TBAY · WEB366

Thiết kế web trọn gói · Bộ công cụ Content/SEO cho Claude

TÀI LIỆU HƯỚNG DẪN

Cài đặt & sử dụng plugin Content Duplicate Audit

Kiểm tra bài viết trùng lặp, ăn thịt từ khóa từ sitemap XML và xuất báo cáo Excel kèm đề xuất xử lý

Video hướng dẫn

Xem từ phút 3:24 trên YouTube

Quét QR hoặc bấm vào thẻ →

Bài viết gốc

Xử lý nội dung trùng lặp trong WordPress

Quét QR hoặc bấm vào thẻ →

CÔNG TY TNHH DỊCH VỤ CÔNG NGHỆ TBAY
119A Diệp Minh Châu, Phường Tân Sơn Nhì, TP.HCM
Hotline: 0852.080383 · Email: hotro@tbay.vn · tbay.vn – website366.com

Phiên bản plugin 1.0.0 · 10/2026

1. Plugin này làm gì?

Viết blog lâu ngày, website rất dễ có nhiều bài **cùng nhắm một ý định tìm kiếm** và tự tranh top với nhau (gọi là *ăn thịt từ khóa* – keyword cannibalization). Plugin **content-duplicate-audit** giúp Claude đọc sitemap, tìm các nhóm bài trùng và xuất một file Excel để đội content xử lý, theo dõi tiến độ.

Nguyên tắc của plugin: **trùng lặp là trùng intent, không phải trùng câu chữ.** Hai tiêu đề khác chữ nhưng cùng trả lời một câu hỏi của người tìm thì vẫn là trùng; ngược lại, giống chữ nhưng khác đối tượng, ngành hay giai đoạn thì không trùng. Thuật toán chỉ lọc ứng viên, Claude là người soát lại và ra quyết định cuối cùng.

Đầu vào	Bắt buộc	Ghi chú
Sitemap XML	Có	File .xml / .xml.gz tải lên, hoặc URL. Nhận cả sitemap_index.xml của Yoast/Rank Math và nhiều file cùng lúc.
Export Search Console	Không	File CSV tab “Trang” (URL, Lượt nhấp, Lượt hiển thị). Có file này thì chọn bài chủ và đề xuất xóa chính xác hơn.
Phạm vi	Không	Mặc định lấy bài viết + trang + sản phẩm, bỏ category/tag/author/phân trang. Có thể giới hạn, ví dụ chỉ /blog/.

Đầu ra: file Excel `bao-cao-trung-lap-<ten-mien>-<ngày>.xlsx` gồm danh sách bài trùng (có link), trùng với bài nào, mức độ tương đồng, đề xuất xử lý cụ thể và danh sách chủ đề nên viết mới.

Nội dung trong gói plugin

```
content-duplicate-audit/
├── .claude-plugin/
│ ├── plugin.json # thông tin plugin
│ └── marketplace.json # đ cài qua /plugin marketplace
├── commands/
│ └── kiem-tra-trung-lap.md # lệnh /kiem-tra-trung-lap
├── skills/content-duplicate-audit/
│ ├── SKILL.md
│ └── scripts/audit.py
└── README.md
```

2. Chuẩn bị trước khi cài

- File **content-duplicate-audit-plugin.zip** (nhận từ TBAY / WEB366).
- Tài khoản Claude có dùng được Plugin/Skill: Claude Code, Claude Desktop (Cowork) hoặc claude.ai có bật Skills.
- Link sitemap của website, thường là `https://tenmien.vn/sitemap_index.xml` (Yoast, Rank Math) hoặc `/wp-sitemap.xml` (WordPress mặc định).
- Tùy chọn: file CSV xuất từ Google Search Console → Hiệu suất → tab **Trang** → Xuất → CSV.
- Môi trường chạy cần Python 3.9+ với `requests`, `beautifulsoup4`, `openpyxl`. Claude tự cài nếu thiếu, bạn không phải làm gì.

3. Cài đặt plugin

Chọn một trong ba cách dưới đây, tùy bạn đang dùng Claude ở đâu.

A Claude Desktop / Cowork – cách dễ nhất

- Mở Claude Desktop, vào phần quản lý **Plugin**.
- Chọn tải plugin lên từ file, chọn file **content-duplicate-audit-plugin.zip**.
- Bật plugin cho cuộc trò chuyện. Lệnh **/kiem-tra-trung-lap** sẽ xuất hiện khi gõ dấu "/".

B Claude Code (máy tính, dòng lệnh)

Giải nén file zip vào một thư mục, ví dụ `D:\claude-plugins\content-duplicate-audit`, rồi gõ trong Claude Code:

```
/plugin marketplace add D:\claude-plugins\content-duplicate-audit
/plugin install content-duplicate-audit@web366-content-tools
```

Khởi động lại Claude Code nếu được yêu cầu. Nếu thư mục plugin được đưa lên GitHub, người khác cài bằng:

```
/plugin marketplace add <user>/<repo>
/plugin install content-duplicate-audit@web366-content-tools
```

C claude.ai – chỉ dùng skill

- Giải nén file zip, nén riêng thư mục `skills/content-duplicate-audit/` thành một file .zip mới.
- Vào **Settings** trên claude.ai, tìm mục **Skills**, tải file .zip đó lên và bật skill.
- Cách này không có lệnh /kiem-tra-trung-lap; bạn gọi skill bằng câu nói tự nhiên (xem mục 4).

> **Mẹo**
> Claude Desktop/Cowork và Claude Code chạy được trên máy có mạng nên tải được tiêu đề, H1, số từ của từng bài – kết quả đầy đủ nhất. Trên claude.ai, mạng thường bị chặn nên nên tải trực tiếp các file sitemap con lên (xem mục 6).

4. Cách sử dụng

Cách 1 – Gọi lệnh

```
/kiem-tra-trung-lap https://tenmien.vn/sitemap_index.xml
/kiem-tra-trung-lap post-sitemap.xml gsc-trang.csv
```

Tham số thứ nhất là URL hoặc file sitemap; nếu đính kèm thêm file .csv, Claude coi đó là dữ liệu Search Console. Gõ lệnh không kèm gì, Claude sẽ tự tìm file sitemap bạn đã tải lên hoặc hỏi lại đúng một câu.

Cách 2 – Nói tự nhiên

Skill tự kích hoạt khi bạn gửi sitemap kèm những câu như:

- “Kiểm tra bài trùng lặp cho sitemap này giúp mình.”
- “Website mình có bài nào ăn thịt từ khóa không? Bài nào nên gộp hay xóa?”
- “Rà soát bài viết cũ của blog, chỉ xét thư mục /blog/.”

Claude làm gì sau khi nhận sitemap

Bước	Việc Claude thực hiện
1. Đọc sitemap	Lấy toàn bộ URL, tự đọc các sitemap con nếu là sitemap index, phân loại Bài viết / Trang / Sản phẩm.
2. Tải dữ liệu trang	Lấy title, H1, meta description, số từ, canonical, noindex, mã HTTP của từng URL.
3. Tìm ứng viên trùng	Bỏ dấu, bỏ từ dừng, bỏ năm, gom từ đồng nghĩa rồi so độ tương đồng tiêu đề + H1 + slug. Rất cao ≥ 75%, Cao ≥ 55%, Cần xem xét 40–55%.
4. Soát lại intent	Bước quan trọng nhất: Claude đọc từng nhóm, tách nhóm gộp sai, thêm cặp trùng nghĩa khác chữ, chọn bài chủ, ghi lý do và việc cần làm cụ thể, gợi ý 5–15 chủ đề viết mới.
5. Xuất Excel	Tạo file .xlsx 5 sheet, có công thức tổng hợp tự cập nhật.
6. Trả kết quả	Gửi file Excel kèm tóm tắt: số nhóm trùng, 3–5 nhóm nên xử lý trước và lý do.

> **Lưu ý thời gian**
> Website trên 1.000 URL có thể mất vài phút ở bước tải dữ liệu. Website trên 5.000 URL nên chạy theo từng sitemap con hoặc từng thư mục để Claude soát kỹ được.

5. Đọc và xử lý file Excel

Sheet	Nội dung
Tổng quan	Số URL, số nhóm trùng, tỷ lệ bài bị trùng, số lượng theo từng đề xuất (tự cập nhật khi sửa), chú thích màu, phương pháp.
Bài trùng lặp	Mỗi dòng một bài: Nhóm, Vai trò (Bài chủ/Bài phụ), URL, Trùng với URL, Độ tương đồng, Intent chung, **Đề xuất**, Lý do, Việc cần làm, Ưu tiên, Trạng thái (Chưa làm / Đang làm / Đã xong / Bỏ qua), Ghi chú.
Tất cả cặp tương đồng	Mọi cặp ≥ 40% và các cặp Claude thêm vào, để bạn kiểm chứng.
Gợi ý viết mới	Tiêu đề, từ khóa, góc khác biệt, lý do và link nên trỏ về.
Tất cả URL	Toàn bộ URL đã phân tích, cờ trùng lặp, mã HTTP, noindex.

Ý nghĩa các đề xuất

Đề xuất	Khi nào	Việc cần làm
Giữ & Cập nhật	Bài chủ của nhóm.	Bổ sung nội dung hay từ bài phụ, cập nhật năm, số liệu; nhận link nội bộ.
Gộp + Redirect 301	Trùng intent rõ, bài phụ có nội dung đáng giữ (lựa chọn an toàn mặc định).	Chuyển phần giá trị sang bài chủ, redirect 301 bài phụ về bài chủ.
Viết lại góc khác	Chồng lấn một phần, có thể tách theo đối tượng, ngành, giai đoạn, định dạng.	Đổi tiêu đề, từ khóa và nội dung theo góc mới ghi trong cột “Việc cần làm”.
Xóa + Redirect 301	Trùng intent, nội dung mỏng (<300 từ) hoặc 0 lượt nhấp.	Xóa bài và redirect 301 về bài chủ. Không bao giờ để lỗi 404.
Giữ nguyên (không trùng)	Thuật toán báo trùng nhưng xét kỹ là khác intent.	Không cần xử lý.

Sau khi xử lý
Luôn redirect 301 khi gộp/xóa (dùng Rank Math, Yoast Premium hoặc plugin Redirection), cập nhật link nội bộ đang trỏ vào bài cũ, rồi kiểm tra lại trong Google Search Console sau 4–8 tuần.

6. Lưu ý và xử lý sự cố

Tình huống	Cách xử lý
Claude báo không tải được sitemap / chạy "chế độ SLUG"	Môi trường đang chặn mạng (hay gặp trên claude.ai). Tải trực tiếp các sitemap con lên: post-sitemap.xml, page-sitemap.xml, product-sitemap.xml. Kết quả chế độ SLUG chỉ dựa vào URL nên kém chính xác; nên chạy lại trên Claude Desktop/Code.
Website đa ngôn ngữ (/en/, /vi/)	Bản dịch không phải trùng lặp. Chạy riêng từng ngôn ngữ, ví dụ "chỉ xét các URL chứa /vi/".
Sản phẩm biến thể (khác màu/size, URL riêng)	Không xử lý như bài viết; nên gom về một trang sản phẩm có biến thể hoặc dùng canonical.
Trang đã có canonical trỏ sang URL khác	Đã xử lý kỹ thuật, plugin ghi chú lại và không đề xuất gộp thêm.
Không có nhóm trùng nào	Vẫn có file Excel (Tất cả URL + Gợi ý viết mới) – website của bạn đang sạch.
Muốn kiểm tra bài bị copy nguyên văn	Plugin so sánh ở mức tiêu đề/H1/URL (intent), không so toàn văn. Dùng Siteliner hoặc Copyscape cho việc này.

7. Tài liệu tham khảo

Đọc bài viết gốc để hiểu cách xử lý nội dung trùng lặp trên WordPress, và xem video để thấy toàn bộ quá trình chạy skill từ sitemap đến file Excel.

Bài viết gốc
Xử lý nội dung bài viết trùng lặp trong website WordPress
https://website366.com/xu-ly-noi-dung-bai-viet-trung-lap-trong-website-wordpress/

Video hướng dẫn dùng skill
Xem phần demo bắt đầu từ phút 3:24
https://youtu.be/cOob2Eiieks?t=204

8. Cần hỗ trợ?

TBAY / WEB366 hỗ trợ cài plugin, chạy kiểm tra trùng lặp và xử lý gộp bài, redirect 301, cập nhật link nội bộ cho website WordPress của bạn.

CÔNG TY TNHH DỊCH VỤ CÔNG NGHỆ TBAY
119A Diệp Minh Châu, Phường Tân Sơn Nhì, TP.HCM
Hotline / Zalo: **0852.080383** · Email: **hotro@tbay.vn**
Website: tbay.vn – website366.com