MCP SERVER | NO CODE | CLOUD HOSTED

HTML to Text Extractor MCP

Strip junk code and get pure text context.

HTML to Text Extractor strips messy web content down to clean, readable plain text. When your agent reads emails or scraped webpages, it often gets bogged down by inline CSS, broken tables, and redundant tags. This MCP instantly removes all that noise, letting you pass only the pure, structural text to your AI client. It saves massive amounts of token context while preserving list structure and essential formatting.

A+ Quality Score 100/100

text-extraction | html-parsing | token-optimization | data-cleaning | web-scraping

MAINTAINED BY VINKIUS CLOUD FOR AI AGENTS · JUNE 2026
vinkius.com/mcp/html-to-text-extractor

SCAN TO VIEW

The infrastructure that powers AI agents in the real world.

Vinkius connects AI to the world's software through secure, enterprise-grade infrastructure — enabling real-world execution at scale, built on the Model Context Protocol (MCP).

Your AI Connections Run Through Vinkius Cloud

The world's largest managed MCP catalog

Vinkius is the cloud infrastructure where AI agents connect 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.

HTML to Text Extractor MCP

1 tools available
Cloud-hosted on Vinkius

Ever noticed how much junk data comes with an email or a scraped article? When an agent pulls content from sources like Zendesk or Gmail, it usually gets dumped into a large chunk of raw HTML—a mess full of CSS code and unused tags. Forcing your AI client to read this garbage burns tokens fast and often confuses the model about what's actually important.

This MCP fixes that problem right away. It converts complex web markup into clean plain text instantly, preserving list layouts and link structure while eliminating all the junk. Think of it as a universal filter for dirty data. You feed it raw HTML, and you get back only the human-readable content. Connecting to this MCP via Vinkius gives your agent an immediate way to cleanse information before any processing happens, making subsequent steps much more reliable.

Core Capabilities

01 — Cleanse Raw Web Content

Takes raw HTML input and strips out all markup, leaving only clean, usable plain text.

02 — Reduce Token Overhead

Saves context window space by eliminating extraneous CSS and scripting tags from large documents.

03 — Maintain Document Structure

Preserves the original spatial layout, including bullet points and section breaks, so the AI client still understands the document's flow.

One Click on Vinkius — From Prompt to Execution

Available at **vinkius.com/mcp/html-to-text-extractor** — connect your AI agent in three steps.

01 Pass the messy HTML content (like a raw email dump or web page snippet) into the MCP.

02 The tool analyzes the markup, stripping away all CSS, tags, and scripts while keeping the core text readable.

03 Receive a clean plain-text string that your AI client can use for accurate context processing.

The bottom line is you get pure data without the digital noise.

Built For

Content operations teams, support engineers, and data analysts who spend their day reading web content or handling customer service tickets. If your work involves taking information from an external source into an automated workflow, this MCP is critical.

Support Engineer

Needs to automatically pull clean text summaries from complex email threads (like Zendesk) before feeding them into a knowledge base.

Data Analyst

Scrapes web pages for reports, needing reliable plain text that ignores page-specific styling and scripts.

Content Manager

Handles large volumes of online content, requiring a way to strip out all HTML remnants so the final output is purely editorial copy.

What Changes When You Connect

01 Saves tokens. Instead of feeding your agent 3MB of raw HTML, you pass only the necessary information, saving up to 95% of your context window space.

02 Handles dirty data. It reliably cleans content from sources like email APIs or web scrapers that dump messy markup into a single string.

03 Keeps structure. The resulting plain text preserves layout elements—like bullet points and section breaks—so the AI client understands the document's original flow.

04 Reduces confusion. By removing confusing CSS, scripts, and redundant tags, your agent spends less time parsing junk and more time generating accurate results.

05 Works across sources. Use this to process content from any web-based source that delivers HTML markup.

Real-World Applications

Summarizing a long customer support ticket

A support engineer pulls a multi-reply email thread containing messy HTML and tables. Instead of feeding the entire raw string to their agent, they use this MCP's `extract_text` tool first. The agent then summarizes only the clean plain text, ignoring all the junk code.

Analyzing a complex webpage for research

A data analyst scrapes an article from a website that uses heavy styling and scripts. They pipe the raw HTML through this MCP to strip out the noise. The agent then processes the clean text to identify key themes, ignoring all the visual clutter.

Cleaning up bulk email imports

A content manager gets a CSV of emails that were exported with full HTML markup. They run the `extract_text` tool on each field before uploading them to the workflow. The agent can then reliably search and categorize the clean, text-only messages.

Building an automated research pipeline

A developer builds a system that pulls data from multiple external APIs. By running this MCP first, they ensure every piece of raw HTML data is normalized into pure plain text before it hits the final AI processing step.

Patterns to Avoid

Treating raw HTML as clean input

✗ AVOID

Sending a massive string containing inline CSS and broken tables directly to the agent, hoping it can figure out what matters.

✓ INSTEAD

Always run the content through this MCP first. Use `extract_text` to convert the messy markup into pure text before your agent sees it. This prevents token waste and improves accuracy.

Relying on LLMs to strip tags

✗ AVOID

Prompting the AI client: 'Please summarize this HTML block, ignoring all tags.' The AI spends tokens trying to interpret the code instead of summarizing.

✓ INSTEAD

Don't ask the agent to clean the data. Use `extract_text` to do the cleaning work mechanically and feed it only the stripped text.

Mixing structured data types

✗ AVOID

Trying to pass a mix of HTML, JSON, and raw text into one prompt without pre-processing.

✓ INSTEAD

Use this MCP on all web content sources. This normalizes the input format, ensuring only clean plain text enters your primary workflow.

The Right Fit

Use this MCP if your data source delivers HTML markup and you need to pass pure, readable context to an agent or workflow. It is essential for any task involving web scraping, email parsing, or documentation review where the raw input is messy. Don't use it if your starting point is already clean text (like a database record). Also, don't rely on this MCP to *structure* data; it only extracts plain text. If you need structured output like JSON or XML, you'll need a different tool after using `extract_text`.

This MCP is purely about cleaning the input stream. It doesn't summarize, categorize, or analyze; it just removes the digital clutter so your agent can do that work accurately and efficiently.

The headache of messy web content

Today, if you pull data from an external source—say, a customer service ticket or a website report—you often get more than just the words. You get tables coded in HTML, inline styling for every paragraph, and tons of CSS code that has nothing to do with the message itself. Manually copying this stuff is tedious; running it through your agent without cleaning it burns thousands of tokens on useless markup.

With this MCP, you don't waste time wrestling with code. You feed the raw HTML string in, and it instantly strips out every single tag and style definition. What you get back is clean plain text that maintains the original flow, letting your AI client focus only on meaning.

Extract Text with `extract_text`

Manual cleanup involves opening developer tools to isolate content or writing complex regex rules just to get rid of the tags. This is fragile and doesn't account for every possible HTML variation.

This MCP handles all that automatically. It’s a reliable, single step that guarantees clean context. Your agent gets pure data, period.

HTML to Text Extractor with 1 Tool

This single tool lets you convert complex, messy HTML markup into pure, readable plain text context.

#	TOOL	DESCRIPTION
01	`extract_text`	Converts raw HTML into clean plain text instantly by stripping away all markup, significantly reducing token usage for agents processing heavy web pages or emails.

See It in Action

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

U Extract the text from this messy HTML email before I summarize it.

Extracted Text: Returned clean plain text successfully.

U Convert this raw HTML page snippet into plain text.

Extracted Text: HTML tags removed, layout preserved.

U Strip all the tables and CSS from this HTML string.

Extracted Text: Stripped output generated.

Frequently Asked Questions

01 **What types of files can the HTML to Text Extractor use?**

It accepts any raw text containing HTML markup, like content dumped from APIs, scraped web snippets, or full email source code. It doesn't care where the data came from, only that it needs cleaning.

02 **Does extract_text save my tokens?**

Yes. By eliminating unnecessary CSS and tags, you drastically reduce the size of the input context window, saving your agent a huge amount of computational cost.

03 **Can I use this MCP to summarize text?**

No. This MCP only extracts plain text; it doesn't perform any summarization or analysis. You must run the content through `extract_text` first, and then pass that clean output to a separate agent for summarizing.

04 **What if my HTML has tables?**

The tool preserves the spatial layout, meaning it keeps structural elements like lists and table divisions intact in the plain text, making them easier for your agent to parse contextually.

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 `"html-to-text-extractor": { "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

HTML to Text Extractor 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 HTML to Text Extractor. 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	HTML to Text Extractor MCP
Server ID	`019e38a9-2de6-70b4-b15f-83cae00991b9`
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/html-to-text-extractor.