

DeepDive

News ▢

Graphify: Open-Source-Tool spart bis zu 4x Tokens bei KI-Workflows

KI-gestützte Videozusammenfassung

Graphify: Open-Source-Tool spart bis zu 4x Tokens bei KI-Workflows

Die 5 Ws

Wer

Benjamín Cordero und die Community um Imperio Agéntico

Was

Vorstellung von Graphify, einem Open-Source-Tool, das durch die Erstellung eines Wissensgraphen aus Dateien den Token-Verbrauch bei KI-Assistenten wie Claude Code drastisch reduziert

Wann

Aktuell (Video nach dem Start von Cloudfamous 5.5 veröffentlicht)

Wo

YouTube-Kanal von Benjamín Cordero und GitHub

Warum

Weil KI-Modelle bei großen Dateimengen viele Tokens für wiederholtes Einlesen verschwenden; Graphify schafft einen navigierbaren Graphen, der nur relevante Abschnitte lädt

Perspectives

Pro

Graphify spart erheblich Tokens und Zeit, da KI-Assistenten nur noch den Graphen durchsuchen müssen. Die lokale Verarbeitung von Code und Audio spart zusätzlich API-Kosten. Die Integration in bestehende Tools wie Claude Code oder Obsidian ist einfach und das Tool ist komplett Open Source.

Contra

Der initiale Indexierungsprozess kann bei großen Datenmengen 20–30 Minuten dauern und verbraucht selbst Tokens. Zudem muss der Graph bei häufigen Änderungen regelmäßig aktualisiert werden, was zusätzliche Kosten verursacht. Für reine Codeprojekte reicht oft ein einfacher grep-Befehl, sodass der Mehrwert geringer sein kann.

Significance

Graphify adressiert ein zentrales Effizienzproblem bei der Arbeit mit KI-Assistenten: den übermäßigen Token-Verbrauch durch wiederholtes Einlesen großer Dateimengen. Das Tool reduziert Token-Kosten nachweislich um den Faktor 4 (etwa 70% Einsparung), wie der Entwickler in einem Live-Test demonstriert. Es kombiniert Code-Analyse, Sprachtranskription und KI-gestützte Indexierung zu einem durchsuchbaren Wissensgraphen. Die Open-Source-Verfügbarkeit und die Kompatibilität mit verschiedenen Agenten und Dateitypen machen es für Entwickler und Power-User attraktiv, die mit großen Datenbeständen arbeiten.

Key Quotes

„Graphify kann dir bis zu 4x weniger Tokens pro Frage sparen – das ist kein Marketing, sondern live gemessen.“

„Stell dir einen Wissensgraphen vor: Jeder Punkt ist eine Idee, jede Linie eine Verbindung – ohne dass ich etwas manuell verbinden musste.“

„Die Kostenersparnis von 70x, die im Internet kursiert, ist falsch. Der reale Faktor liegt bei etwa 4x, was immer noch enorm ist.“

„Graphify ist gratis, Open Source und funktioniert mit Claude Code, Codex, Cursor und vielen anderen Agenten.“

„Du erstellst einmal einen Index und aktualisierst nur die geänderten Dateien – das minimiert den laufenden Token-Verbrauch.“

Action Items

1. Installiere Graphify über den GitHub-Link (<https://github.com/your-repo>) und führe die Indexierung für dein Projekt aus.
2. Teste die Token-Ersparnis mit einer Beispielabfrage: Vergleiche Token-Verbrauch mit und ohne Graphify.
3. Tritt der Community „Imperio Agéntico“ bei, um weitere Optimierungen und den geplanten „Motor Agente“ zu erhalten.
4. Aktualisiere deinen Knowledge-Graphen regelmäßig mit dem Befehl ``graphify update``, um aktuelle Änderungen zu berücksichtigen.
5. Teile deine Erfahrungen in den Kommentaren oder auf GitHub, um das Tool weiterzuentwickeln.

Follow-up Questions

- Wie schneidet Graphify im Vergleich zu RAG-Systemen bei großen Dokumentenmengen ab?
- Welche Strategien gibt es, um die initiale Indexierung zu beschleunigen?
- Wie lässt sich Graphify in CI/CD-Pipelines integrieren, um den Graphen automatisch zu aktualisieren?
- Unterstützt Graphify auch andere Modelle als Claude, z.B. Gemini oder Llama?
- Wann wird der angekündigte „Motor Agente“ mit visueller Kostenanzeige veröffentlicht?