MCP SERVER | NO CODE | CLOUD HOSTED

Long-Tail Extractor MCP for AI Agents

Discover High-Potential Keywords and Optimize Content Strategy

Long-Tail Extractor identifies recurring word sequences (n-grams) in large texts. It's an analysis engine that helps you find high-potential long-tail keyword candidates for content strategy and SEO by scanning documents for specific 3, 4, and 5-word phrases.

A+ Quality Score 100/100

n-grams | keywords | text-analysis | seo-tools | pattern-recognition | data-mining

MAINTAINED BY VINKIUS CLOUD FOR AI AGENTS · JULY 2026
vinkius.com/mcp/long-tail-extractor

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.

Long-Tail Extractor MCP

3 tools available
Cloud-hosted on Vinkius

This MCP gives your AI client the ability to perform deep textual analysis, uncovering patterns in large bodies of writing. Instead of manually sifting through pages of text, you can feed the material into your agent, and it will automatically locate recurring word sequences—specifically those that are 3, 4, or 5 words long. This process is crucial for SEO, helping users pinpoint high-potential long-tail keyword phrases they might otherwise miss. Once the patterns are found, your agent can use Vinkius to apply a frequency threshold filter, isolating only the most significant candidates. You can also calculate how common these specific sequences are relative to the document's total word count. This capability moves text analysis from tedious manual labor to precise data output.

Core Capabilities

01 — Identify Recurring Phrases

The MCP scans a document and returns lists of recurring word phrases along with how many times they appeared and where in the text they were found.

02 — Filter by Frequency Threshold

It filters out common noise, leaving only patterns that appear above a minimum frequency count you define.

03 — Measure Pattern Density

The MCP calculates the specific density score of keyword sequences relative to the entire word count of the source material.

One Click on Vinkius — From Prompt to Execution

Available at **vinkius.com/mcp/long-tail-extractor** — connect your AI agent in three steps.

01 You provide your AI client with a large text document or article that needs analysis.

02 Your agent runs the text through the MCP's pattern recognition engine, which identifies all potential 3, 4, and 5-word n-grams.

03 Finally, you instruct your agent to run filtering tools against the results, allowing you to pinpoint high-density, recurring keyword phrases.

The bottom line is that it automates a complex process of manual text analysis, delivering precise data on phrase frequency and location instantly.

Built For

Content strategists who are tired of guessing what keywords will hit big. Copywriters needing to write for specific search intent. SEO analysts drowning in raw text files that need pattern recognition. This MCP gives you the data backbone for content decisions.

SEO Analyst

Uses this MCP to process competitor websites or massive client reports, identifying hidden long-tail phrases they are ranking for.

Content Strategist

Feeds the MCP large swaths of research data to ensure every piece of content targets naturally occurring keyword clusters.

Technical Writer

Runs this on technical manuals or white papers to find repetitive terminology that should be standardized across all documentation.

What Changes When You Connect

01 Pinpoint exactly which 3, 4, or 5-word phrases are repeating in your text. Using `extract_ngram_sequences` gives you the location and count for every candidate.

02 Stop wasting time on low-value keywords. The MCP lets you use `filter_high_frequency_patterns` to isolate only the most impactful phrases based on frequency.

03 Gain a metric understanding of keyword saturation. By running `calculate_pattern_density`, you know precisely how common a phrase is relative to the whole document, improving your SEO targeting.

04 Automate manual text review. Instead of reading documents page by page, your agent processes massive datasets and delivers clean data on keyword patterns.

05 Improve content depth and relevance. You can feed this MCP into your workflow to ensure every piece addresses naturally occurring user search language.

Real-World Applications

Analyzing Competitor Content

An SEO analyst needs to know what phrases competitors are repeating in their top-ranking articles. They feed the text into your agent and use `extract_ngram_sequences` to pull out all recurring 3, 4, and 5-word sequences they should target.

Structuring a New Product Guide

A technical writer is building a user manual. They run the text through the MCP, using `filter_high_frequency_patterns` to identify key terms that must be highlighted and standardized throughout all chapters.

Validating Keyword Research Data

A content strategist has compiled a massive research report. They use your agent with the MCP to run `calculate_pattern_density`, proving which phrases are truly pervasive across the data, not just mentioned once or twice.

Mining Internal Corporate Documents

A compliance officer needs to ensure that a specific phrase is consistently used in all internal policy documents. They run `extract_ngram_sequences` against the whole corpus and check for consistency.

Patterns to Avoid

Treating keywords as single words

✗ AVOID

Simply searching a document for 'quick brown fox' without checking if it appears frequently enough or where exactly it is located.

✓ INSTEAD

Don't just search. Use the MCP to first run `extract_ngram_sequences` to find all 3-word candidates, then use `filter_high_frequency_patterns` to narrow down only those that appear more than twice.

Ignoring context or word order

✗ AVOID

Assuming a keyword is good just because it appears in the document. You might miss variations of sequence or misidentify importance.

✓ INSTEAD

The MCP fixes this by looking at contiguous blocks of words, giving you precise data on both frequency and relative pattern density using `calculate_pattern_density`.

Overloading the AI with raw text

✗ AVOID

Giving your agent a 10,000-word article and asking it to 'find keywords' without guiding it on how.

✓ INSTEAD

Tell your agent to use `extract_ngram_sequences` first. This limits the output to only specific n-gram candidates you care about (3, 4, or 5 words).

The Right Fit

Use this MCP if your goal is finding repeating, multi-word phrases in large text bodies for SEO research, content validation, or data mining. This tool excels when simple keyword counting isn't enough; you need to know the sequence and frequency of specific word groups (3-5 words). Don't use it if you only want a list of single keywords—a basic vocabulary lookup is better for that. Also, don't rely on this MCP alone for content *writing*; it provides data, not finished articles. You still need human strategy to guide your agent in using the discovered patterns.

Long-Tail Extractor MCP: Discovering Keyword Patterns in SEO Research

Right now, finding effective long-tail keywords is a massive manual drag. You copy huge blocks of text—like competitor articles or market reports—and spend hours manually highlighting phrases and cross-referencing them to see if they repeat often enough to be worth writing about.

With this MCP, your agent handles the heavy lifting. It identifies all recurring 3-, 4-, and 5-word sequences automatically, giving you a clean list of candidates so you know exactly which phrases are repeating across massive datasets.

Using Long-Tail Extractor for Precise Content Density Analysis

The biggest time sink is filtering the noise. You find a good phrase, but then you have to estimate how often it appears relative to the whole document's word count, which is guesswork at best.

Now you get objective data. The MCP allows your agent to calculate pattern density, giving you a specific score that proves exactly how pervasive and relevant those phrases are across the entire source text.

Long-Tail Extractor: 3 Tools for Keyword Pattern Recognition

Use these tools to analyze text patterns, filter keyword candidates by frequency, and calculate the density of long-tail phrases in your documents.

#	TOOL	DESCRIPTION
01	`extract_ngram_sequences`	Identifies and returns recurring word phrases from a text, providing their total counts and specific positions within the document.
02	`filter_high_frequency_patterns`	Filters out low-value patterns, keeping only n-grams that meet or exceed a specified minimum frequency threshold.
03	`calculate_pattern_density`	Calculates the precise density score of word sequences to measure how prevalent they are compared to the total word count.

See It in Action

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

U Find all recurring 3-word sequences in this article about gardening equipment.

Found 2 recurring patterns:

- 'pruning shears guide' (3 occurrences)
- 'soil drainage system' (1 occurrence)

The sequence 'pruning shears guide' is the most consistent candidate.

U Show me only phrases that appear at least 4 times in this list of n-grams.

The following patterns met the minimum threshold of 4 occurrences:

PHRASE	COUNT
'sustainable farming practice'	5
'local food sourcing guide'	4

Use these phrases to build your next content cluster.

U Calculate the density for 'natural light window' in a 500-word text where it appeared 7 times.

Density Score: 0.014.

This score means that for every 100 words, your phrase appears about 1.4 times. This is a strong indicator of high topical relevance.

Frequently Asked Questions

01 **How does the Long-Tail Extractor MCP help me find better keywords?**

It finds long-tail keywords by analyzing common word sequences (3, 4, and 5 words) that people are actually writing about. This gives you a data-backed list of phrases that have high search potential.

02 **Is the Long-Tail Extractor MCP just for SEO, or can I use it elsewhere?**

You can use it anywhere there's text. Besides SEO, it’s great for legal compliance, technical documentation, or academic research where you need to find repeating terminology across huge files.

03 **Does the Long-Tail Extractor MCP only look at 3-word phrases?**

No. The MCP is designed to scan and identify sequences of words that are 3, 4, or 5 words long, giving you comprehensive coverage for keyword candidates.

04 **What if I don’t know what frequency threshold I need?**

You can experiment with the MCP’s filtering tools. Start by setting a low threshold to see all patterns, then raise it until your agent only presents the most consistent and valuable phrases.

05 **Will Long-Tail Extractor tell me if a phrase is actually high volume?**

It doesn't provide search volume numbers. However, by calculating pattern density, it tells you how *prevalent* the phrase is within your specific body of text, which is a key metric for content planning.

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 `"long-tail-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

Long-Tail 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 Long-Tail 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	July 2026
MCP Server	Long-Tail Extractor MCP
Server ID	`019f13f2-999a-729c-9f12-2a7922928ff9`
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/long-tail-extractor.