**I. Hướng dẫn sử dụng chi tiết**

1. Giới thiệu về WordPress Content Crawler

WordPress Content Crawler là plugin WordPress hỗ trợ thu thập nội dung từ hầu hết các trang web trên Internet. Plugin này sử dụng các bộ chọn CSS để định vị và trích xuất nội dung từ mã nguồn của trang mục tiêu. Bộ chọn CSS dễ học và hiệu quả trong việc xác định nội dung cần lấy. Plugin đi kèm công cụ Visual Inspector, cho phép bạn nhấp chọn các phần tử trên trang để tự động tạo bộ chọn CSS. Công cụ này còn có khả năng nhận diện các phần tử tương tự, giúp lấy tất cả các URL trong một trang danh mục.

Plugin cung cấp hơn 200 tùy chọn cài đặt, giúp bạn thu thập và tùy chỉnh nội dung theo ý muốn. Một số tính năng nổi bật bao gồm:

* Thay đổi hoặc xóa các thành phần HTML
* Điều chỉnh thuộc tính của phần tử
* Tìm và thay thế nội dung
* Chuyển đổi giá trị giữa hai thuộc tính
* Gán các thành phần vào mã ngắn để sử dụng trong các mẫu bài viết

Plugin hoạt động dựa trên WP-Cron, một tính năng giúp WordPress lập lịch các tác vụ chạy ngầm vào thời điểm nhất định. WP-Cron được sử dụng để thu thập, cập nhật và xóa nội dung tự động, đảm bảo plugin hoạt động liên tục và hiệu quả.

**1.1. Các tính năng chính**

 Lưu chi tiết bài đăng: Plugin hỗ trợ lưu tiêu đề, trích đoạn, nội dung, thẻ, danh mục, slug, ngày đăng, meta tùy chỉnh, phân loại, từ khóa meta, mô tả meta, hình ảnh nổi bật, hình ảnh bài viết, trạng thái và nhiều thông tin khác.

 Thu thập bài viết tự động: Tự động tìm và thu thập bài viết từ các URL được chỉ định sau khi thiết lập.

 Xóa bài viết: Tự động xóa các bài viết cũ theo cài đặt.

 Tạo danh mục và phân loại: Tự động tạo danh mục và phân loại từ nội dung thu thập được hoặc nhập thủ công.

 Meta bài đăng tùy chỉnh: Lưu các giá trị meta theo bộ chọn CSS hoặc nhập trực tiếp.

 Bộ chọn thay thế: Hỗ trợ nhiều bộ chọn CSS để đảm bảo thu thập dữ liệu từ các trang có thiết kế khác nhau.

 Thu thập bài viết phân trang: Lưu nội dung từ các bài viết có nhiều trang.

 Loại bỏ nội dung không cần thiết: Loại bỏ quảng cáo, bình luận hoặc các phần tử dư thừa bằng bộ chọn CSS.

 Phân loại bài đăng: Lưu bài viết dưới dạng bài viết, trang, sản phẩm hoặc loại bài đăng tùy chỉnh.

 Bảo vệ mật khẩu: Thiết lập mật khẩu cho bài viết để giới hạn người xem.

 Kiểm tra bài viết trùng lặp: Xác minh trùng lặp qua URL, tiêu đề hoặc nội dung.

 Lưu hình ảnh thành thư viện: Lưu hình ảnh từ bài viết dưới dạng thư viện hoặc thư viện WooCommerce.

 Proxy: Hỗ trợ proxy hoặc nhiều proxy để vượt qua giới hạn IP.

 Thu thập số lượng lớn: Thiết lập giới hạn số bài viết thu thập trong mỗi phiên.

 JSON Parsing: Thu thập dữ liệu từ JSON bằng bộ chọn CSS.

 Dịch tự động: Tích hợp các dịch vụ dịch thuật API như Google, Microsoft và Yandex (một số dịch vụ có tính phí).

 Quay tự động: Viết lại nội dung bài viết để tối ưu SEO thông qua các dịch vụ Spin Rewriter hoặc Türkçe Spin (trả phí).

 WooCommerce: Lưu sản phẩm, giá cả, tồn kho, thuộc tính và các cài đặt WooCommerce.

 Quản lý tệp: Đổi tên, di chuyển và gán thông tin cho tệp phương tiện.

 Lưu nhanh: Giảm thời gian lưu cài đặt mà không cần tải lại trang.

 Thuộc tính “srcset” cho ảnh: Tối ưu hóa hình ảnh trên các thiết bị khác nhau.

 Cảnh báo lỗi: Hiển thị thông báo chi tiết về lỗi để khắc phục nhanh chóng.

 Điều hướng dễ dàng: Lưu vị trí hiện tại khi chuyển tab và khôi phục khi quay lại.

 Thêm URL thủ công: Nhập URL thủ công để thu thập dữ liệu theo lịch trình.

 Nhập/Xuất cài đặt: Sao lưu và khôi phục cài đặt bằng mã được tạo sẵn.

 Bảng điều khiển chi tiết: Giám sát quá trình thu thập dữ liệu và trạng thái hoạt động.

 PHP mới nhất: Tương thích với các phiên bản PHP mới nhất.

 Hướng dẫn tương tác: Cung cấp hướng dẫn từng bước cấu hình plugin.

**1.2. Yêu cầu**

Để plugin WordPress Content Crawler hoạt động đúng cách, hệ thống cần đáp ứng các yêu cầu sau:

1. Phiên bản PHP

* PHP **>= 8.1**  
  Hầu hết các dịch vụ lưu trữ hiện nay đều hỗ trợ thay đổi phiên bản PHP qua bảng điều khiển. Nếu máy chủ của bạn không đáp ứng yêu cầu này, hãy liên hệ với nhà cung cấp dịch vụ lưu trữ để được hỗ trợ nâng cấp.

2. Phần mở rộng PHP

Plugin yêu cầu các phần mở rộng PHP sau:

* **json:** Mặc định được bật trên hầu hết các máy chủ.
* **mbstring:** Được kích hoạt sẵn trên hầu hết các máy chủ. Nếu chưa bật, bạn có thể yêu cầu nhà cung cấp dịch vụ hỗ trợ.
* **curl:** Thường được bật sẵn. Plugin cần tính năng này để thực hiện các yêu cầu HTTP.
* **dom:** Dùng để xử lý tài liệu HTML/XML, thường được kích hoạt mặc định.
* **fileinfo:** Có sẵn từ PHP 5.3.0 và thường được bật mặc định trên máy chủ.

3. WP-Cron

* WP-Cron là hệ thống lập lịch của WordPress, cần thiết để tự động thu thập, cập nhật và xóa bài viết.
* Hầu hết các máy chủ đều hỗ trợ WP-Cron, nhưng một số dịch vụ lưu trữ giá rẻ có thể vô hiệu hóa tính năng này. Nếu WP-Cron không hoạt động, plugin sẽ không thể tự động thực thi các tác vụ.

**1.3. Cài đặt và kích hoạt**

**Quá trình cài đặt WP Content Crawler tương tự như các plugin khác trên WordPress. Dưới đây là các bước chi tiết:**

**Bước 1: Kiểm tra yêu cầu hệ thống**

**Đảm bảo hệ thống của bạn đáp ứng đầy đủ các Yêu cầu của Plugin trước khi cài đặt.**

**Bước 2: Đăng nhập vào WordPress**

* **Truy cập vào Bảng quản trị WordPress của bạn.**

**Bước 3: Cài đặt Plugin**

1. **Trong thanh bên, vào mục Plugins > Add New.**
2. **Nhấp vào nút Upload Plugin ở đầu trang.**
3. **Nhấn Choose File, chọn tệp .zip nhận được sau khi mua plugin.**
4. **Nhấn Install Now để bắt đầu quá trình cài đặt.**

**Bước 4: Kích hoạt Plugin**

**Sau khi cài đặt thành công, hãy kích hoạt plugin.**

**Bước 5: Kích hoạt Giấy phép**

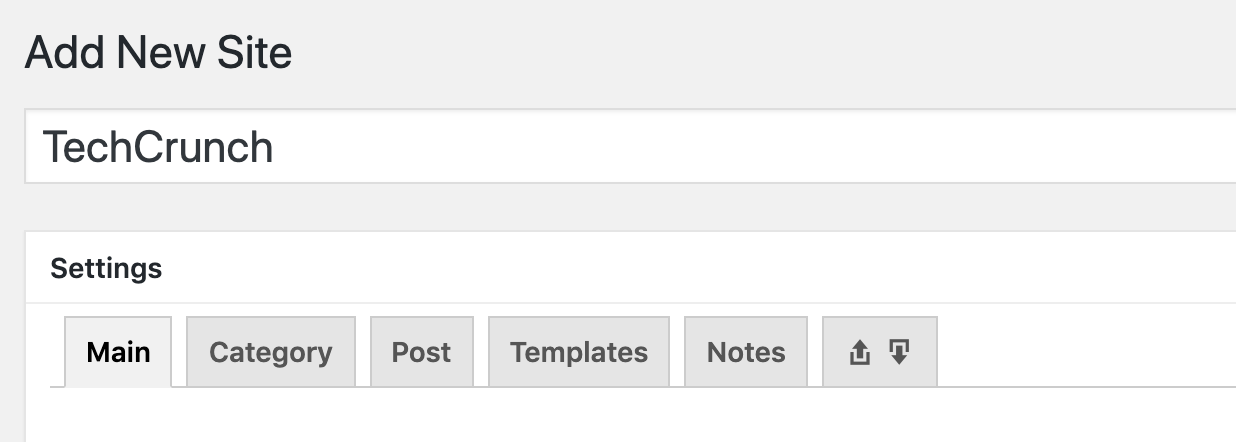
1. **Vào Settings > Content Crawler License Settings.**
2. **Nhập Mã mua hàng bạn nhận được.**
3. **Nhấn Save Changes để lưu và kích hoạt giấy phép.**

**2. Hướng dẫn tạo web với WP CONTENT CRAWLER**

**2.1. Thu thập dữ liệu trang web đầu tiên của bạn**

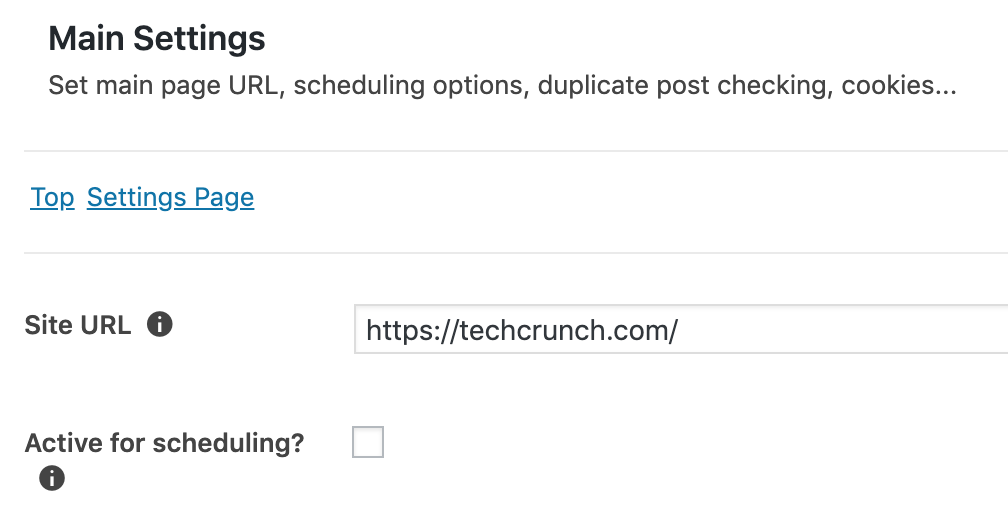
**Trong hướng dẫn này, bạn sẽ học cách thu thập dữ liệu trang web đầu tiên của mình. Trang web chúng ta sẽ thu thập dữ liệu là**[**TechCrunch**](https://techcrunch.com/)**. Trước tiên, hãy tạo một trang web mới. Từ thanh bên của bảng quản trị, hãy nhấp vào .Content Crawler > Add New**

**Hãy đặt tên cho trang web để chúng ta có thể dễ dàng tìm thấy nó trong**[**Trang danh sách trang web**](https://docs.wpcontentcrawler.com/site-list/index.html#site-listing-page)**. Nhập TechCrunchvào thông tin như được hiển thị trong**[**Hình 2.1**](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-site-name)**.Enter site name here**



**Hình 2.1 *Nhập tên trang web***

Tiếp theo, chúng ta cần điền thiết lập ( [URL Site](https://docs.wpcontentcrawler.com/site-settings/tabs/main-tab/main.html#main-page-url) ) vì đây là một trong những thiết lập bắt buộc phải điền. Như đã giải thích trong [Tạo Site Mới](https://docs.wpcontentcrawler.com/site-settings/creating-new-site.html#creating-new-site) , chúng ta nên nhập URL đầy đủ của trang chính của site. Hãy nhập như trong [Hình 2.2](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-site-url) .Main > Site URLhttps://techcrunch.com/



Hình 2.2 *URL trang web*

**2.1.1. Thu thập URL bài đăng**

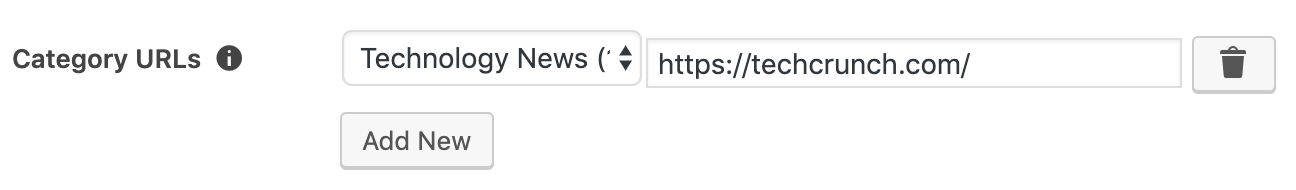
Bây giờ, chúng ta sẽ cấu hình Categorycài đặt ( [Tab Category](https://docs.wpcontentcrawler.com/site-settings/tabs/category-tab/index.html#tab-category) ). Mục đích của cài đặt category là hướng dẫn plugin tìm URL của các bài đăng mà chúng ta muốn lưu. Nó được gọi như vậy vì các category trong các trang web chứa các liên kết đến các bài đăng. Categories về cơ bản là một tập hợp các URL bài đăng. Do đó, chúng ta sẽ cho plugin biết trong các trang category nào nó sẽ thu thập URL bài đăng.category settings

Đối với ví dụ này, hãy để plugin thu thập URL bài đăng từ trang chính của TechCrunch và từ Appsdanh mục. Hãy vào phần cài đặt ( [URL danh mục](https://docs.wpcontentcrawler.com/site-settings/tabs/category-tab/category.html#category-map) ) và nhập URL đầu tiên. Vì chúng ta muốn thu thập URL bài đăng từ trang chính, chúng ta chỉ cần nhập URL của trang chính, đó là , như được hiển thị trong [Hình 2.3](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-category-url-main) .Category > Category URLshttps://techcrunch.com/



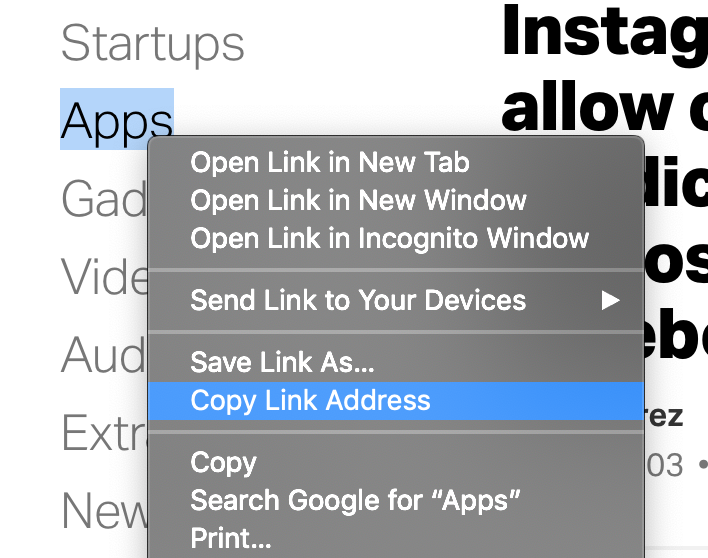
Hình 2.3 *Nhập URL trang chính làm URL danh mục*

Có thể đã nhận thấy, có một mục nhập chọn bên cạnh mục nhập URL trong cài đặt [URL danh mục](https://docs.wpcontentcrawler.com/site-settings/tabs/category-tab/category.html#category-map) , hiện đang hiển thị Uncategorized. Nó hiển thị các danh mục bài đăng khả dụng trong trang web của bạn. Khi bạn chọn một danh mục trong mục nhập chọn này, các URL được thu thập từ trang danh mục sẽ được lưu dưới dạng bài đăng vào danh mục đó. Trong trang web của tôi, tôi đã tạo một danh mục có tên là . Tôi muốn lưu các bài đăng trên trang chính của trang web mục tiêu, TechCrunch, vào . Do đó, tôi chọn danh mục này trong mục nhập chọn như thể hiện trong [Hình 2.4](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-category-category-main) .Technology NewsTechnology News

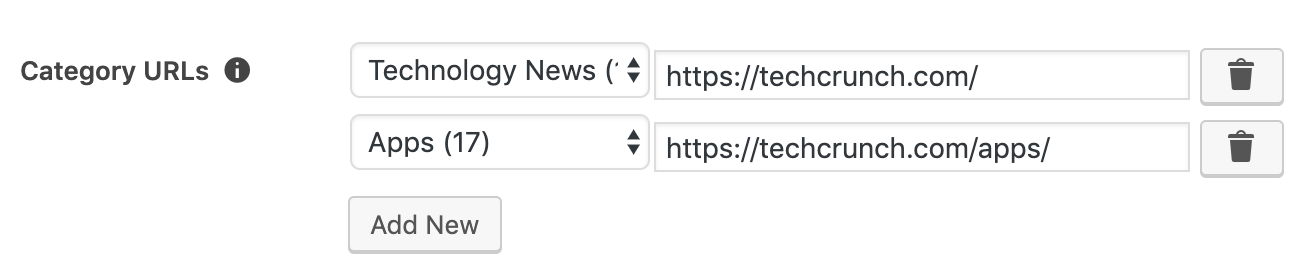


Hình 2.4 *Chọn danh mục mà tôi muốn lưu các bài đăng được thu thập từ URL danh mục mục tiêu*

Bây giờ chúng ta hãy thêm một URL danh mục mới. Nhấp đểThêm mớinút để thêm nhóm đầu vào mới. Thể loại thứ hai mà chúng ta muốn thu thập bài đăng là Appsthể loại. Hãy sao chép URL của nó từ trang đích ( [Hình 2.5](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-copy-apps-category-url) ) và dán vào đầu vào URL của nhóm đầu vào mới được thêm vào ( [Hình 2.6](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-category-config-apps) ). Tôi cũng đã chọn Appsthể loại mà tôi đã tạo trước đó trong trang của mình để các bài đăng từ Appsthể loại TechCrunch sẽ được lưu vào Appsthể loại của trang của tôi.

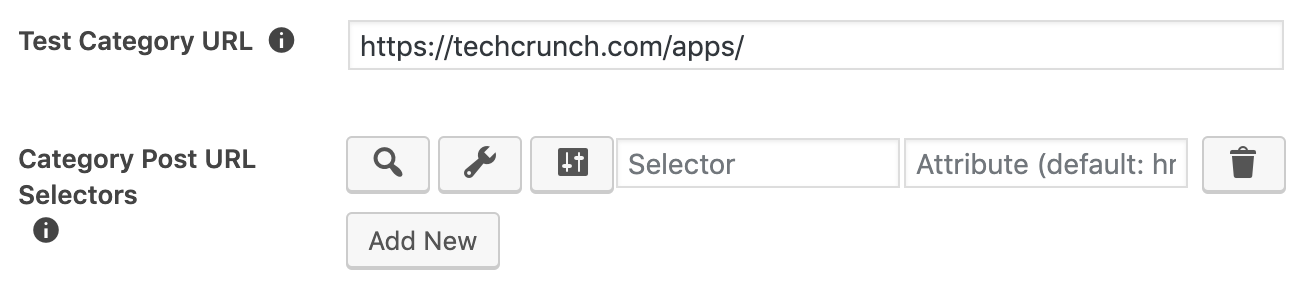


Hình 2.5 *Sao chép URL của Appsdanh mục*



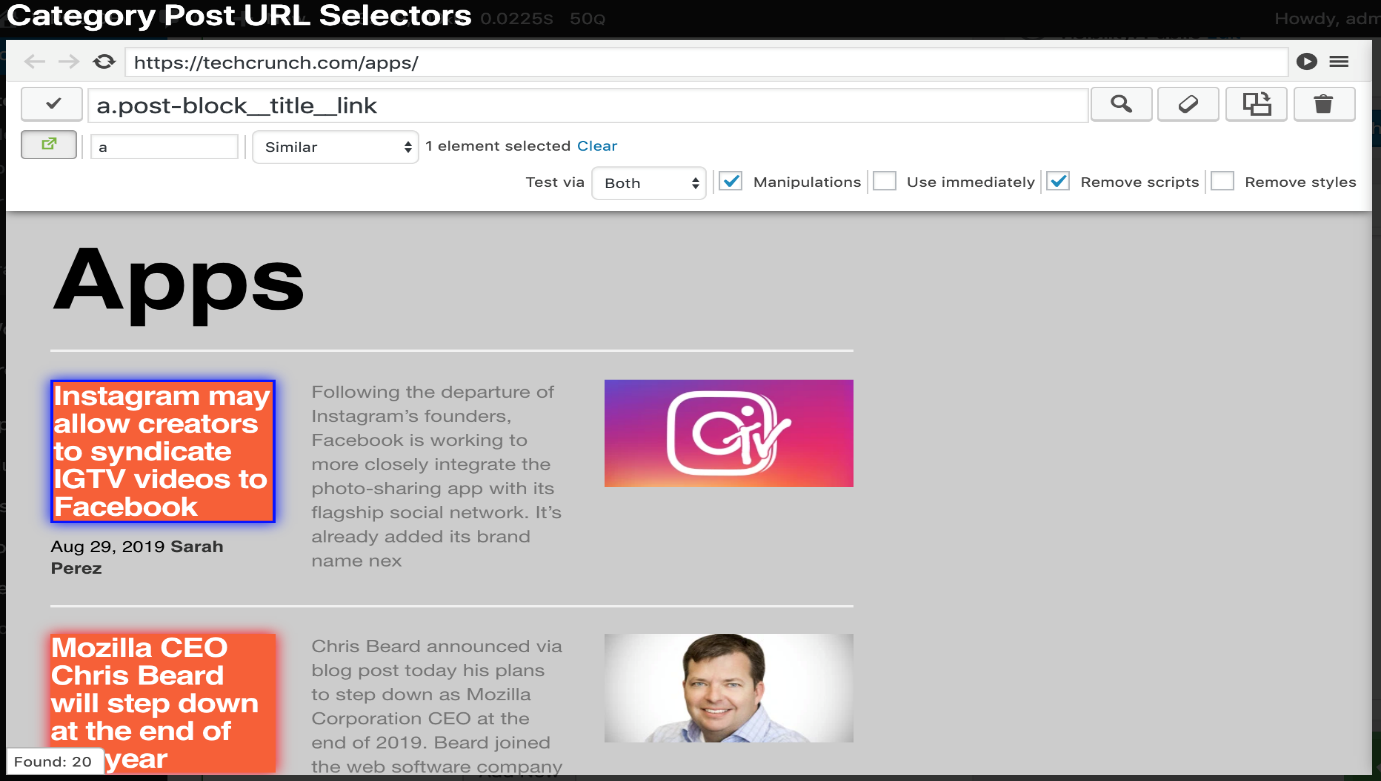
Hình 2.6 *Thêm Appsdanh mục vào cài đặtCategory > Category URLs*

Bây giờ chúng ta đã xác định URL danh mục mà chúng ta muốn thu thập URL bài đăng, hãy cho plugin biết cách tìm URL bài đăng trong các trang danh mục này. Dán URL của Appsdanh mục vào đầu vào ( [Hình 2.7](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-test-category-url) , [Kiểm tra URL danh mục](https://docs.wpcontentcrawler.com/site-settings/tabs/category-tab/category.html#test-url-category) ). Từ bây giờ, các nút kiểm tra có trong [Tab danh mục](https://docs.wpcontentcrawler.com/site-settings/tabs/category-tab/index.html#tab-category) sẽ sử dụng URL này để kiểm tra bộ chọn CSS.Category > Test Category URL



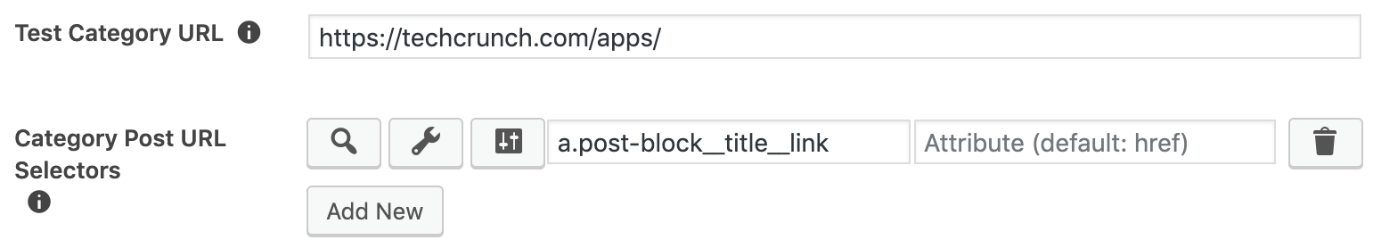
Hình 2.7 *Nhập URL danh mục thử nghiệm*

Để tìm bộ chọn CSS cho URL bài đăng, hãy nhấp vàonút của thiết lập [Post URL Selectors](https://docs.wpcontentcrawler.com/site-settings/tabs/category-tab/category.html#category-post-link-selectors) . Nút này sẽ mở [Visual Inspector](https://docs.wpcontentcrawler.com/site-settings/visual-inspector/visual-inspector.html#visual-inspector) . Vì chúng ta đã nhập URL danh mục thử nghiệm, nên visual inspector sẽ tự động tải URL đó. Khi bạn mở visual inspector cho [Post URL Selectors](https://docs.wpcontentcrawler.com/site-settings/tabs/category-tab/category.html#category-post-link-selectors) , nó được tự động cấu hình để tìm similar acác phần tử. Mặc dù nó đã được kích hoạt, hãy đảm bảo [Toggle hover select](https://docs.wpcontentcrawler.com/site-settings/visual-inspector/visual-inspector.html#toggle-hover-select) () đang hoạt động. Bây giờ, hãy nhấp vào một trong các tiêu đề được hiển thị trong trình kiểm tra trực quan. Ngay sau khi bạn thực hiện thao tác này, trình kiểm tra trực quan sẽ tìm thấy một bộ chọn CSS khớp với tất cả các URL bài đăng như được hiển thị trong [Hình 2.8](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-visual-inspector-finding-post-urls) .



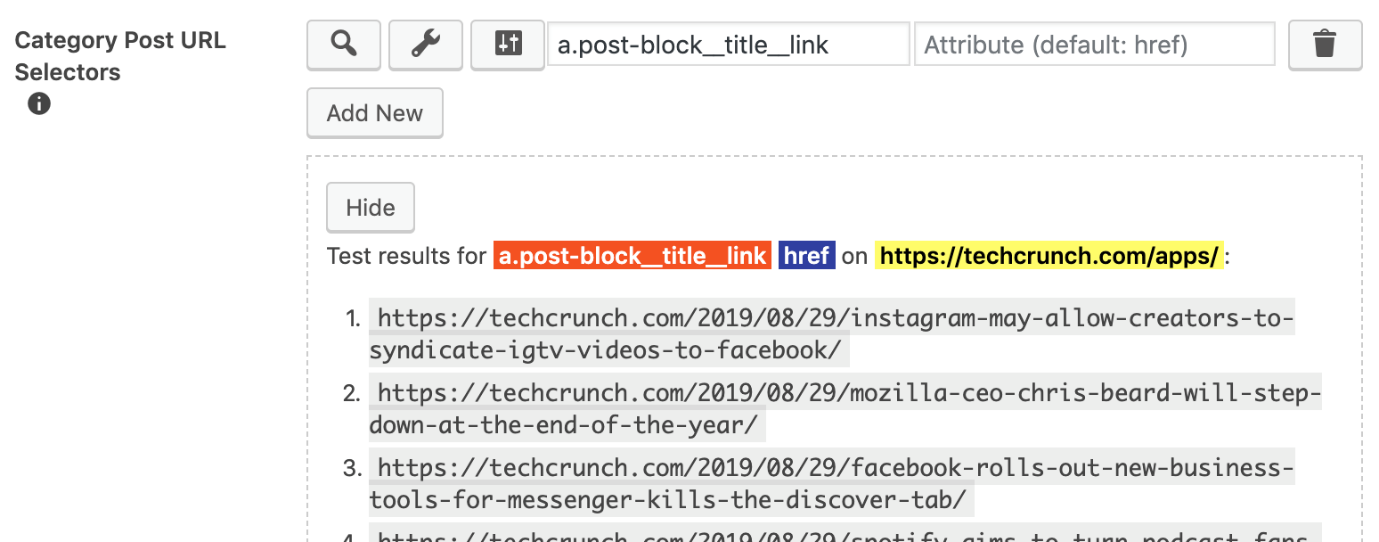
Hình 2.8 *Nhấp vào phần tử tiêu đề (được tô sáng bằng đường viền màu xanh) để tìm tất cả 20các URL bài đăng trong Appstrang TechCrunch*

Bây giờ, hãy nhấp vào nút [Sử dụng bộ chọn CSS](https://docs.wpcontentcrawler.com/site-settings/visual-inspector/visual-inspector.html#use-css-selector) () để sao chép bộ chọn CSS đã tìm thấy vào đầu vào bộ chọn CSS của thiết lập [Bộ chọn URL bài đăng](https://docs.wpcontentcrawler.com/site-settings/tabs/category-tab/category.html#category-post-link-selectors) . Bạn sẽ thấy bộ chọn CSS được sao chép như thể hiện trong [Hình 2.9](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-css-selector-copied-for-category-url-selectors) .



Hình 2.9 *Đã tìm thấy bộ chọn CSS được sao chép vào đầu vào bộ chọn CSS của cài đặt.Category Post URL Selectors*

Plugin cho phép bạn kiểm tra mọi bộ chọn CSS được nhập vào đầu vào bộ chọn CSS của cài đặt bằng cách nhấp vào nút. Bây giờ, hãy nhấp vào nút này để kiểm tra xem bộ chọn CSS của chúng ta có thể tìm thấy URL bài đăng trong Appstrang hay không. Bạn sẽ thấy kết quả kiểm tra như được hiển thị trong [Hình 2.10](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-testing-category-url-selector) . Như bạn có thể thấy, có các URL bài đăng được hiển thị trong kết quả kiểm tra. Bạn có thể sao chép một trong các URL và mở nó trong một tab mới để đảm bảo URL là chính xác.



Hình 2.10 *Kiểm tra bộ chọn URL danh mục*

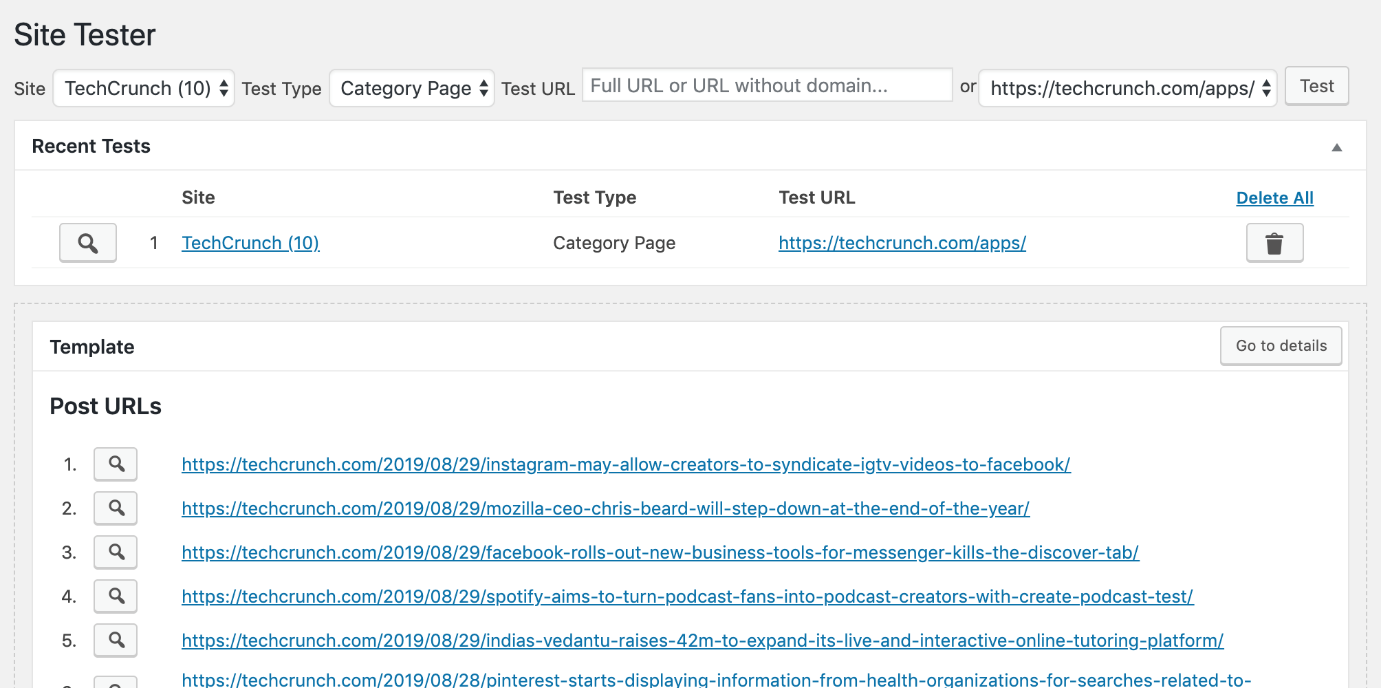
Bước cuối cùng và quan trọng nhất là kiểm tra cài đặt của bạn. Plugin có một trang có tên là [Tester Page](https://docs.wpcontentcrawler.com/tester/index.html#wp-content-crawler-site-tester) . Trang này được sử dụng để kiểm tra cài đặt trang web của bạn và đảm bảo mọi thứ đều theo thứ tự trước khi bật cài đặt trang web của bạn để tự động thu thập dữ liệu. Bây giờ, hãy nhanh chóng kiểm tra cài đặt danh mục của chúng tôi bằng cách sử dụng [Tester Page](https://docs.wpcontentcrawler.com/tester/index.html#wp-content-crawler-site-tester) .

Từ thanh bên của bảng điều khiển quản trị, hãy nhấp vào . Thao tác này sẽ mở [Tester Page](https://docs.wpcontentcrawler.com/tester/index.html#wp-content-crawler-site-tester) . Cấu hình các đầu vào của trang này như sau:Content Crawler > Tester

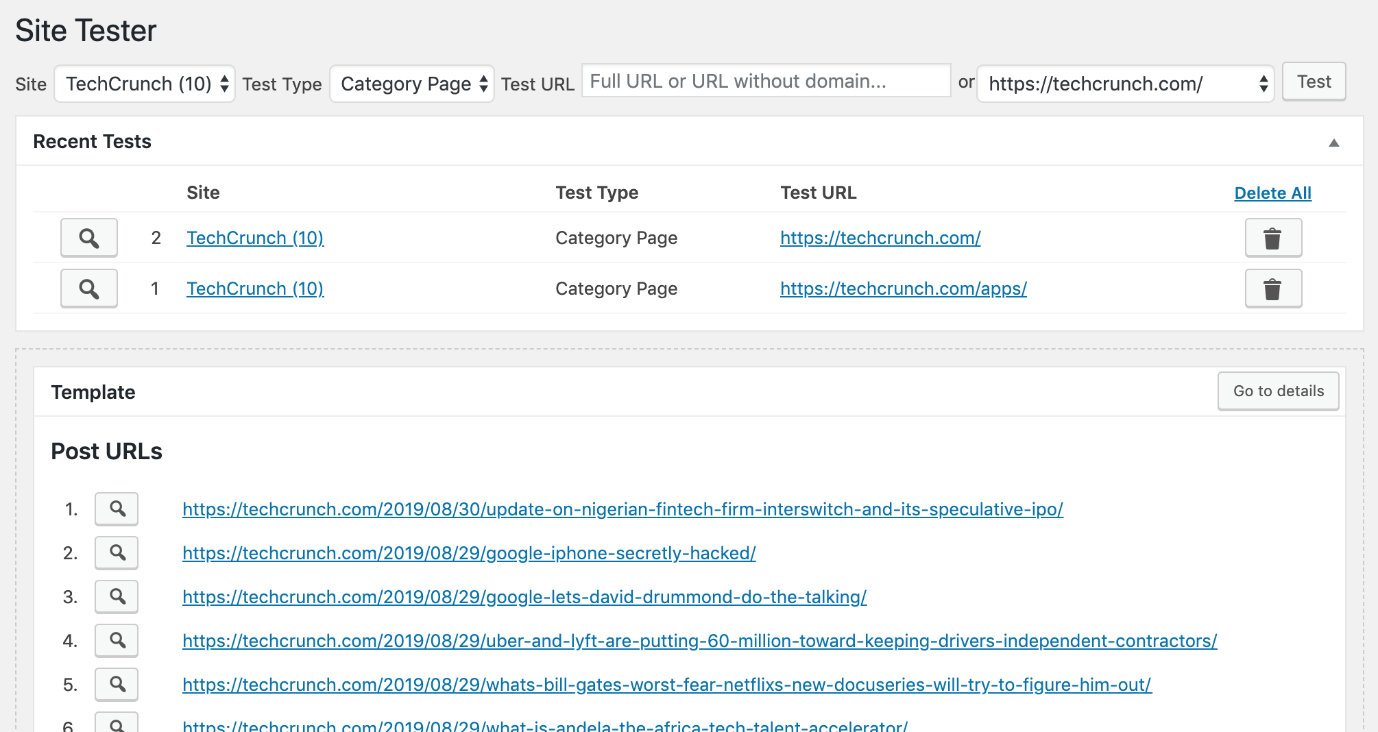
|  |  |
| --- | --- |
| **Địa điểm:** | Chọn trang web có tên làTechCrunch |
| **Loại thử nghiệm:** | Category page |
| **URL kiểm tra:** | Để trống ô nhập và chọn https://techcrunch.com/apps/từ tùy chọn chọn. |

Bây giờ, nhấp vào đểBài kiểm tranút. Sau khi chúng được tải, kết quả sẽ tương tự như kết quả hiển thị trong [Hình 2.11](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-testing-category-settings) . Nếu không có kết quả nào, điều đó có nghĩa là cài đặt trang danh mục của bạn không tìm thấy URL bài đăng và bạn nên sửa cài đặt của mình. Nếu kết quả hiển thị URL của bài đăng bạn muốn lưu, thì cài đặt danh mục của bạn đã được cấu hình đúng.

Cuối cùng, chúng ta hãy thử nghiệm trang chính. Như được hiển thị trong [Hình 2.12](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-testing-category-settings-2) , URL bài đăng được tìm thấy trong trang chính của TechCrunch. Điều này có nghĩa là cài đặt danh mục của chúng tôi được cấu hình đúng.



Hình 2.11 *Kết quả thử nghiệm cho AppsloạiTechCrunch*

Hình 2.12 *Kết quả thử nghiệm cho trang chính củaTechCrunch*

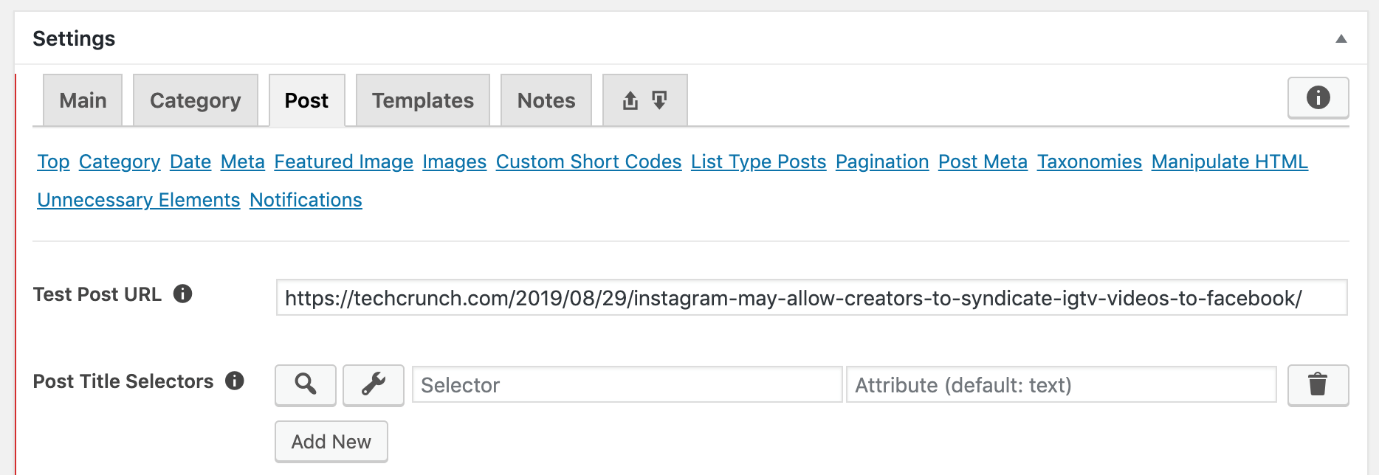
**2.1.2. Cấu hình thông tin nào sẽ được lấy từ các trang bài đăng**[**ℑ**](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#configuring-what-information-should-be-retrieved-from-post-pages)

Hiện tại, plugin có thể tìm URL của các bài đăng từ hai danh mục mà chúng tôi đã nhập vào cài đặt [Category URLs](https://docs.wpcontentcrawler.com/site-settings/tabs/category-tab/category.html#category-map) . Bây giờ, đã đến lúc cho plugin biết cách tìm thông tin chi tiết về bài đăng, chẳng hạn như tiêu đề bài đăng và nội dung bài đăng.

Trong thanh bên của bảng quản trị, hãy nhấp vào và nhấp để mở cài đặt.Content Crawler > All SitesTechCrunch

Mặc dù plugin có nhiều cài đặt, chúng ta không cần phải điền vào từng cài đặt. Hiện tại, chúng ta chỉ quan tâm đến việc lưu tiêu đề và nội dung của bài đăng. Do đó, cấu hình các cài đặt này là đủ.

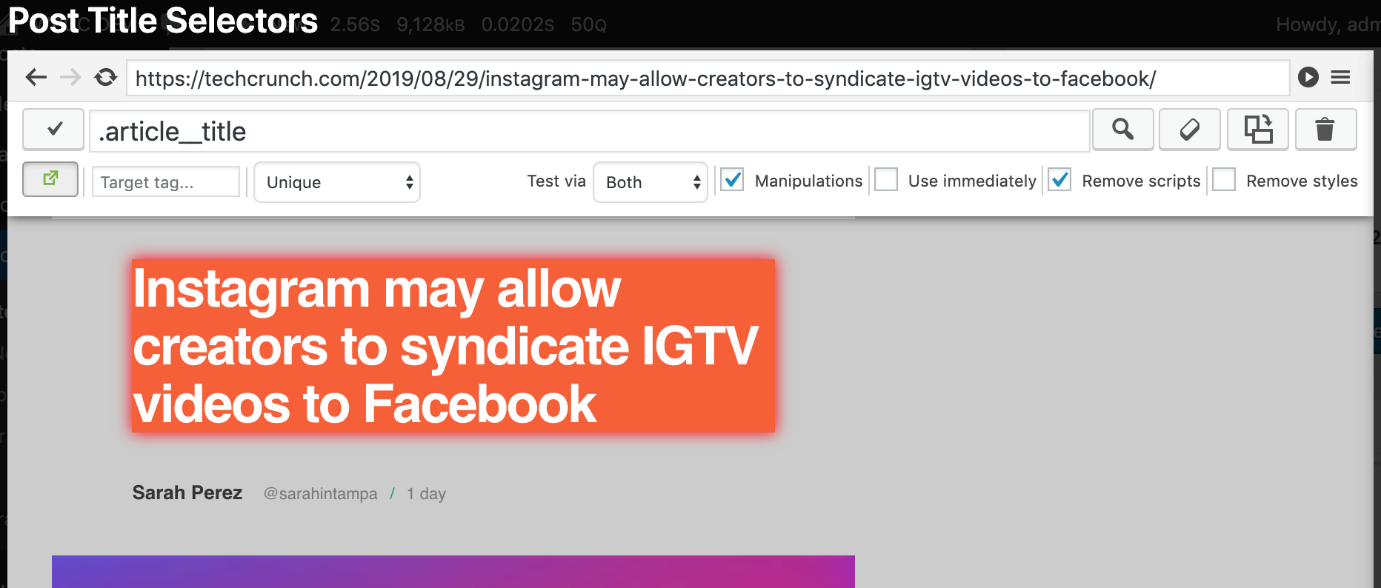
Trước hết, chúng ta cần một URL bài đăng mà chúng ta sẽ sử dụng để cấu hình các thiết lập. Hãy sao chép URL đầu tiên được hiển thị trong kết quả Appskiểm tra danh mục ( [Hình 2.11](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-testing-category-settings) ) và dán vào thiết lập ( [Test Post URL](https://docs.wpcontentcrawler.com/site-settings/tabs/post-tab/post.html#test-url-post) ) như được hiển thị trong [Hình 2.13](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-test-post-url) . Plugin sẽ sử dụng URL bài đăng này khi kiểm tra các bộ chọn CSS trong [Tab Bài đăng](https://docs.wpcontentcrawler.com/site-settings/tabs/post-tab/index.html#tab-post) và, trong một số trường hợp, trong [Tab Mẫu](https://docs.wpcontentcrawler.com/site-settings/tabs/templates-tab/index.html#tab-templates) .Post > Test Post URL



Hình 2.13 *URL của bài đăng từ trang web đích được nhập vào cài đặt.Test Post URL*

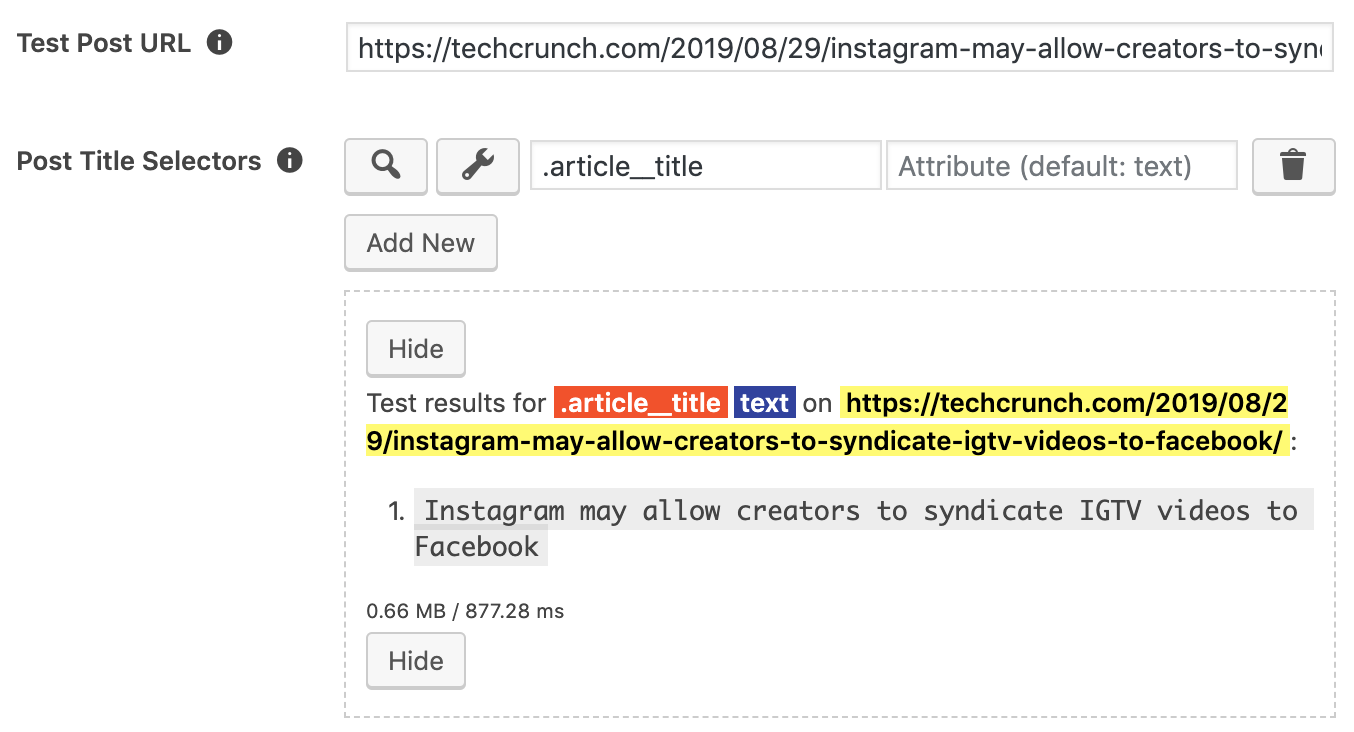
Hãy tìm bộ chọn CSS cho tiêu đề bài đăng để plugin có thể tìm thấy và lưu khi tạo bài đăng. Như chúng ta đã làm khi tìm bộ chọn CSS cho [Bộ chọn URL bài đăng](https://docs.wpcontentcrawler.com/site-settings/tabs/category-tab/category.html#category-post-link-selectors) , chúng ta sẽ sử dụng lại [Visual Inspector](https://docs.wpcontentcrawler.com/site-settings/visual-inspector/visual-inspector.html#visual-inspector) . Nhấp đểnút cài đặt ( [Chọn tiêu đề bài đăng](https://docs.wpcontentcrawler.com/site-settings/tabs/post-tab/post.html#post-title-selectors) ). Thao tác này sẽ mở [Visual Inspector](https://docs.wpcontentcrawler.com/site-settings/visual-inspector/visual-inspector.html#visual-inspector) và tải ngay [URL bài đăng thử nghiệm](https://docs.wpcontentcrawler.com/site-settings/tabs/post-tab/post.html#test-url-post) .Post > Post Title Selectors

Visual inspector sẽ được cấu hình tự động để tìm một uniquephần tử (Xem: [Selection behavior](https://docs.wpcontentcrawler.com/site-settings/visual-inspector/visual-inspector.html#selection-behavior) ), vì chỉ được có một tiêu đề bài đăng. Bây giờ, hãy nhấp vào tiêu đề của bài đăng như được hiển thị trong [Hình 2.14](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-visual-inspector-finding-title-selector) . Visual inspector sẽ tìm một bộ chọn CSS duy nhất tìm tiêu đề bài đăng trong trang bài đăng mục tiêu và hiển thị nó trong đầu vào [bộ chọn CSS](https://docs.wpcontentcrawler.com/site-settings/visual-inspector/visual-inspector.html#css-selector) .



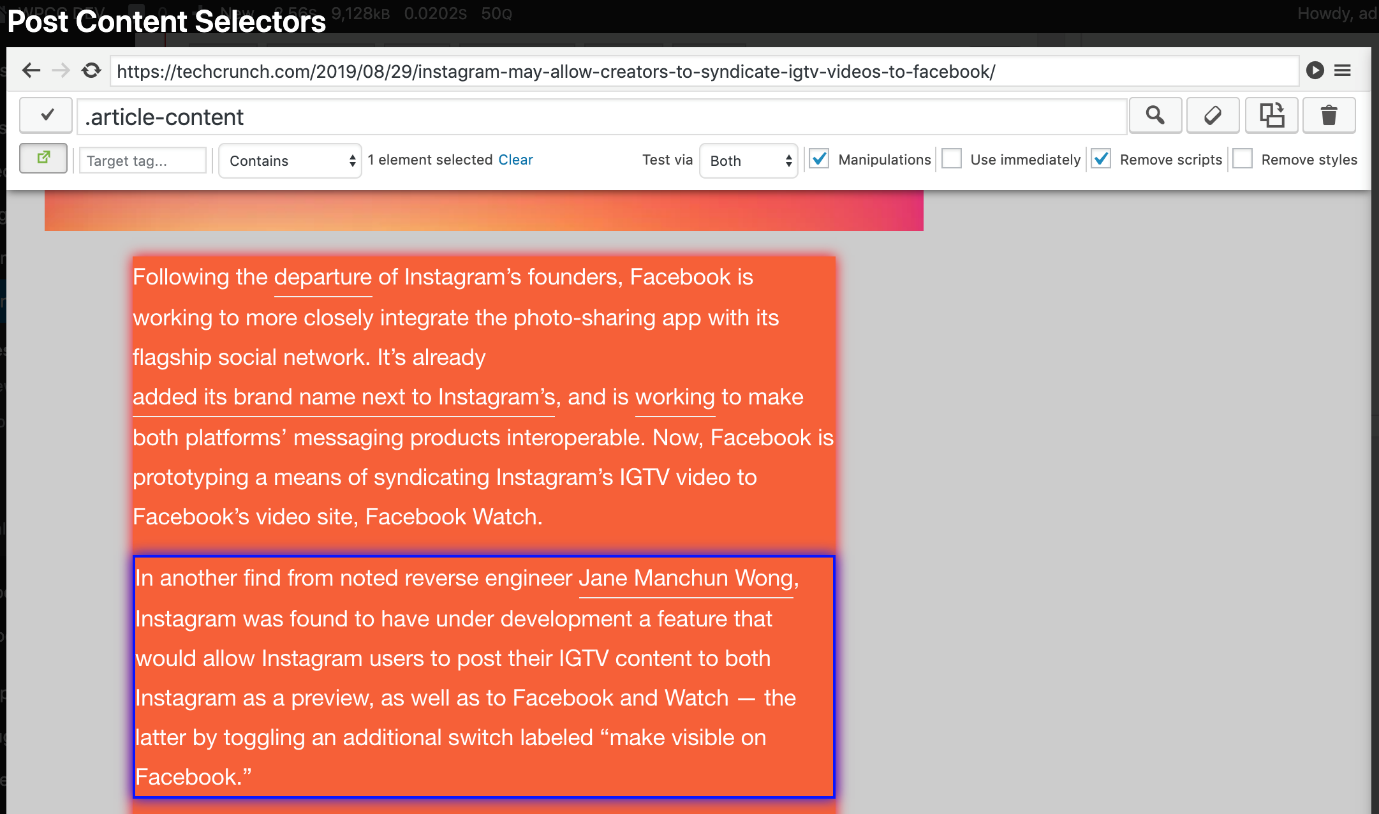
Hình 2.14 *Nhấp vào tiêu đề trong trình kiểm tra trực quan để tìm bộ chọn CSS của nó*

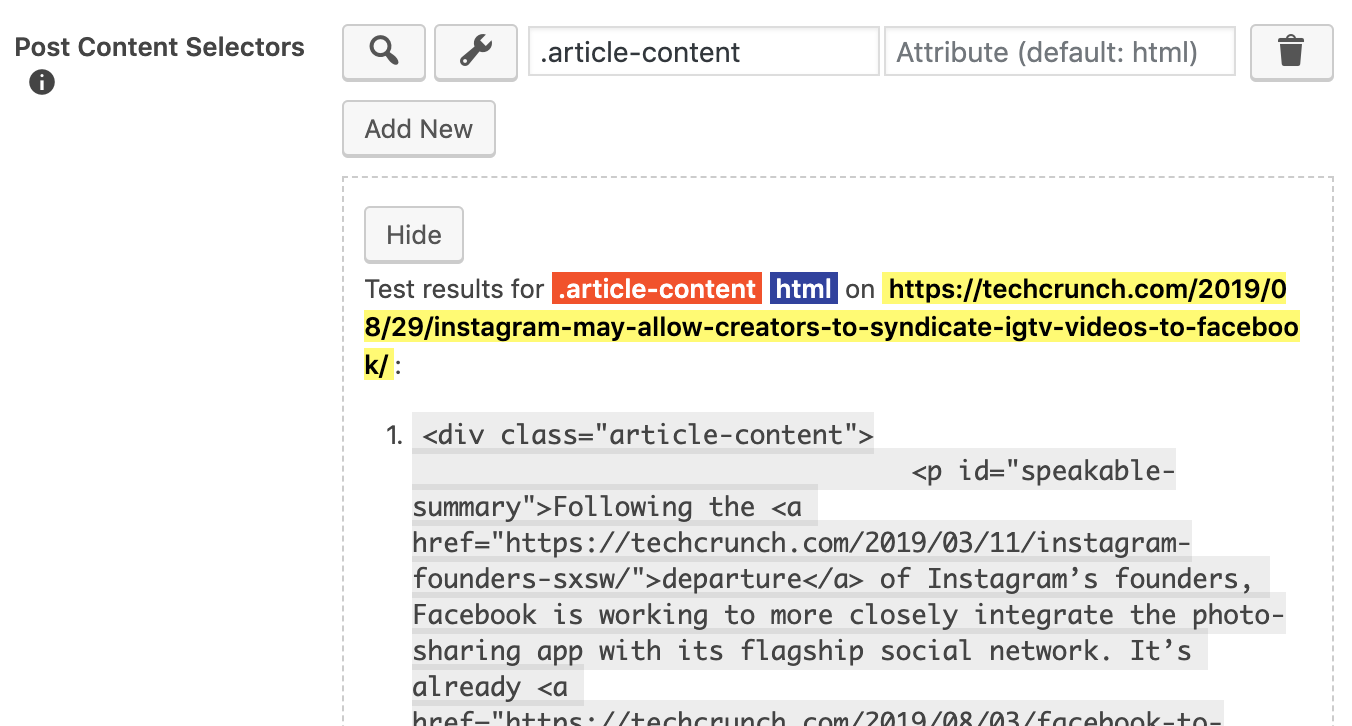
Bây giờ, hãy nhấp vào nút [Sử dụng bộ chọn CSS](https://docs.wpcontentcrawler.com/site-settings/visual-inspector/visual-inspector.html#use-css-selector) () để sao chép bộ chọn CSS đã tìm thấy vào đầu vào bộ chọn CSS của thiết lập [Bộ chọn tiêu đề bài đăng](https://docs.wpcontentcrawler.com/site-settings/tabs/post-tab/post.html#post-title-selectors) . Bạn sẽ thấy bộ chọn CSS được sao chép như thể hiện trong [Hình 2.15](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-post-title-selector-test) . Nhấp vào để kiểm tra bộ chọn CSS. Bạn sẽ thấy tiêu đề bài đăng trong kết quả kiểm tra như thể hiện trong [Hình 2.15](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-post-title-selector-test) .



Hình 2.15 *Đã tìm thấy bộ chọn CSS được sao chép vào đầu vào bộ chọn CSS của thiết lập*[*Bộ chọn tiêu đề bài đăng*](https://docs.wpcontentcrawler.com/site-settings/tabs/post-tab/post.html#post-title-selectors)*. Bộ chọn được kiểm tra bằng cách nhấp vàocái nút.*

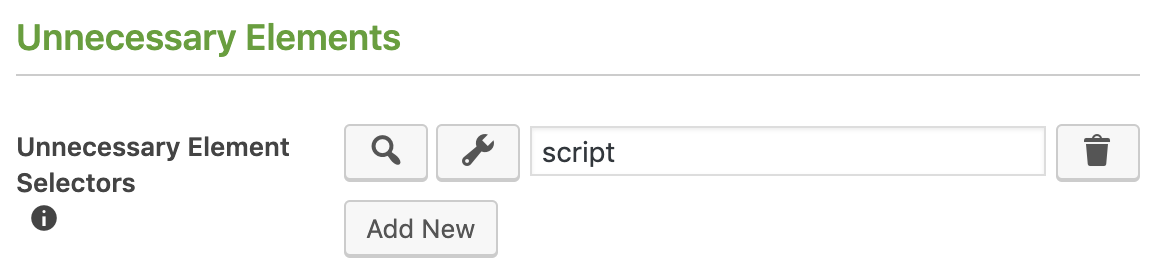
Bây giờ, hãy yêu cầu plugin lưu nội dung bài đăng. Đi tới ( [Post Content Selectors](https://docs.wpcontentcrawler.com/site-settings/tabs/post-tab/post.html#post-content-selectors) ) và nhấp vàoPost > Post Content Selectorsnút để mở thanh tra trực quan. Thanh tra trực quan được cấu hình tự động để tìm một phần tử mà contains phần tử đã chọn (Xem: [Hành vi lựa chọn](https://docs.wpcontentcrawler.com/site-settings/visual-inspector/visual-inspector.html#selection-behavior) ). Nhấp vào một đoạn văn của nội dung bài đăng như được hiển thị trong [Hình 2.16](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-visual-inspector-finding-content-selector) . Thanh tra trực quan sẽ tìm một bộ chọn CSS cho phần tử chứa đoạn văn đã nhấp và tô sáng các phần tử khớp với bộ chọn CSS đã tìm thấy. Nếu phần tử được tô sáng không chứa tất cả nội dung bài đăng, bạn cũng có thể nhấp vào một đoạn văn khác. Khi bạn thực hiện việc đó, thanh tra trực quan sẽ tìm một bộ chọn CSS khớp với phần tử chứa hai đoạn văn đã chọn. Sau khi bạn hài lòng với kết quả, hãy nhấp vàonút để sao chép bộ chọn CSS vào đầu vào bộ chọn CSS của thiết lập [Bộ chọn Nội dung Bài đăng](https://docs.wpcontentcrawler.com/site-settings/tabs/post-tab/post.html#post-content-selectors) như được hiển thị trong [Hình 2.17](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-testing-post-content-selectors) . Bạn có thể nhấp đểnút để kiểm tra bộ chọn CSS và đảm bảo kết quả như bạn mong đợi.

Hình 2.16 *Nhấp vào một đoạn văn (được đánh dấu bằng khung màu xanh) trong nội dung bài đăng để tìm bộ chọn CSS cho phần tử cha của đoạn văn đó.*



Hình 2.17 *Kiểm tra bộ chọn CSS phù hợp với nội dung bài đăng.*

Điều cuối cùng cần làm khi tạo cài đặt bài đăng là đảm bảo bạn ngăn chặn scriptcác thành phần xuất hiện trong nội dung bài đăng của mình. Điều này rất quan trọng vì việc thêm mã JavaScript không xác định vào trang web của bạn khiến trang web của bạn dễ bị tổn thương về mặt bảo mật. Việc ngăn chặn các thành phần tập lệnh khá dễ dàng. Đi tới ( [Unnecessary Element Selectors](https://docs.wpcontentcrawler.com/common-options/unnecessary-elements.html#x-unnecessary-element-selectors) ) và nhập như thể hiện trong [Hình 2.18](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-unnecessary-elements-script) . Thao tác này sẽ xóa tất cả các thành phần khỏi mã nguồn của trang đích.Post > Unnecessary Elements > Unnecessary Element Selectorsscriptscript

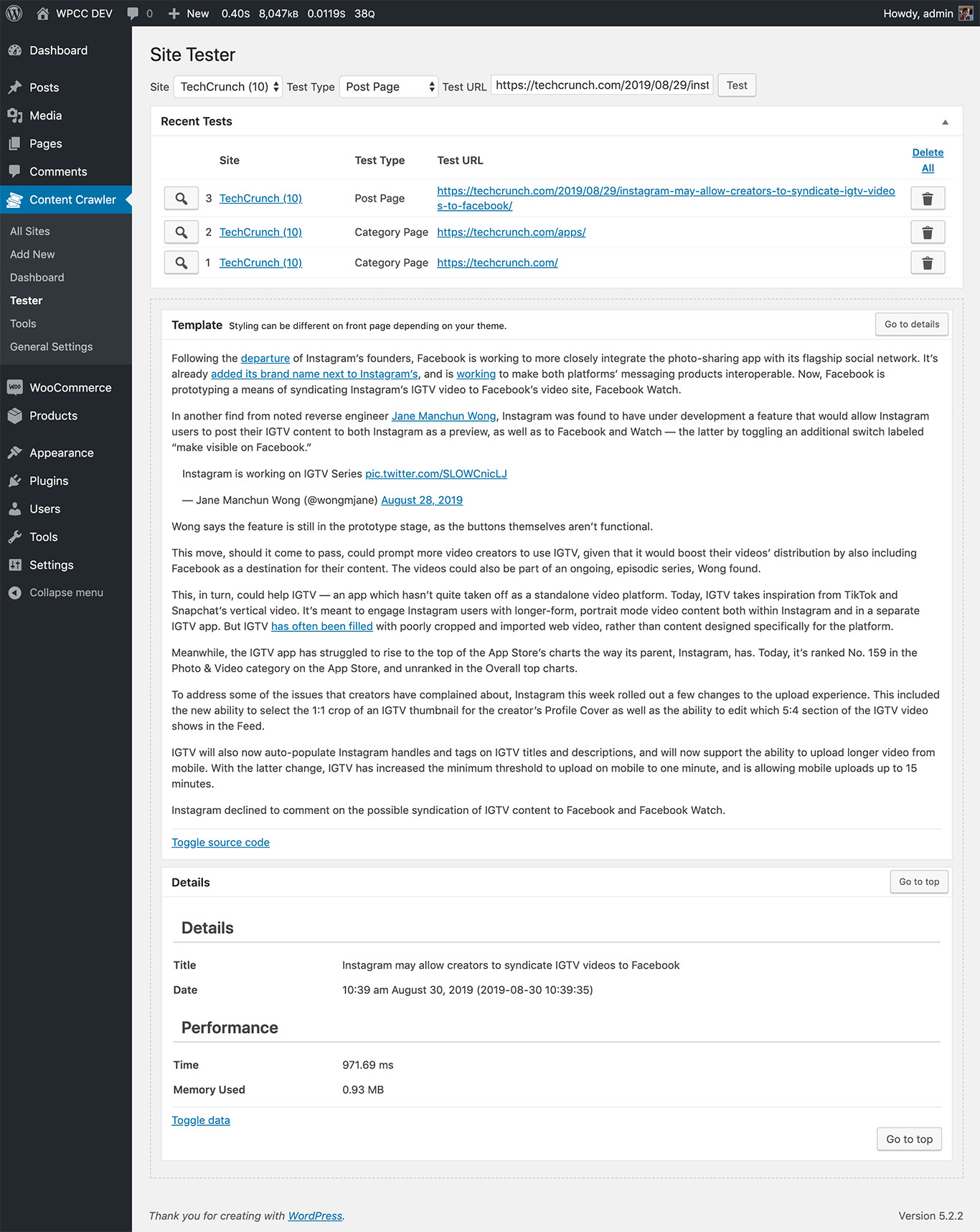


Hình 2.18 *Xóa scriptcác phần tử*

Vậy là xong. Lưu cài đặt (Xem: [Lưu cài đặt](https://docs.wpcontentcrawler.com/site-settings/saving-settings.html#saving-settings) ) và từ thanh bên của bảng quản trị, nhấp để đảm bảo cài đặt được cấu hình đúng. Trong [Tester Page](https://docs.wpcontentcrawler.com/tester/index.html#wp-content-crawler-site-tester) , cấu hình đầu vào như sau:Content Crawler > Tester

|  |  |
| --- | --- |
| **Địa điểm:** | Chọn trang web có tên làTechCrunch |
| **Loại thử nghiệm:** | Post page |
| **URL kiểm tra:** | Nhập URL chúng tôi đã nhập vào cài đặtTest Post URL |

Nhấp đểBài kiểm trađể kiểm tra cấu hình trang bài đăng của bạn. Bạn sẽ thấy kết quả tương tự như kết quả hiển thị trong [Hình 2.19](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-tester-post-page) . Trong [Phần Mẫu](https://docs.wpcontentcrawler.com/tester/index.html#tester-page-section-template) của kết quả, nội dung bài đăng được hiển thị. Tiêu đề bài đăng được hiển thị trong [Phần Chi tiết](https://docs.wpcontentcrawler.com/tester/index.html#tester-page-section-details) . Kết quả cho thấy chúng tôi đã cấu hình thành công cài đặt trang bài đăng để lấy nội dung bài đăng và tiêu đề bài đăng.



Hình 2.19 *Kết quả kiểm tra trang bài viết*

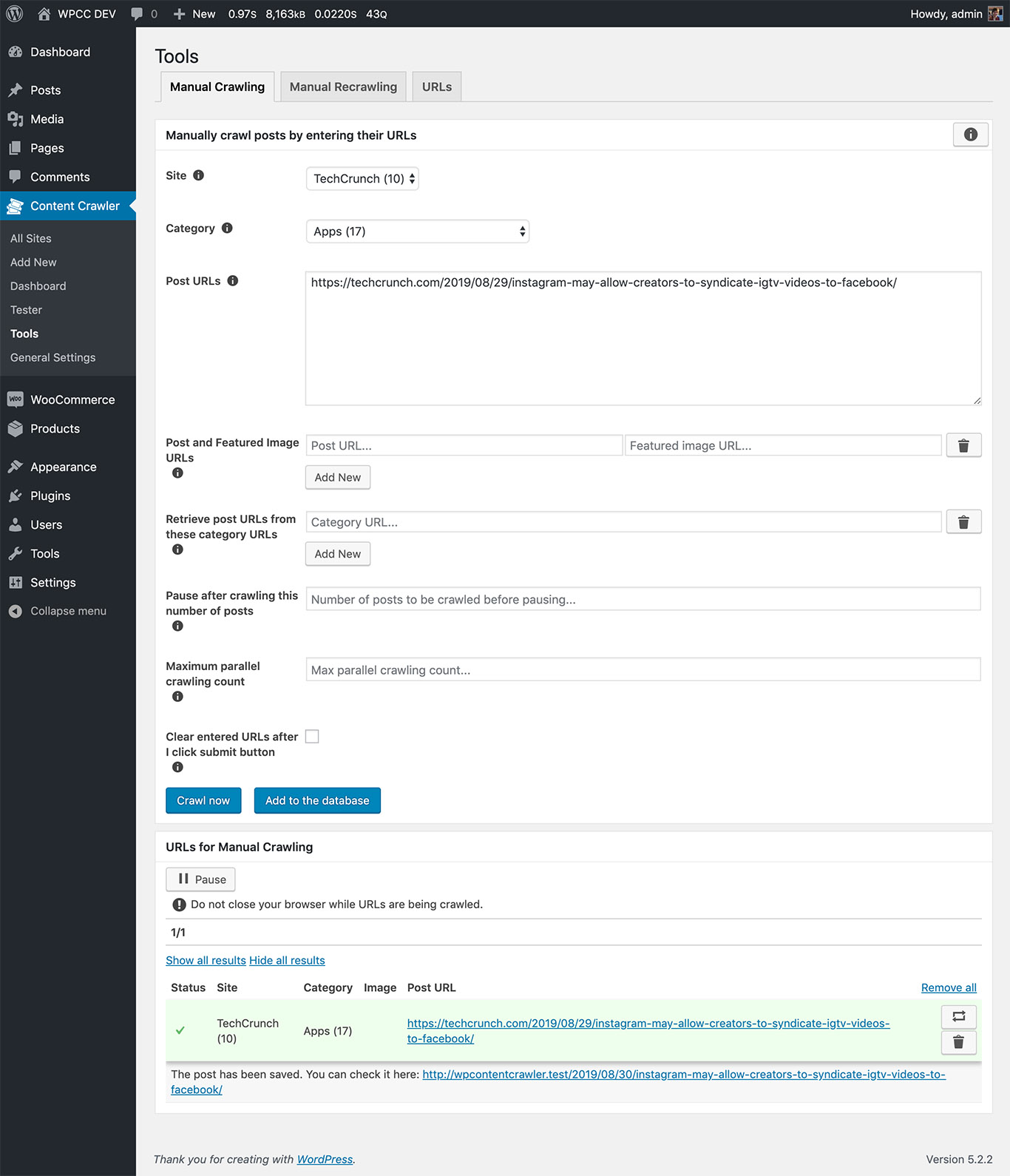
**2.1.3. Lưu bài viết thử nghiệm**[**ℑ**](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#saving-a-test-post)

Bây giờ chúng ta đã cấu hình cài đặt trang bài đăng và đảm bảo rằng chúng được cấu hình đúng bằng cách kiểm tra chúng bằng [Tester Page](https://docs.wpcontentcrawler.com/tester/index.html#wp-content-crawler-site-tester) , chúng ta đã sẵn sàng để lưu bài đăng. Chúng ta sẽ sử dụng [Manual Crawling Tool](https://docs.wpcontentcrawler.com/tools/manual-crawling.html#tool-manual-crawling) .

Trong thanh bên của bảng điều khiển quản trị, hãy nhấp vào ( [Trang Công cụ](https://docs.wpcontentcrawler.com/tools/index.html#wp-content-crawler-tools) ). Đảm bảo tab đang hoạt động. Đây là công cụ thu thập thông tin thủ công. Nó có thể lưu một hoặc nhiều bài đăng khi URL của chúng được cung cấp cho nó. Cấu hình cài đặt của công cụ này như sau:Content Crawler > ToolsManual Crawling

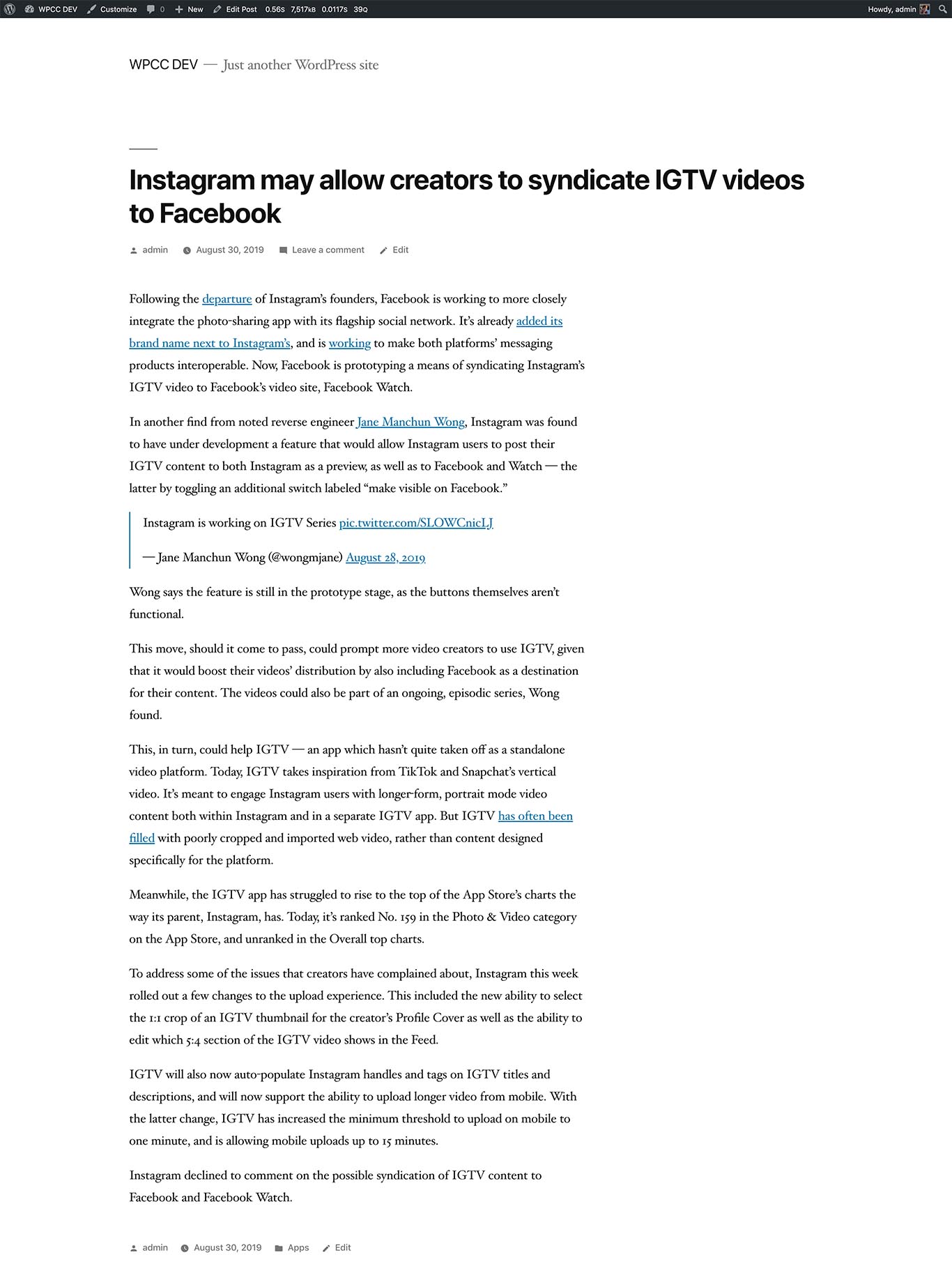
|  |  |
| --- | --- |
| **Địa điểm:** | Lựa chọnTechCrunch |
| **Loại:** | Chọn Appsdanh mục. Nếu danh mục này không có sẵn trên trang web của bạn, bạn có thể để danh mục mặc định hiển thị trong mục nhập chọn. |
| **URL bài viết:** | Dán [URL Bài đăng thử nghiệm](https://docs.wpcontentcrawler.com/site-settings/tabs/post-tab/post.html#test-url-post) mà chúng ta đã sử dụng khi cấu hình cài đặt trong [Tab Bài đăng](https://docs.wpcontentcrawler.com/site-settings/tabs/post-tab/index.html#tab-post) . |

Sau khi thiết lập các giá trị này, hãy nhấp vào nút. Ngay khi bạn nhấp vào nút này, công cụ sẽ bắt đầu thu thập dữ liệu URL bạn đã nhập vào trường. Trong vài giây, quá trình thu thập dữ liệu sẽ hoàn tất. Bạn sẽ thấy kết quả như được hiển thị trong [Hình 2.20](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-manual-crawling) .Crawl NowPost URLs



Hình 2.20 [*Công cụ thu thập dữ liệu thủ công*](https://docs.wpcontentcrawler.com/tools/manual-crawling.html#tool-manual-crawling)*với các thiết lập được cấu hình. Kết quả thu thập dữ liệu được hiển thị ở cuối trang.*

Kết quả của quá trình thu thập dữ liệu được hiển thị ở phần dưới cùng của trang. Hãy kiểm tra bài đăng được lưu vào trang web của chúng tôi. URL của bài đăng được hiển thị trong kết quả thu thập dữ liệu. Bạn có thể nhấp trực tiếp vào bài đăng đó để mở bài đăng trong một tab mới. Bạn có thể thấy bài đăng được thu thập dữ liệu thủ công trong [Hình 2.21](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-manually-crawled-post) .



Hình 2.21 *Bài viết được thu thập thông tin bằng cách sử dụng*[*Công cụ thu thập thông tin thủ công*](https://docs.wpcontentcrawler.com/tools/manual-crawling.html#tool-manual-crawling)*.*

Sau khi bạn quan sát bài đăng được thu thập thủ công và hài lòng với kết quả, giờ là lúc bật tính năng thu thập tự động, tức là scheduling.

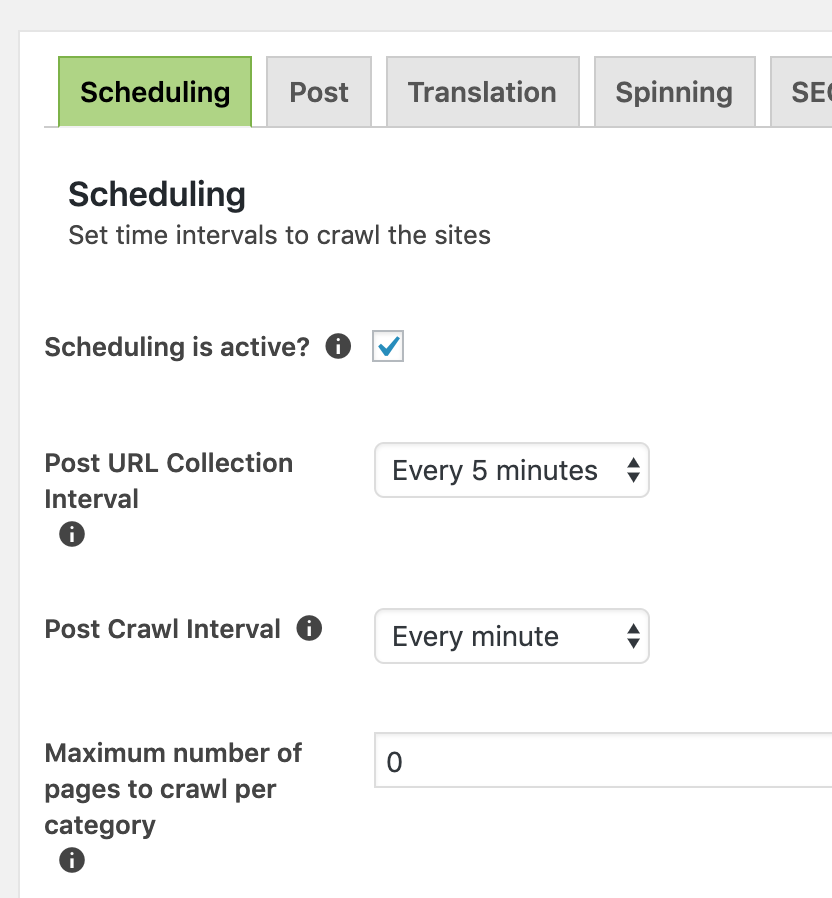
**2.1.4. Cho phép thu thập dữ liệu tự động**[**ℑ**](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#enabling-automatic-crawling)

Nói cách khác, schedulingphải bật tính năng thu thập dữ liệu tự động trong [Trang Cài đặt chung](https://docs.wpcontentcrawler.com/general-settings/index.html#general-settings-page) của plugin và trong [Trang Cài đặt trang web](https://docs.wpcontentcrawler.com/site-settings/index.html#site-settings-page) . Nếu tắt tính năng lập lịch trong [Trang Cài đặt chung](https://docs.wpcontentcrawler.com/general-settings/index.html#general-settings-page) , sẽ không có trang web nào được thu thập dữ liệu tự động ngay cả khi tính năng lập lịch được kích hoạt trong [Trang Cài đặt trang web](https://docs.wpcontentcrawler.com/site-settings/index.html#site-settings-page) . Vì vậy, trước tiên hãy bật tính năng lập lịch từ cài đặt chung của plugin.

Trong thanh bên của bảng điều khiển quản trị, hãy nhấp và đảm bảo [Tab Lên lịch](https://docs.wpcontentcrawler.com/general-settings/tabs/scheduling-tab/index.html#tab-gs-scheduling) được bật trong trang đã mở. Tiếp theo, hãy kiểm tra hộp kiểm ( [Lên lịch đang hoạt động?](https://docs.wpcontentcrawler.com/general-settings/tabs/scheduling-tab/scheduling.html#wpcc-is-scheduling-active) ). Khi bạn thực hiện điều đó, hai cài đặt bổ sung sẽ hiển thị:Content Crawler > General SettingsScheduling > Scheduling is active

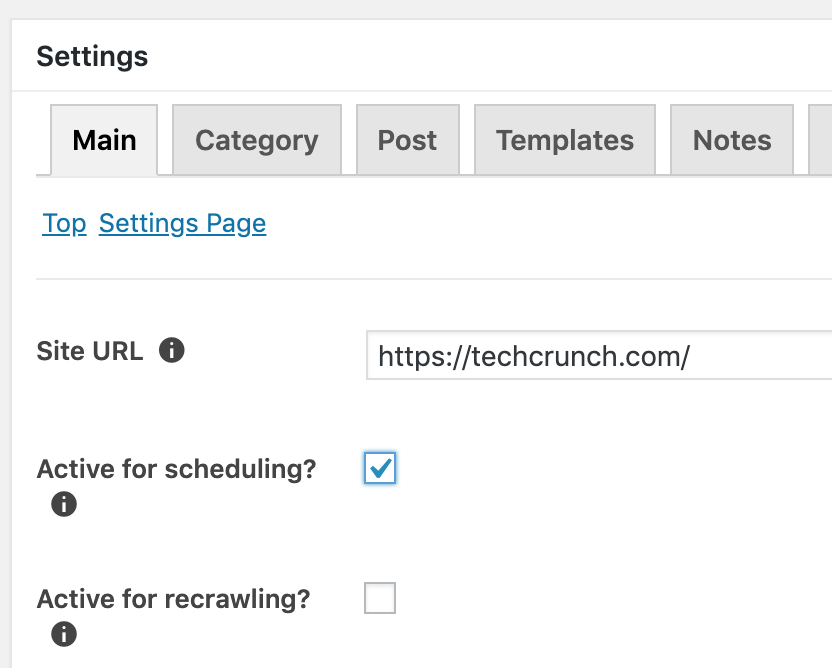
* [Post URL Collection Interval](https://docs.wpcontentcrawler.com/general-settings/tabs/scheduling-tab/scheduling.html#wpcc-interval-url-collection) : Chọn khoảng thời gian để thu thập URL. Ví dụ, nếu bạn chọn, plugin sẽ thu thập một trang của trang danh mục được xác định trong cài đặt cứ sau 5 phút để tìm URL bài đăng mới và lưu chúng vào cơ sở dữ liệu của bạn. Các URL bài đăng được lưu trong cơ sở dữ liệu của bạn được gọi là, nghĩa là các URL bài đăng đang chờ được lưu.Every 5 minutesCategory > Category URLsqueue
* [Post Crawl Interval](https://docs.wpcontentcrawler.com/general-settings/tabs/scheduling-tab/scheduling.html#wpcc-interval-post-crawl) : Chọn khoảng thời gian để thu thập dữ liệu bài đăng. Ví dụ, nếu bạn chọn, plugin sẽ thu thập dữ liệu URL bài đăng cũ nhất trong hàng đợi và lưu vào trang web của bạn dưới dạng bài đăng mới.Every minute

Chúng ta hãy chọn Post [URL Collection Interval](https://docs.wpcontentcrawler.com/general-settings/tabs/scheduling-tab/scheduling.html#wpcc-interval-url-collection) và Post [Crawl Interval](https://docs.wpcontentcrawler.com/general-settings/tabs/scheduling-tab/scheduling.html#wpcc-interval-post-crawl) . Tiếp theo, nhấp vào nút để lưu các thiết lập chung. Sau khi lưu các thay đổi, bạn sẽ thấy trang như trong [Hình 2.22](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-enable-scheduling-general-settings) .Every 5 minutesEvery minuteSave Changes



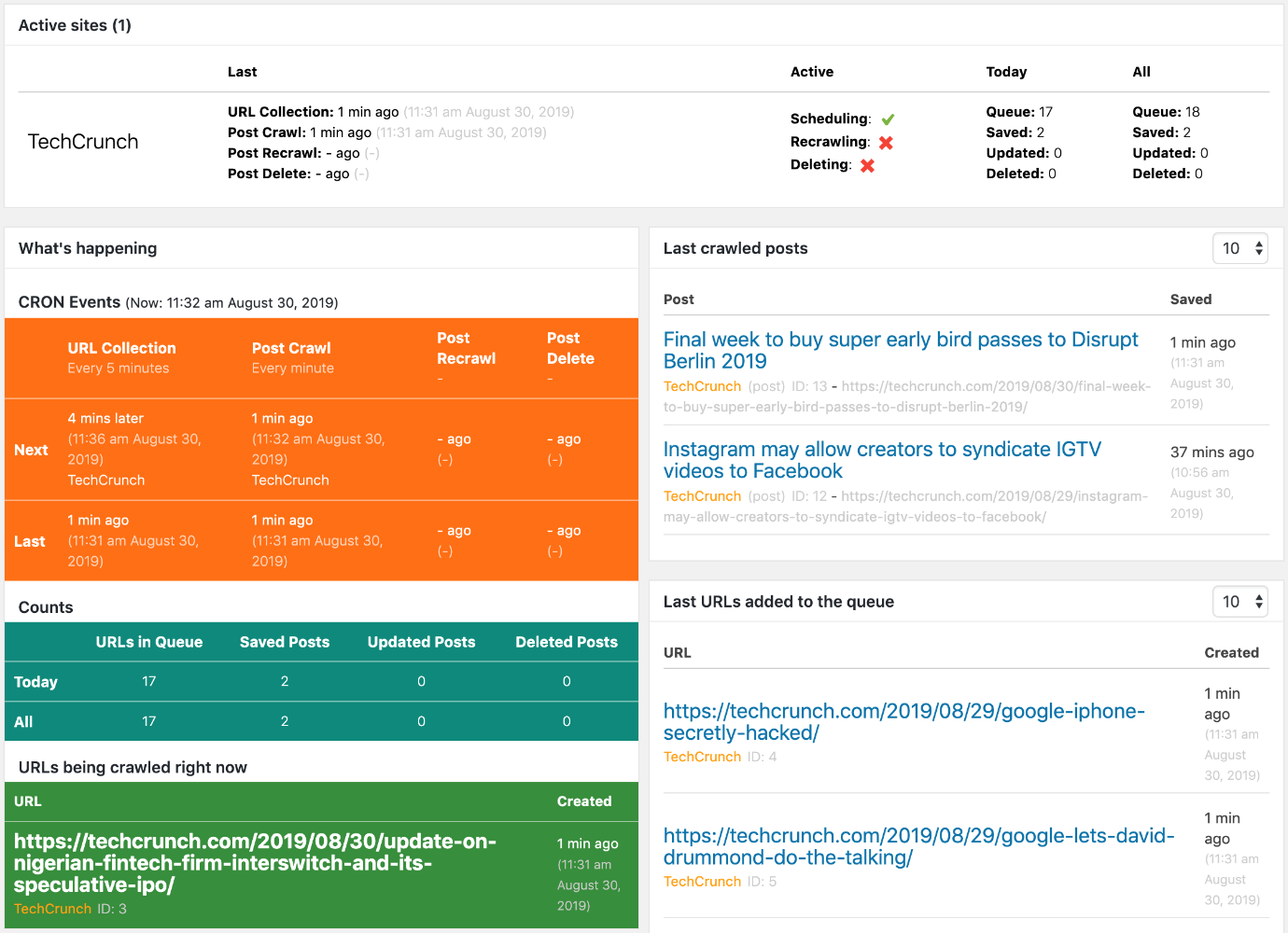
Hình 2.22 *Lên lịch được bật trong*[*Trang Cài đặt chung*](https://docs.wpcontentcrawler.com/general-settings/index.html#general-settings-page)*và các thay đổi được lưu lại.*

Chúng ta cũng hãy bật tính năng lập lịch trong [Trang Cài đặt Trang web](https://docs.wpcontentcrawler.com/site-settings/index.html#site-settings-page) . Trong thanh bên của bảng điều khiển quản trị, hãy nhấp và nhấp vào khi trang được tải. Bây giờ, hãy kiểm tra cài đặt ( [Hình 2.23](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-enable-scheduling-site-settings) , [Hoạt động để lập lịch?](https://docs.wpcontentcrawler.com/site-settings/tabs/main-tab/main.html#active) ) và lưu cài đặt (Xem: [Lưu Cài đặt](https://docs.wpcontentcrawler.com/site-settings/saving-settings.html#saving-settings) ).Content Crawler > All SitesTechCrunchMain > Active for scheduling



Hình 2.23 *Lên lịch được bật trong .Site SettingsTechCrunch*

Bây giờ, từ thanh bên của bảng điều khiển quản trị, hãy nhấp vào ( [Trang bảng điều khiển](https://docs.wpcontentcrawler.com/dashboard/index.html#wp-content-crawler-dashboard) ). Bạn sẽ thấy trong [Phần Trang web đang hoạt động](https://docs.wpcontentcrawler.com/dashboard/index.html#dashboard-page-section-active-sites) . Điều này có nghĩa là bạn đã bật thành công tính năng lập lịch. Sau vài phút, bạn sẽ thấy có một vài URL trong hàng đợi và một vài bài đăng đã được lưu, như được hiển thị trong [Hình 2.24](https://docs.wpcontentcrawler.com/quick-start/crawling-your-first-site/crawling-your-first-site.html#fig-dashboard-after-enabling-scheduling) .Content Crawler > DashboardTechCrunch



Hình 2.24 [*Trang bảng điều khiển*](https://docs.wpcontentcrawler.com/dashboard/index.html#wp-content-crawler-dashboard)*sau khi bật tính năng lập lịch và chờ vài phút*

Miễn là tính năng lập lịch vẫn hoạt động, plugin sẽ tiếp tục thu thập URL bài đăng mới và lưu bài đăng vào trang web của bạn.

**II. Hướng dẫn kết hợp ChatGPT**

**1. Sử dụng ChatGPT**

Plugin WP Content Crawler hỗ trợ tích hợp ChatGPT và các mô hình OpenAI khác để tự động tạo nội dung trong quá trình thu thập dữ liệu từ các trang đích. Dưới đây là các bước thực hiện:

Bước 1: Truy cập cài đặt trang web

* Vào Content Crawler > Add New hoặc chỉnh sửa một trang web đã có sẵn.

Bước 2: Chọn mẫu nội dung

* Chuyển đến một trường mẫu nội dung nơi bạn muốn sử dụng ChatGPT (ví dụ: Nội dung bài viết hoặc Tiêu đề bài viết).
* Nhấp vào nút [openai-gpt] bên cạnh trường mẫu.

Bước 3: Cấu hình chế độ ChatGPT

* Chọn chế độ Chế độ trò chuyện để cho phép AI tạo văn bản dựa trên nội dung thu thập được.
* Tùy chỉnh các tùy chọn như:
  + Prompt (Lời nhắc): Mô tả nội dung bạn muốn AI tạo ra.
  + Nhiệt độ (Temperature): Điều chỉnh mức độ sáng tạo của AI.
  + Ngôn ngữ đầu ra: Chọn ngôn ngữ cho nội dung.

Bước 4: Sao chép mã ngắn

* Sau khi cấu hình, plugin sẽ tạo một đoạn mã ngắn (shortcode).
* Sao chép mã ngắn này.

Bước 5: Chèn mã vào mẫu nội dung

* Dán mã ngắn vừa sao chép vào vị trí phù hợp trong mẫu của bạn.

Bước 6: Lưu cài đặt

* Nhấn Save để lưu cấu hình trang web.

**2. Chuỗi nhắc nhở với ChatGPT**

Chuỗi nhắc (Prompt Chaining) là kỹ thuật giúp chia nhỏ một nhiệm vụ phức tạp thành nhiều nhiệm vụ con đơn giản hơn, giúp cải thiện chất lượng và độ chính xác của nội dung được tạo bởi ChatGPT trong plugin WP Content Crawler.

Lợi ích của Chuỗi nhắc

* Cải thiện chất lượng nội dung.
* Giảm lỗi trong quá trình tạo văn bản.
* Tăng tính linh hoạt trong việc tùy chỉnh nội dung.

Cách tạo Chuỗi nhắc trong WP Content Crawler

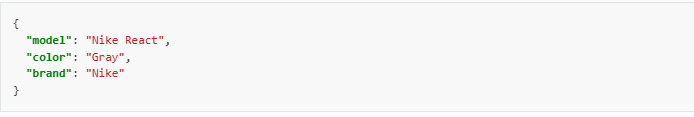
1. Truy cập Cài đặt Plugin
   * Vào Content Crawler > Add New hoặc chỉnh sửa một trang web có sẵn.
2. Chọn Mẫu Nội dung
   * Chuyển đến trường nơi bạn muốn sử dụng ChatGPT.
3. Tạo Chuỗi nhắc
   * Nhấn vào nút [openai-gpt] và chọn Chế độ trò chuyện.
   * Trong phần cấu hình, viết từng bước của nhiệm vụ bằng cách chia nhỏ thành các câu lệnh đơn giản. Ví dụ:
     + Bước 1: Tóm tắt nội dung bài viết.
     + Bước 2: Tạo đoạn giới thiệu hấp dẫn.
     + Bước 3: Viết phần kết thúc với lời kêu gọi hành động.
4. Sao chép Mã ngắn
   * Sao chép từng mã ngắn được tạo ra từ mỗi bước.
5. Chèn vào Mẫu Nội dung
   * Dán các mã ngắn vào mẫu nội dung theo thứ tự bạn muốn.
6. Lưu Cài đặt
   * Nhấn Save Changes để lưu lại cấu hình.

**3. Cấu trúc văn bản với ChatGPT**

“Cấu trúc văn bản” có nghĩa là chuyển đổi văn bản thuần túy thành định dạng có cấu trúc như JSON. Ví dụ, văn bản này:



có thể được chuyển đổi thành JSON sau:



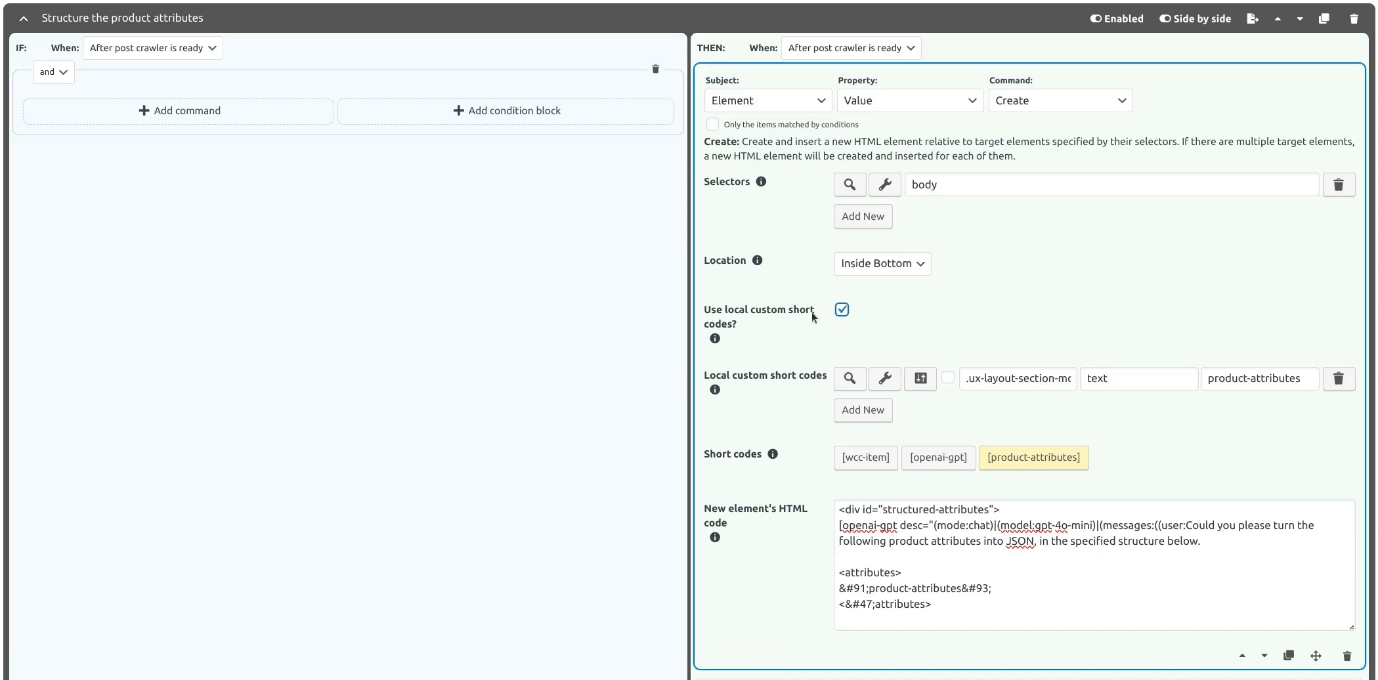
Bằng cách này, có thể trích xuất màu sắc của giày một cách đáng tin cậy, chẳng hạn, chỉ bằng cách lấy brand giá trị trường từ JSON.

Để thực hiện điều này, bạn có thể làm theo các bước sau:

1. Vào [Tab Bài đăng](https://docs.wpcontentcrawler.com/site-settings/tabs/post-tab/index.html#tab-post) > [Phần Bộ lọc](https://docs.wpcontentcrawler.com/site-settings/tabs/post-tab/filters-section.html#section-filters-post) > Cài đặt [bộ lọc trang bài đăng](https://docs.wpcontentcrawler.com/site-settings/tabs/post-tab/filters-section.html#post-page-filters) và thêm bộ lọc mới.
2. Để ifphần đó trống
3. Thêm lệnh mới vào thenphần này và cấu hình như sau:

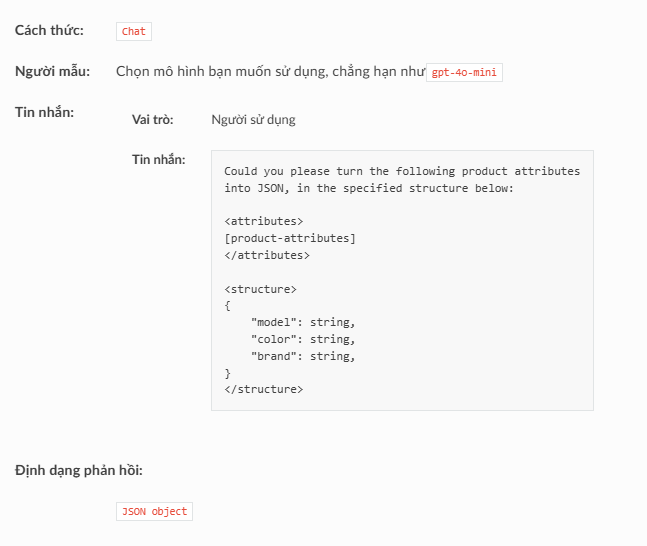


Một ví dụ về cấu hình này có thể được thấy trong 3.1.

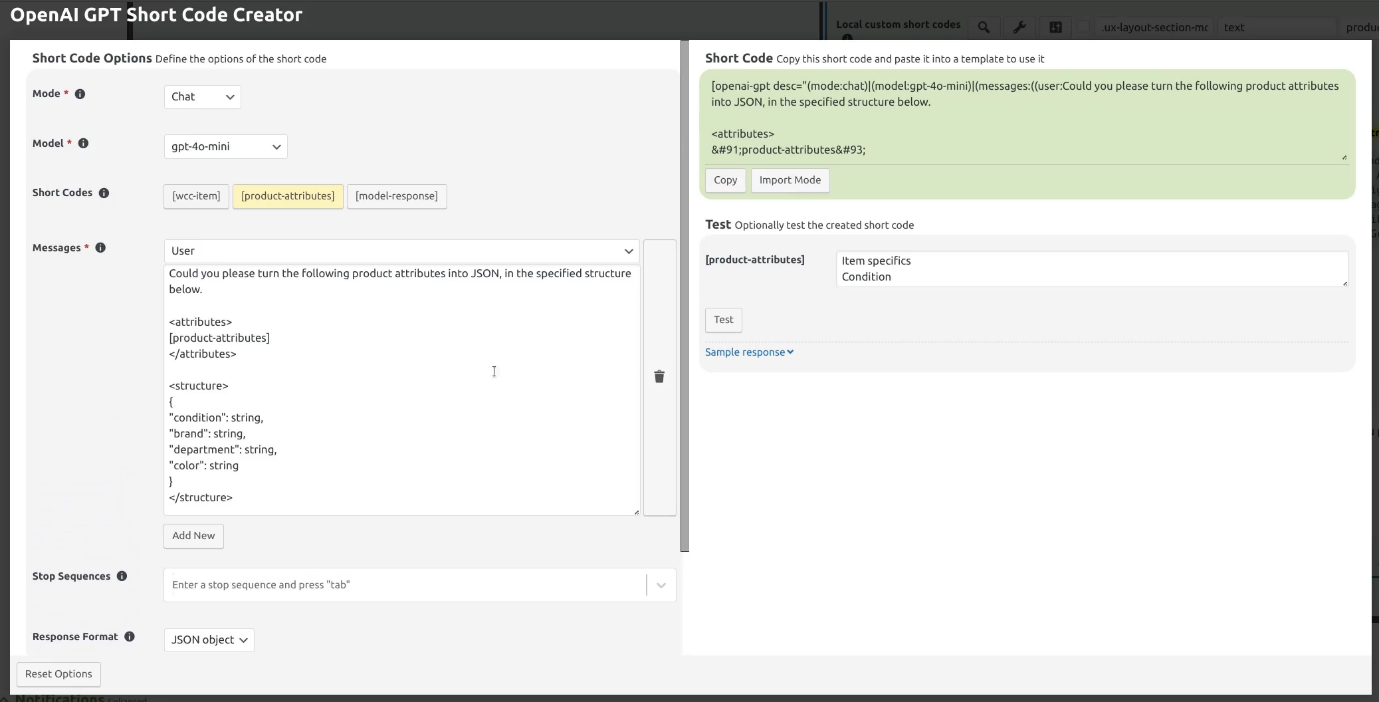


Hình 3.1 *Ví dụ về cấu hình cho lệnh*[*Create*](https://docs.wpcontentcrawler.com/common-options/filter-commands/action-commands/action-commands.html#cmd-element-create)*. Mở hình ảnh trong một tab mới để xem hình ảnh lớn hơn.*

Để tạo một [openai-gpt]mã ngắn, hãy nhấp vào[openai-gpt]nút hiển thị trong tùy chọn của lệnh và cấu hình nó như thế này:Short codes

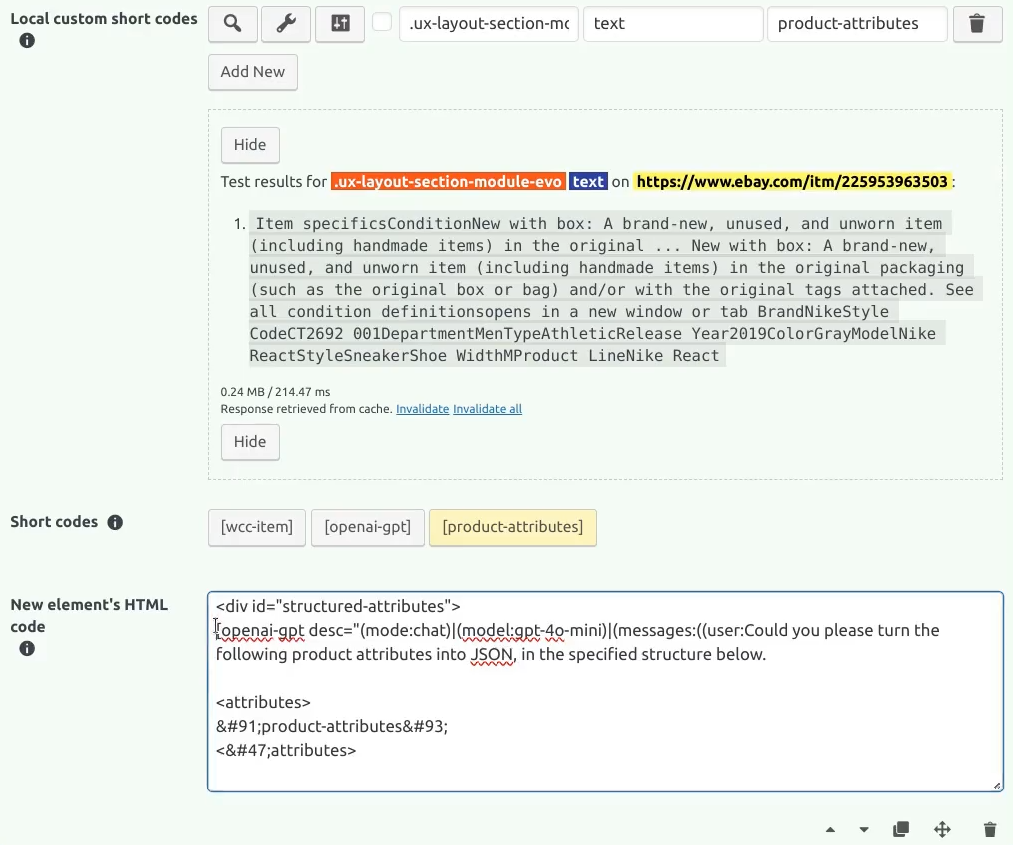


Một ví dụ về cấu hình này có thể được thấy trong hình 3.2 .



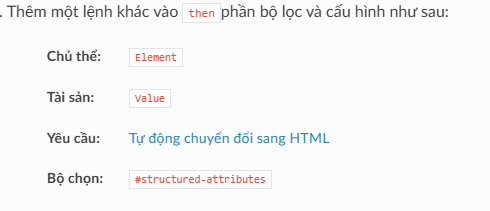
Hình 3.2 *Ví dụ về [openai-gpt]cấu hình mã ngắn để cấu trúc thuộc tính sản phẩm. Mở hình ảnh trong tab mới để xem hình ảnh lớn hơn.*

Sau đó, sao chép mã ngắn bằng cách nhấp vàoSao chéptrong [Phần Mã ngắn](https://docs.wpcontentcrawler.com/common-options/short-codes/openai-gpt/openai-gpt.html#openai-gpt-section-short) và chèn nó vào tùy chọn được hiển thị ở bước trước, như thể hiện trong hình 3.3 .New element's HTML code.

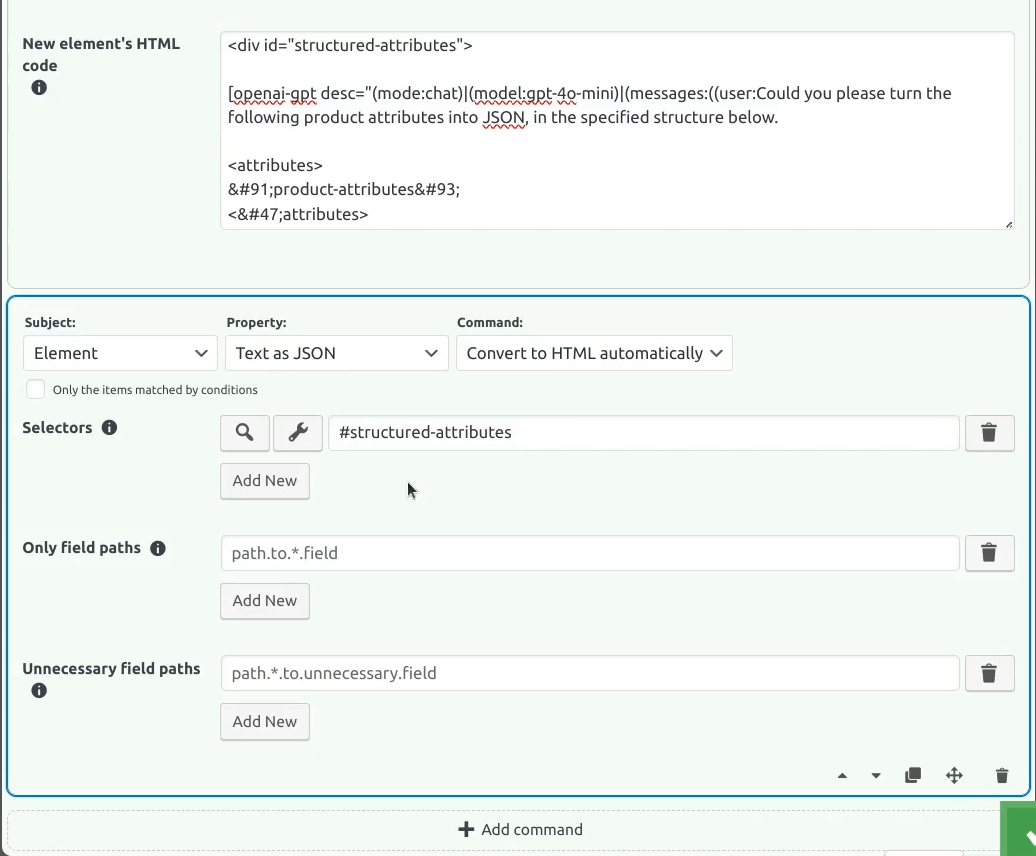


Hình 3.3 *Giá trị ví dụ cho tùy chọn để cấu trúc thuộc tính sản phẩm. Mở hình ảnh trong tab mới để xem lớn hơn.New element's HTML code*

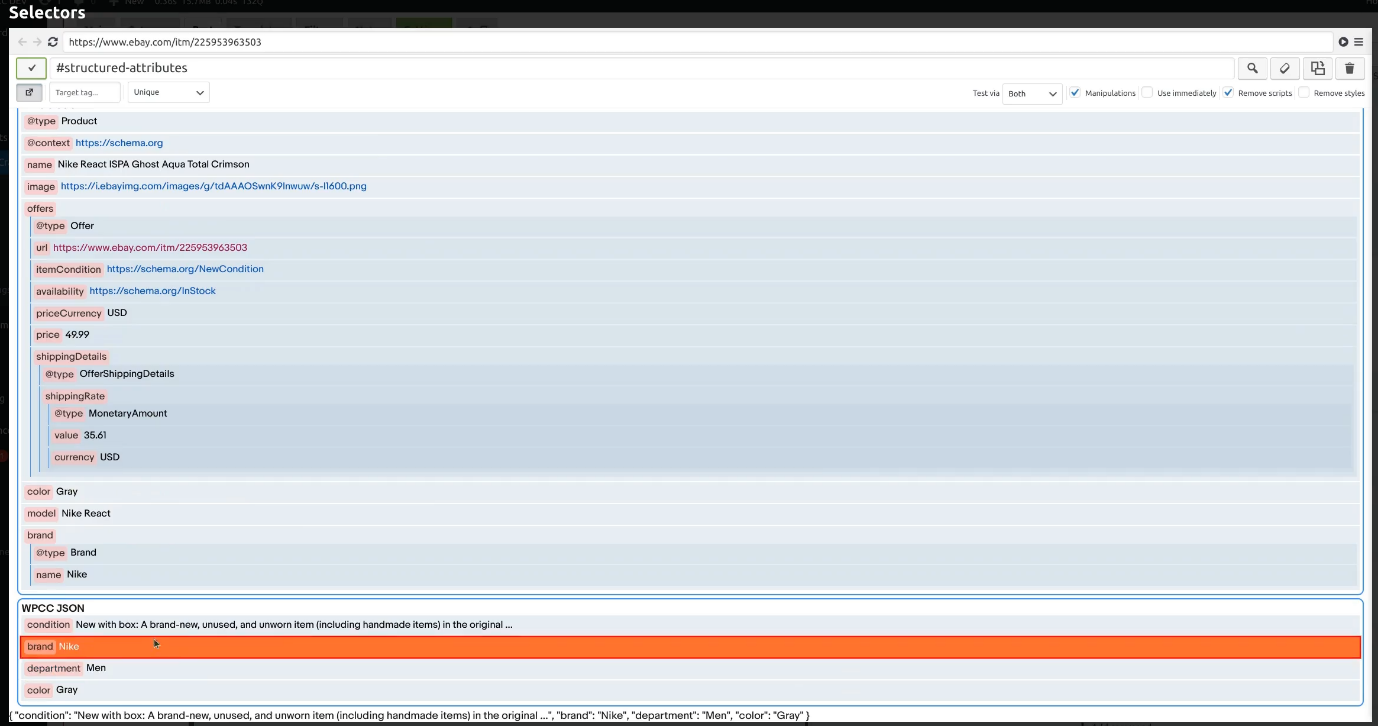
Bây giờ các thuộc tính sản phẩm sẽ được cấu trúc thành định dạng JSON, bạn có thể sử dụng các tính năng phân tích cú pháp JSON của plugin để lấy các giá trị. Ví dụ, bạn có thể tự động chuyển đổi JSON sang HTML và sau đó chọn bất kỳ trường nào giống như bất kỳ phần tử HTML nào khác. Để thực hiện việc này, hãy làm theo các bước sau:



Một ví dụ được thể hiện trong hình 3.4.

Hình 3.4 *Tự động chuyển đổi dữ liệu có cấu trúc thành HTML. Mở hình ảnh trong tab mới để xem lớn hơn.*

Đi đến cài đặt bộ chọn CSS, mở [Visual Inspector](https://docs.wpcontentcrawler.com/site-settings/visual-inspector/visual-inspector.html#visual-inspector) của nó , sau đó làm mới nó. Bạn có thể thấy dữ liệu ở cuối trang, như thể hiện trong hình 3.5 . Bạn có thể chọn chúng để sử dụng, giống như bất kỳ phần tử nào khác.



Hình 3.5 *Chọn các trường JSON được chuyển đổi sang HTML thông qua*[*Visual Inspector*](https://docs.wpcontentcrawler.com/site-settings/visual-inspector/visual-inspector.html#visual-inspector)*. Mở hình ảnh trong một tab mới để xem hình ảnh lớn hơn.*