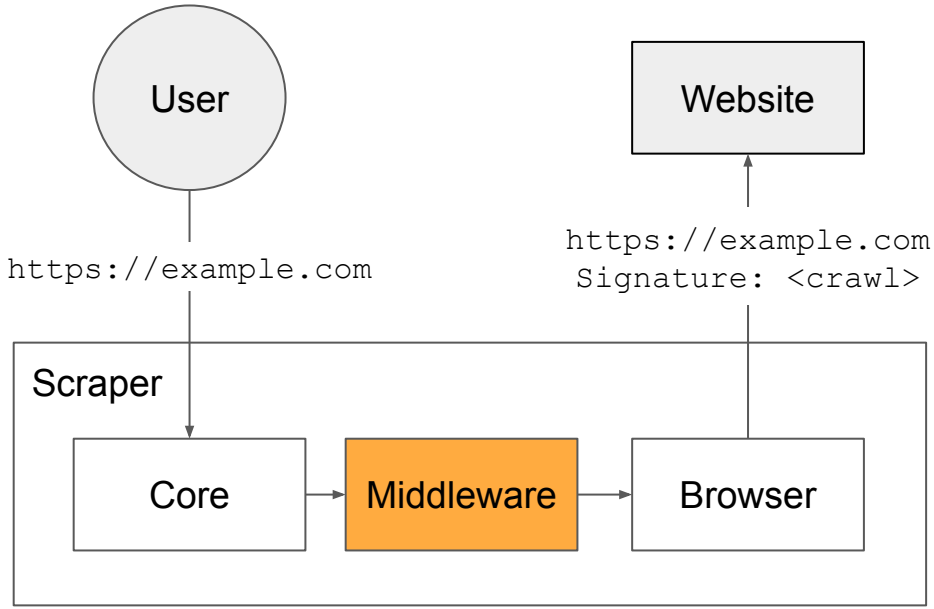
Simplify scraper integration

1. crawl4ai: hook
2. scrapy: download middleware

What we do

1. Develop bot-auth package
2. Examples for both scrapers
3. WIP exposure with CoAP

Demo!!!!!



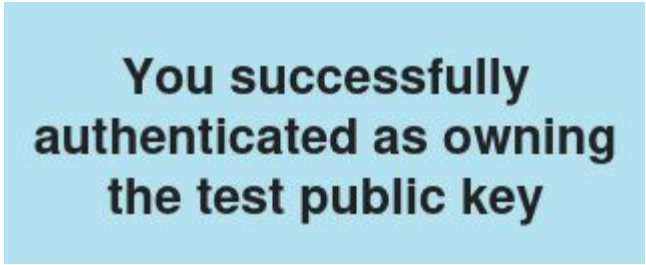
$ uv run main.py https://http-message-signatures-example.research.cloudflare.com/debug
[INIT].... → Crawl4AI 0.7.1
[HOOK] before_goto - Navigate to https://http-message-signatures-example.research.cloudflare.com/debug
[FETCH]... ↓ https://http-message-signatures-example.research.cloudflare.com/debug | ✓ | ⏱: 0.38s
[SCRAPE].. ◆ https://http-message-signatures-example.research.cloudflare.com/debug | ✓ | ⏱: 0.00s
[COMPLETE] ● https://http-message-signatures-example.research.cloudflare.com/debug | ✓ | ⏱: 0.39s
[RESPONSE] You successfully authenticated as owning the test public key
user-agent: Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 Chrome/116.0.0.0 Safari/537.36
signature:
pyhms=:jC4nC7mwq9qhgcpn169kda6FOOO7yF5vUdgu0JBalTAiDxdb57PKU6b8OpAXBbisVyIMRD6RKTASyM4CBLotBg==:
signature-agent: http-message-signatures-example.research.cloudflare.com
signature-input: pyhms=("@authority" "signature-agent");created=1753005232;keyid="compute-jwk-
thumbprint";expires=1753005532;tag="web-bot-auth";alg="ed25519"

Conclusion

✅ Interop successful

🐍 Python learnings: uv, ruff, black

🤝 New participants + remote

👨‍💻 cyberstormdotmu/Bot-Authentication

📅 BoF on Monday at 17:00 (UTC+2)