MCP SERVER | NO CODE | CLOUD HOSTED

Internet Archive Search MCP

Deeply filter 40M+ historical documents and media.

Internet Archive Search lets your agent perform advanced research across the world's largest digital library. You can query over 40 million items—everything from books and films to music, software, and historical documents. Filter content by specific decades, media type, creator, or topic using complex queries (AND, OR, NOT). It's built for deep, focused archival discovery.

A+ Quality Score 98.33/100

digital-library | search-engine | archival-data | information-retrieval | content-discovery | open-access

MAINTAINED BY VINKIUS CLOUD FOR AI AGENTS · JUNE 2026
vinkius.com/mcp/internet-archive-search

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.

Internet Archive Search MCP

12 tools available
Cloud-hosted on Vinkius

Think of this MCP as a massive research assistant that indexes everything from the last century. Instead of just searching keywords, you tell it what kind of content you need and when. Your agent can handle complicated queries using operators like AND or NOT to narrow results down instantly. Need to find all articles about civil rights published in the 1960s? Or maybe only pre-war films shot on film stock? This MCP handles that complexity across texts, videos, and audio recordings alike.

When you connect this Internet Archive Search through Vinkius, your agent gets access to a highly structured workflow. You can refine results by publisher or narrow the search down just to specific collections like NASA's records. It's about precision discovery; it helps you bypass the noise of general web searches and go straight for primary source material.

Core Capabilities

01 — Filter Search Results

Limit results by format, such as texts, movies, or software, to focus your research.

02 — Analyze Content Composition

Determine what types of content are present in a set of search results using JSON faceting syntax.

03 — Isolate Specific Creators

Find all works associated with an author, organization, or notable person.

04 — Target Historical Periods

Restrict searches to content created within a specific start and end year range.

05 — Perform Deep Text Searches

Run full-text queries across item descriptions and metadata for highly specific terms.

06 — Identify by Subject Matter

Search content using curated, assigned topics like 'world war 2' or 'jazz music'.

One Click on Vinkius — From Prompt to Execution

Available at **vinkius.com/mcp/internet-archive-search** — connect your AI agent in three steps.

01 You provide your agent with a complex research query and specify required filters, such as the date range, media type, or creator.

02 The MCP executes this multi-faceted search across the Internet Archive's 40 million+ items.

03 Your agent receives structured data that pinpoints relevant results based on all the applied criteria.

The bottom line is you get precise answers to highly specific historical or academic questions, without needing to manually click through dozens of search result pages.

Built For

Academics and researchers who need primary source data. Think historians sifting through decades of raw material, journalists verifying obscure facts, or students doing deep literature reviews across multiple formats.

Academic Researcher

Uses the MCP to combine date range filtering with subject search to build a bibliography on niche historical topics.

Historical Journalist

Runs multiple searches by creator or publisher to trace the evolution of a figure's ideas over time, cross-referencing texts and images.

Archivist/Curator

Applies media type filtering and faceted search tools to analyze large collections and determine content gaps or popular material.

What Changes When You Connect

01 Pinpoint content by exact era: Use `search_by_date_range` to pull all material from a specific decade, eliminating modern noise.

02 Analyze vast collections with precision. Run `faceted_search` to understand how results are distributed across different formats or topics automatically.

03 Trace the work of individuals using `search_by_creator`. Find every article or book by an author without having to search their name manually dozens of times.

04 Focus your media type: Use `search_by_mediatype` to pull only films, leaving out millions of irrelevant documents. Or vice versa.

05 Stay current on research topics using `search_recent`. See what has been added to the archive since you last ran a query.

Real-World Applications

Tracking Early Cinema History

A film student needs to find all silent films from the 1920s directed by German masters. They ask their agent, specifying `search_by_mediatype` (film) and using `search_by_date_range` (1920-1929), guaranteeing they don't miss any key works.

Verifying Corporate History

A journalist needs to prove a company changed its name and branding. They use `search_by_publisher` combined with `search_fulltext` to find all mentions of the old name in their annual reports, year by year.

Researching Global Food Trends

A global food researcher wants to see how rice farming was discussed in different languages. They use `search_by_subject` (agriculture) and combine it with `search_by_language` for Mandarin, Spanish, and English.

Identifying Rare Software Manuals

A tech historian needs to find documentation for obsolete operating systems. They run a search using `search_by_mediatype` (software) and then filter the results by known manufacturers via `search_by_publisher`.

Patterns to Avoid

General keyword searching

✗ AVOID

Asking your agent simply to 'find articles about World War II.' This returns millions of results spanning all times and formats.

✓ INSTEAD

You need to combine tools. Use `search_by_subject` (world war 2) and pair it with `search_by_date_range` (1939-1945) for maximum focus.

Forgetting the format

✗ AVOID

Searching for 'NASA images' but getting a mix of texts, audio recordings, and films. You waste time manually sifting through junk.

✓ INSTEAD

Always use `search_by_mediatype` (image) to narrow results immediately after your initial search.

Searching too broadly

✗ AVOID

Using only the general `search` tool without limiting fields or dates. The result set is massive and overwhelming.

✓ INSTEAD

Use specific tools first. Start with `search_by_collection` (NASA) then refine the scope using a limited search like `search_by_mediatype`.

The Right Fit

Use this MCP if your research requires deep, structured filtering across multiple decades and content formats. If you need to find everything by 'George Orwell' *and* it must be related to 'dystopia,' you combine `search_by_creator` with `search_by_subject`. Don't use it if you are just looking for a single piece of general information; then, a simple web search is fine. But if your goal is academic discovery—finding the *breadth* and *depth* of historical content—this MCP is required. Never rely on basic keyword searches alone; always layer in `search_by_date_range` or `faceted_search` to get usable data.

Dealing with the digital dust pile of human history

Today, deep research feels like wading through a giant dumpster fire. You start by typing keywords into Google Scholar or academic databases, and you're immediately hit with thousands of links. Then you open them up; some are paywalled, some are outdated, and most require you to manually check the publication date, author, and format before you even know if they matter.

With this MCP, that manual labor vanishes. You tell your agent exactly what you're looking for—say, films from the 1940s about civil rights—and it filters everything down in one go. You get a clean list of actionable results across text, film, and audio.

Internet Archive Search: Precise Discovery

You no longer have to jump between publisher websites or manually check the year on every single search result. The MCP handles the metadata checks for you, applying filters like `search_by_publisher` and `search_by_date_range` instantly.

What changes is that your agent gives you actionable intelligence instead of just a link dump. You get curated results ready for synthesis.

Internet Archive Search: 12 Discovery Tools

These twelve specialized tools allow your agent to perform highly targeted searches, filters, and analyses across the entire digital archive.

#	TOOL	DESCRIPTION
01	faceted_search	Analyzes the composition of a result set by breaking down categories like media type, collection, or creator.
02	search_by_collection	Limits results to content housed within specific themed archives or community collections.
03	search_by_creator	Finds all available works from a designated author, organization, or notable figure.
04	search_by_date_range	Narrows down results to only include items published within a defined start and end year range.
05	search_fulltext	Performs a broad search across all 40 million items, supporting complex queries and wildcards.
06	search_by_language	Retrieves content that is published specifically in a requested language, such as French or Spanish.
07	search_by_mediatype	Filters the search to only show items of one specific format type, like audio or film.
08	search_by_publisher	Identifies all content that originated from a particular publishing house.
09	search_recent	Retrieves the most recently uploaded materials to see what new items have been added to the archive.
10	search_by_subject	Searches for content using curated, general topics like 'science fiction' or 'civil rights'.
11	search	Supports AND, OR, NOT, wildcards (*), and field searches. Use this for broad discovery. Optional: fields (e.g., "identifier,title,mediatype"), rows (1-100), page for pagination, and sort (e.g., "date desc"). Universal search across 40M+ items in the Internet Archive
12	search_top_downloads	Finds the most popular or frequently downloaded content within specific formats like texts or movies.

See It in Action

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

U Search for public domain films from the 1940s.

Found 12,847 films from the 1940s in Prelinger Archives and community collections, including WWII propaganda, educational shorts, and home movies.

U Show me the most downloaded items.

Top downloads include: Big Buck Bunny (movie), various Project Gutenberg ebooks, NASA Apollo mission photos, and classic software from the softwarelibrary collection.

U Search for NASA images.

Found 185,000+ NASA items including Apollo mission photographs, Hubble Space Telescope images, satellite imagery, and Space Shuttle documentation.

Frequently Asked Questions

01 **How do I find content from specific years using the Internet Archive Search MCP?**

You use the `search_by_date_range` tool by providing a start year and an end year, alongside your main search query. This limits results to only that time period.

02 **Can I filter by media type using Internet Archive Search?**

Yes. You use `search_by_mediatype` to restrict your search to one format, like movies or audio recordings, making the result set much smaller and more targeted.

03 **What is the best way to find all work by a specific person?**

Use `search_by_creator`. This tool gathers every item associated with that author or organization name across the entire archive, regardless of date or format.

04 **Is Internet Archive Search good for finding rare software manuals?**

Absolutely. Use `search_by_mediatype` and then refine by keywords in the title using the general `search` tool to locate old digital artifacts.

05 **How do I find content about a topic without knowing the exact year?**

Start with `search_by_subject`. This uses curated topics like 'world war 2,' allowing you to gather all related materials across different time periods and formats.

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 `"internet-archive-search": { "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

Internet Archive Search 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

DOCUMENT INFORMATION

Generated	June 2026
MCP Server	Internet Archive Search MCP
Server ID	`019d75b6-5bc5-700f-ae62-a3a74eeda8c1`
Platform	Vinkius Cloud for AI Agents
Endpoint	`https://edge.vinkius.com/{token}/mcp`

LICENSE & USAGE