MCP SERVER | NO CODE | CLOUD HOSTED

Web Scraper MCP

Give your AI agent live web access and clean data extraction.

Web Scraper is an MCP that gives your AI agent direct read access to live web pages. It lets your agent pull clean, usable text from any URL, stripping away ads and site clutter. You can also extract structured metadata like titles and links, or crawl entire documentation sites up to ten pages deep.

A+ Quality Score 98.33/100

web-crawling | markdown-conversion | data-extraction | reader-view | content-parsing | url-fetching

MAINTAINED BY VINKIUS CLOUD FOR AI AGENTS · JUNE 2026
vinkius.com/mcp/web-scraper

SCAN TO VIEW

The connectivity layer between AI and the world's software.

Vinkius sits between AI and every application. All communication passes through Vinkius Cloud via the Model Context Protocol (MCP) — with governance, observability, and security at every layer.

Your AI Connections Run Through Vinkius Cloud

The world's largest managed MCP catalog

Vinkius is the connectivity layer where AI connects to the software your business already runs. We handle the hosting, the security, the credentials, the uptime — you get agents that actually do things.

We operate the world's largest managed MCP catalog. Major SaaS platforms, CRMs, databases, and cloud providers — running, monitored, production-ready. This MCP server is hosted and maintained by the Vinkius Cloud for AI Agents.

> *The agent doesn't manage credentials, doesn't manage uptime, doesn't manage security. Vinkius does.*
>
> — Architecture principle

Four Pillars of the Vinkius Runtime

01 — Security by design

Credentials stay encrypted at rest via AES-256. The AI agent never touches raw keys — they're injected into a sandboxed V8 isolate at runtime. Actions are logged, and connections have an emergency kill switch.

02 — Built on MCP Fusion

This MCP server was built with **MCP Fusion**, the open-source framework (Apache 2.0) that powers the entire Vinkius catalog. Schema-as-firewall strips undeclared fields, compiled PII redaction runs at zero overhead, and cryptographic lockfiles produce git-diffable audit trails.

03 — Deterministic observability

Eight immutable metrics per endpoint: request volume, p95 latency, error rate, active connections, cost attribution. A live payload feed logs every tool call with mutation detection.

04 — Autonomous operations

Servers are deployed, monitored, and patched autonomously. New capabilities and security patches ship weekly. Zero-downtime deployments ensure continuous availability across all managed MCP servers.

AES-256	Ed25519	24h TTL	V8 Isolate
Encryption at rest	PKI vault signatures	Ephemeral session keys	Sandboxed execution

One Token. Instant Access.

Every MCP server on Vinkius is accessed through a **Connection Token**. Tokens are generated in the cloud dashboard and produce a unique MCP endpoint URL. Paste this URL into any MCP-compatible client — no SDK required.

A single token can serve **multiple AI clients simultaneously**, or you can issue separate tokens per client for granular access control. Each token tracks its own request count, last activity timestamp, and can be individually enabled or revoked.

MCP ENDPOINT

```
https://edge.vinkius.com/{token}/mcp
```

Claude · Cursor · VS Code · Windsurf · Grok · Gemini

Security Is the Architecture

Security in Vinkius is not a feature — it's the foundation of the runtime. The gateway enforces multiple independent protection layers between AI agents and third-party APIs.

01 — Ed25519 PKI Vault

Every workspace has an Ed25519 Master Key. Session keys are generated ephemerally (24h TTL) and signed by the Master Key. Credentials never leave the vault boundary.

02 — V8 Isolate Sandboxing

Tool code runs inside isolated-vm V8 isolates with 64 MB memory caps and per-request timeouts. No filesystem access, no network access except through the SSRF-guarded fetch bridge.

03 — SSRF Guard

All outbound HTTP requests are DNS-resolved and validated before execution. Private IP ranges (10.x, 172.16-31.x, 192.168.x, AWS metadata 169.254.x) are blocked at the network layer.

04 — DLP & PII Redaction

A ResponseGuard pipeline intercepts every tool response. Configurable redaction patterns strip sensitive fields (emails, SSNs, card numbers) before data reaches the AI agent.

05 — Cryptographic Audit Trail

Every request is signed into a SHA-256 hash chain with Ed25519 signatures. Events form a tamper-proof, SIEM-exportable forensic record.

06 — Honeytoken Trap System

Phantom credentials are injected into isolated environments. If a honeytoken is used outside Vinkius infrastructure, the server is quarantined instantly.

Emergency Kill Switch

EU AI Act Art. 14(1) Compliant

The kill switch is an **emergency halt** mechanism — not a simple toggle. When triggered, it executes three actions atomically:

01 — Server deactivated

The MCP server is immediately taken offline across the entire cluster.

02 — All tokens revoked

Every connection token is invalidated. Total lockout — reconnection blocked until new tokens are issued.

03 — WebSocket connections killed

Active connections terminated via Redis pubsub broadcast. Propagates to every runtime node in the cluster.

Full Visibility. Zero Guesswork.

The Vinkius cloud dashboard includes a full MCP Governance suite — real-time analytics and security controls for production AI operations.

Control Plane

KPI dashboard with request volume, latency, success rate, token consumption, and AI-generated operational briefings.

Agent Activity

Which AI clients are connecting, how often, and what they're doing — real-time session tracking.

FinOps

Cost tracking per tool, payload compression savings, budget optimization signals, and consumption trends.

Tool Health

Slowest and most error-prone tools, with actionable root-cause insights and performance baselines.

Firewall & DLP

PII redaction activity, sensitive data protection counters, and security event timeline.

Incident Log

Error trends, failure rates, status-code breakdowns, and forensic audit trail access.

> Get started at **cloud.vinkius.com** — connect your AI agent in under 60 seconds.

Web Scraper MCP

5 tools available
Cloud-hosted on Vinkius

Stop letting your agent guess facts. This MCP connects your AI client directly to the public internet, giving it a real-time source of information. Instead of hallucinating, your agent reads live articles, parses complex technical documentation, and pulls clean text from any link you provide. You can convert cluttered webpages into pristine Markdown using Mozilla Readability logic. Need to compare sources? Use the batch reading feature to pull up to ten different URLs simultaneously. For developers, this means pointing your agent at a new library's API docs and having it write code based on the absolute latest syntax. The whole catalog of tools is hosted and managed by Vinkius, so you connect once and get access to all web data capabilities.

Core Capabilities

01 — Clean Article Reading

Your agent strips away ads and site navigation from any webpage, returning only the main article content as clean Markdown.

02 — Metadata Collection

The tool extracts structured data like SEO titles, descriptions, canonical links, and all outbound hyperlinks without downloading the page body.

03 — Site Deep Crawling

Your agent automatically navigates a starting URL, crawling up to ten pages deep to map out an entire documentation site or wiki.

04 — Bulk Data Fetching

You can process multiple web sources at once, fetching and comparing content from up to ten different URLs in parallel.

One Click on Vinkius — From Prompt to Execution

Available at **vinkius.com/mcp/web-scraper** — connect your AI agent in three steps.

01 Subscribe to this MCP on Vinkius. No API keys or authentication are required.

02 Simply paste a web link into your chat and tell your agent what to do, such as 'read this URL' or 'crawl this documentation'.

03 Your AI client executes the request, fetching the data and returning it directly to your conversation.

The bottom line is you get real-time web data delivered into your workflow without any setup or keys.

Built For

This MCP is for anyone whose job requires knowing what's happening right now on the live web. Think of researchers who need to synthesize information from multiple sources, developers needing the latest API syntax, or SEO specialists auditing sites in real-time.

Technical Writer

They use this MCP to pull documentation from a new SDK's website and feed it directly into their agent, ensuring all tutorials reflect the current best practices.

Research Analyst

They drop a handful of disparate links—from academic journals or news sites—and ask the AI to synthesize a comprehensive summary from all the fetched data.

Front-End Developer

They point their agent at a new framework's documentation and use the MCP to make sure the generated code uses the absolute latest syntax available on the web.

What Changes When You Connect

01 Instead of relying on stale training data, you let the agent read real-time articles. This eliminates factual hallucinations entirely.

02 The `read` tool converts any messy website into pristine Markdown. You get readable content instantly, perfect for documentation or blog posts.

03 Need to audit a site? Use the `extract` tool to pull only the metadata—titles, descriptions, and OG tags—without downloading the whole page body.

04 Comparing sources is easy with `batch_read`. You can feed up to ten URLs at once, allowing your agent to compare concepts or summarize multiple articles in one go.

05 For deep research, use the `crawl` tool. Give it a single documentation hub link and let your agent map out every related page automatically.

Real-World Applications

Comparing two product architectures

A developer needs to know if 'React-first' or 'HTML-first' is better for their client. They ask their agent to run `batch_read` on both competing articles, allowing the AI to compare them side-by-side and give a definitive recommendation.

Auditing competitor websites

An SEO specialist uses `extract` on five competitor sites. The agent quickly pulls all metadata—titles, descriptions, canonical tags—enabling the specialist to identify weak spots in their own site's optimization.

Researching a niche topic

A researcher drops 15 links related to quantum computing. They ask the agent to use `read` on each, and then summarize the entire collection of clean Markdown text into one coherent report.

Mapping out an old wiki

An internal team uses `crawl` on their company's legacy documentation hub. The agent maps every related page up to ten deep, giving the team a complete structure map before migrating the content.

Patterns to Avoid

Using the AI for general knowledge

✗ AVOID

Asking your agent, 'What is the current best practice for modern web design?' and getting an answer based on data from 2021.

✓ INSTEAD

Always use this MCP. Point your agent to a specific documentation link and ask it to `read` or `extract` the content there. This guarantees you're using real-time information.

Trying to get metadata piecemeal

✗ AVOID

Asking for titles, then separately asking for links, resulting in two disconnected steps and multiple API calls.

✓ INSTEAD

Use the `extract` tool first. It pulls all the structured data—titles, descriptions, *and* link counts—in one single, efficient request.

Forgetting about site depth

✗ AVOID

Asking the agent to summarize a massive wiki but only getting content from the landing page.

✓ INSTEAD

Use the `crawl` tool. This tells your agent not just to read the starting URL, but to automatically map and process related pages up to 10 levels deep.

The Right Fit

Use this MCP if you need verifiable information from a live web source. If the core of your task is reading articles, pulling structured metadata (titles/links), or mapping out documentation structures, this tool is what you want. Don't use it if you just need to process data that already exists within a private database—use a dedicated database connector instead. Also, don't rely on it for general knowledge; always provide the target URL first. If you only need to pull links from one page and nothing else, `list_links` is sufficient. But if you need both the clean text *and* the links, you need this MCP.

Finding facts on the live web used to be a mess of tabs.

Before connecting your agent to a web scraper MCP, gathering information meant opening ten different browser tabs. You'd copy text from one article, paste it into a spreadsheet, and then manually open another site just to check the metadata or see if there were links pointing elsewhere. It was tedious clicking and constant context switching.

Now, you tell your agent which URL to read. The MCP handles all that legwork—it strips away the ads, the sticky footers, and the navigation bars. You get clean Markdown text immediately, letting your agent work on pure, actionable data without the junk.

Web Scraper MCP delivers real-time web data directly to your workflow.

The biggest time sink used to be manually gathering multiple sources for comparison. You'd have to visit URL A, then copy the summary; open URL B, and repeat the process. This made synthesizing complex reports slow and error-prone.

With this MCP, you simply tell your agent to use `batch_read`. It pulls both sites simultaneously. Your agent compares the two sources—the 'React' article versus the 'HTML' article—and delivers a single, synthesized answer. The data is always current.

Web Scraper with 5 Tools

These tools let you pull content, links, structured data, and maps of entire websites using your AI client.

#	TOOL	DESCRIPTION
01	`read`	Pulls any public webpage into clean Markdown format, stripping away ads and clutter for readable content.
02	`extract`	Gathers structured metadata from a page, pulling out the title, description, OG tags, and link count.
03	`list_links`	Pulls every single outbound hyperlink found across an entire web page's source code.
04	`batch_read`	Fetches and processes content from up to ten different URLs simultaneously for comparison or summary.
05	`crawl`	Automatically crawls a website starting at a given URL, mapping out the content of subsequent pages (up to 10).

See It in Action

Real prompts you can use once this MCP is connected to your AI agent through Vinkius Cloud.

U Read https://en.wikipedia.org/wiki/Artificial_intelligence and summarize its history.

I've fetched the Wikipedia page. The history of AI spans back to antiquity with myths of artificial beings, but the formal field was founded in 1956 at Dartmouth College. It experienced cycles of immense optimism followed by disappointment ('AI winters'), eventually leading to the modern deep learning revolution fueled by huge datasets and compute power.

U Extract the links from https://news.ycombinator.com/

I've extracted the outbound links. The site currently links out to 30 primary article sources including domains like github.com, weired.com, and nytimes.com, along with many internal navigational links to user profiles and comment threads.

U Compare these two links: url1.com and url2.com

Using the batch reading tool, I've loaded both URLs simultaneously. URL 1 discusses a 'React-first' architecture and uses component styling. URL 2 advocates for 'HTML-first', server-rendered patterns. While both aim to increase web performance, they take fundamentally opposite approaches to client-side hydration.

Frequently Asked Questions

01 **How does the Web Scraper MCP handle complex documentation sites?**

It uses the crawl tool to map out entire documentation hubs. You give it the starting URL, and your agent automatically navigates up to ten related pages so you don't miss any linked content.

02 **Can I compare articles from different websites at once?**

Yes, use `batch_read`. This tool fetches multiple URLs in parallel, allowing your agent to process and compare the content of up to ten sources simultaneously. It's ideal for comparative analysis.

03 **Do I need any special keys or authentication to use Web Scraper?**

No. You don't need API keys or any specific credentials. Once you subscribe to this MCP, you just paste the link into your chat and tell your agent what task it needs to perform.

04 **Is the content from the Web Scraper always clean?**

Yes. The primary reading tool converts messy webpages using Mozilla Readability logic, which strips out boilerplate code, ads, and navigation bars so you only get pristine text.

05 **How do I find all links on a page?**

Use the `list_links` tool. It systematically pulls every single outbound hyperlink from the web page without needing to download or parse the full body content, giving you just the list.

Go Live in 60 Seconds

Get your connection token from **cloud.vinkius.com**, then paste the endpoint URL into any MCP-compatible client.

YOUR MCP ENDPOINT

```
https://edge.vinkius.com/[TOKEN]/mcp
```

CLIENT	WHERE TO CONFIGURE
Claude AI	Profile → Customize → Connectors → "+" → Add custom connector → Paste endpoint
Cursor	Settings → Features → MCP Servers → "+ Add New MCP Server" → Type: SSE → Paste endpoint
VS Code	Ctrl/Cmd+Shift+P → "MCP: Add Server" → add `"web-scraper": { "url": "…" }`
Windsurf	MCP Settings → `mcp_settings.json` → Add endpoint URL
ChatGPT	Settings → Tools & plugins → Add MCP server → Paste endpoint
Gemini	Extensions → Add MCP Server → Paste endpoint URL

ASK AN AI ABOUT THIS

Let your preferred AI explain this MCP server

- Ask ChatGPT
- Ask Claude
- Ask Perplexity
- Ask Gemini
- Ask Grok

READY TO CONNECT

Web Scraper is live on Vinkius Cloud.

Get your connection token, paste it into your AI agent, and start building. No SDK. No deployment. Just results.

Start at cloud.vinkius.com →

vinkius.com · support@vinkius.com

INDEPENDENT PLATFORM DISCLAIMER

Vinkius is an independent platform and is not affiliated with, endorsed by, sponsored by, verified by, or otherwise authorized by Web Scraper. All third-party trademarks, logos, and brand names are the property of their respective owners. Their use in this document is strictly for informational purposes to identify service compatibility and interoperability.

DOCUMENT INFORMATION

Generated	June 2026
MCP Server	Web Scraper MCP
Server ID	`019d7604-5402-7173-a44b-24cf5d07da40`
Platform	Vinkius Cloud for AI Agents
Endpoint	`https://edge.vinkius.com/{token}/mcp`

LICENSE & USAGE

This document is generated automatically by the Vinkius PDF Engine. Content reflects the MCP server configuration at the time of generation and may change as updates are deployed. For the most current information, visit vinkius.com/mcp/web-scraper.