

**TRƯỜNG ĐẠI HỌC CẦN THƠ**  
**KHOA CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG**



**NIÊN LUẬN CƠ SỞ**  
**NGÀNH KHOA HỌC MÁY TÍNH**

**Đề tài**  
**WEBSITE TIN TỨC BẰNG CƠ CHẾ LẤY TIN TỰ ĐỘNG**  
**CÓ HỖ TRỢ GIỌNG ĐỌC MÁY**

**Giảng viên hướng dẫn:**  
**Th.S Huỳnh Ngọc Thái Anh**

**Sinh viên thực hiện:**  
**Ôn Đình Khang**  
**MSSV: B1812346**  
**Khóa : 44**

**Cần Thơ, 11/2021**

[illegible]

Giảng viên hướng dẫn

## LỜI CẢM ƠN

Để có được bài niên luận này, em xin được bày tỏ lòng biết ơn chân thành và sâu sắc đến Thầy Huỳnh Ngọc Thái Anh – người đã trực tiếp tận tình hướng dẫn, giúp đỡ em. Trong suốt quá trình thực hiện niên luận, nhờ những sự chỉ bảo và hướng dẫn quý giá đó mà bài niên luận này được hoàn thành một cách tốt nhất.

Em cũng xin gửi lời cảm ơn chân thành đến các Thầy Cô Giảng viên Đại học Cần Thơ, đặc biệt là các Thầy Cô ở Khoa CNTT & TT, những người đã truyền đạt những kiến thức quý báu trong thời gian qua.

Em cũng xin chân thành cảm ơn bạn bè cùng với gia đình đã luôn động viên, khích lệ và tạo điều kiện giúp đỡ trong suốt quá trình thực hiện để em có thể hoàn thành bài niên luận một cách tốt nhất.

Tuy có nhiều cố gắng trong quá trình thực hiện niên luận, nhưng không thể tránh khỏi những sai sót. Em rất mong nhận được sự đóng góp ý kiến quý báu của quý Thầy Cô và các bạn để bài niên luận hoàn thiện hơn.

Cần Thơ, ngày 10 tháng 12 năm 2021

Người viết

Ôn Đình Khang

---

## MỤC LỤC

<b>PHẦN GIỚI THIỆU .....</b>	<b>7</b>
1. Đặt vấn đề .....	7
2. Lịch sử giải quyết vấn đề.....	7
3. Mục tiêu đề tài.....	9
4. Đối tượng và phạm vi nghiên cứu .....	10
5. Phương pháp nghiên cứu .....	10
6. Kết quả đạt được.....	10
7. Bố cục .....	11
<b>PHẦN NỘI DUNG .....</b>	<b>12</b>
<b>CHƯƠNG 1: MÔ TẢ HỆ THỐNG.....</b>	<b>12</b>
1. Tổng quan về bài toán .....	12
2. Vấn đề và giải pháp liên quan đến website.....	13
<b>CHƯƠNG 2: THIẾT KẾ VÀ CÀI ĐẶT.....</b>	<b>18</b>
1. Thiết kế hệ thống.....	18
2. Cài đặt hệ thống .....	24
<b>CHƯƠNG 3: KIỂM THỬ VÀ ĐÁNH GIÁ .....</b>	<b>38</b>
1. Kiểm thử .....	38
2. Đánh giá các cơ chế.....	38
<b>CHƯƠNG 4: KẾT QUẢ .....</b>	<b>40</b>
1. Giao diện .....	40
2. Cơ chế lấy tin .....	41
3. Cơ chế giọng đọc .....	41
<b>PHẦN KẾT LUẬN.....</b>	<b>42</b>
1. Kết quả đạt được.....	42

<b>2. Hướng phát triển.....</b>	<b>42</b>
<b>PHẦN TÀI LIỆU THAM KHẢO.....</b>	<b>43</b>

## DANH MỤC HÌNH

Hình 1: Trang chủ VnExpress .....	8
Hình 2: Trang chủ Baomoi.com .....	9
Hình 3: Theme MagazineBook .....	18
Hình 4: Plugin SiteOrigin CSS.....	19
Hình 5: Trang chủ WP Content Crawler .....	20
Hình 6: Plugin ResponsiveVoice Text To Speech .....	22
Hình 7: Plugin Insert Headers and Footers .....	23
Hình 8: Cài đặt WordPress .....	24
Hình 9: Kích hoạt theme MagazineBook .....	25
Hình 10: Bố cục của them Magazine Book.....	25
Hình 11: Cài đặt SiteOrigin CSS.....	26
Hình 12: Custom Css bằng giao diện code.....	26
Hình 13: Custom Css bằng giao diện UI.....	27
Hình 14: Trang chủ Website khi hoàn thành.....	27
Hình 15: Trang chủ Website .....	28
Hình 16: Appearance -> Menu .....	28
Hình 17: Tạo Menu cho trang .....	29
Hình 18: Add New Content Crawler .....	29
Hình 19: Add new Site .....	30
Hình 20: Tab Category .....	31
Hình 21: Tab Post.....	31
Hình 22: Featured Image .....	32
Hình 23: Save Image .....	32

Hình 24:Custom Short Code .....	33
Hình 25:Templates .....	33
Hình 26:Active for scheduling .....	34
Hình 27: ResponsiveVoice Text To Speech.....	34
Hình 28: Key API.....	35
Hình 29:Login ResponsiveVoice .....	35
Hình 30 Cấu hình trang .....	35
Hình 31: Script ResponsiveVoice .....	36
Hình 32: Nhập API Key .....	36
Hình 33: Insert Headers and Footers .....	37
Hình 34: Button giọng đọc .....	37
Hình 35: Single Page .....	40
Hình 36: Danh mục quân sự .....	40
Hình 37: Plugin Content Crawler.....	41
Hình 38: Plugin ResponsiveVoice .....	41

## ABSTRACT

Today, information technology is constantly developing along with the 4.0 technology revolution. It has ended up supporting all fields and industries to develop together. In which Website is an area with a rapid growth rate.

The topic "Building a news website with an automatic news retrieval mechanism that supports machine reading" wishes to build and design a website to help readers capture more and more diverse information from many different sources. Next is building a friendly interface to help readers grasp information quickly and clearly. Finally, the machine's voice helps viewers even if they are busy to hear and understand the article's information.

The website is built on top of WordPress, a free and open source content management system written in PHP programming language, comes with a MySQL database, along with WordPress plugins to build the interface and combine functions on the Web. The mechanism to automatically get news from many other official websites helps to diversify the website's information, update articles quickly and automatically. Use XAMPP as Web Server.



## TÓM TẮT

Ngày nay, công nghệ thông tin đã và đang phát triển không ngừng cùng với cuộc cách mạng công nghệ 4.0 nó đã thúc đẩy tất cả các lĩnh vực và ngành nghề cùng phát triển. Trong đó Website là một lĩnh vực với tốc độ phát triển vượt bậc.

Đề tài **“Xây dựng website tin tức cơ chế lấy tin tự động có hỗ trợ giọng đọc máy”** mong muốn xây dựng và thiết kế nên một trang web giúp ích cho người đọc nắm bắt nhiều thông tin thêm đa dạng từ nhiều nguồn trang website khác nhau. Kế tiếp đó là xây dựng giao diện thân thiện giúp người đọc nắm bắt thông tin một cách nhanh chóng và rõ ràng. Sau cùng là giọng đọc máy giúp người xem dù bận rộn nhưng vẫn có thể nghe được và hiểu được thông tin bài báo.

Website được xây dựng dựa trên WordPress hệ quản trị nội dung mã nguồn mở miễn phí viết bằng ngôn ngữ lập trình PHP đi cùng với cơ sở dữ liệu MySQL, cùng với các plugin của WordPress để xây dựng giao diện và kết hợp các chức năng trên trang web. Cơ chế lấy tin tự động từ nhiều Website chính thống khác giúp đa dạng thông tin của trang web, cập nhật bài viết nhanh chóng và tự động. Sử dụng XAMPP làm Web Server.

## PHẦN GIỚI THIỆU

### 1. Đặt vấn đề

Hiện nay, mặc dù các website tin tức có cơ chế lấy tin tự động ngày càng nhiều và phát triển. Đáp ứng nhu cầu sử dụng các website tin tức với nội dung chính xác, nhanh chóng đem lại thông tin hữu ích cho người sử dụng.

Chúng ta đều biết tin tức có ý nghĩa vô cùng quan trọng với đời sống con người. Để nắm bắt được những thông tin mới nhất về xã hội cũng như các vấn đề nổi bật. Nghiên cứu được tiến hành để xây dựng nên một web tin tức được cập nhật nhanh chóng theo chủ đề của các trang báo lớn khác, cùng với cơ chế đọc tin bằng giọng đọc nhân tạo giúp cho thông tin được cập nhật đến người dùng một cách dễ dàng và chính xác nhất.

Mục tiêu của báo cáo này là sử dụng hệ quản trị nội dung WordPress cùng với một số plugin hỗ trợ khác giúp cho việc lấy tin tự động trở nên nhanh chóng và đơn giản, song song đó là sử dụng một số api có sẵn để tích hợp giọng đọc máy cho bài báo để mọi người có thể tiếp cận dễ đến các tin tức hiện có một cách đơn giản và nhanh chóng.

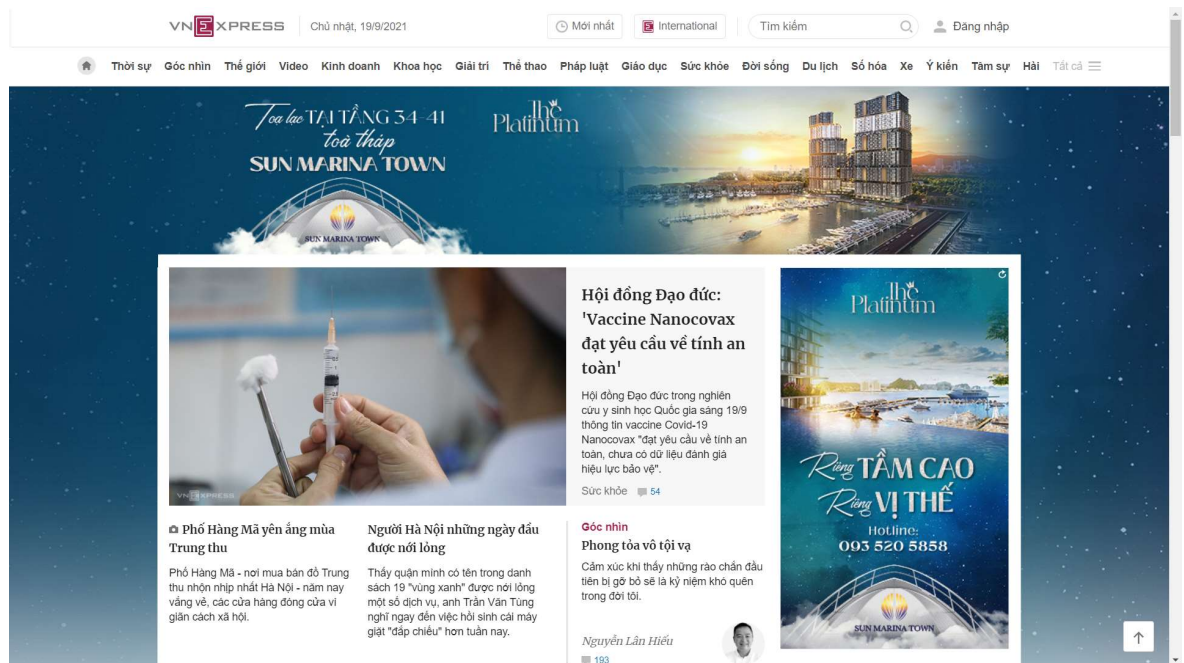
### 2. Lịch sử giải quyết vấn đề

Ngày 6-8-1991 thì trang web đầu tiên được ra đời, kể từ đó đến nay tốc độ phát triển của các website đã tăng lên theo cấp số nhân, tính đến nay đã hơn 1 triệu mũ mười số trang web được đặt trên các máy chủ trên khắp thế giới. Một website cơ bản cấu thành từ ba ngôn ngữ chính là HTML, CSS, Javasripts. Cùng với tốc độ phát triển nhanh như thế thì một số framework được ra đời để hỗ trợ cho việc xây dựng một trang web nhanh, điển hình về phía Front-end chúng ta có React, Angular, Vuejs.. để giúp xây dựng một giao diện nhanh chóng. Về phía Backend chúng ta có Lavaral (PHP), Spring (Java) hay NodeJS ( Javasripts)....

Song song với việc tự xây dựng từ đầu một website thì hiện nay các hệ thống quản trị nội dung (Content Management System – CMS) đã ngày càng phổ biến

giúp cho chúng ta có thể phát triển một website cực kì nhanh chóng mà không cần nhiều kiến thức chuyên sâu về các ngôn ngữ lập trình. Giúp các nhà sáng tạo nội dung có thể dễ dàng chia sẻ ý tưởng của mình đến với mọi người thông qua website.

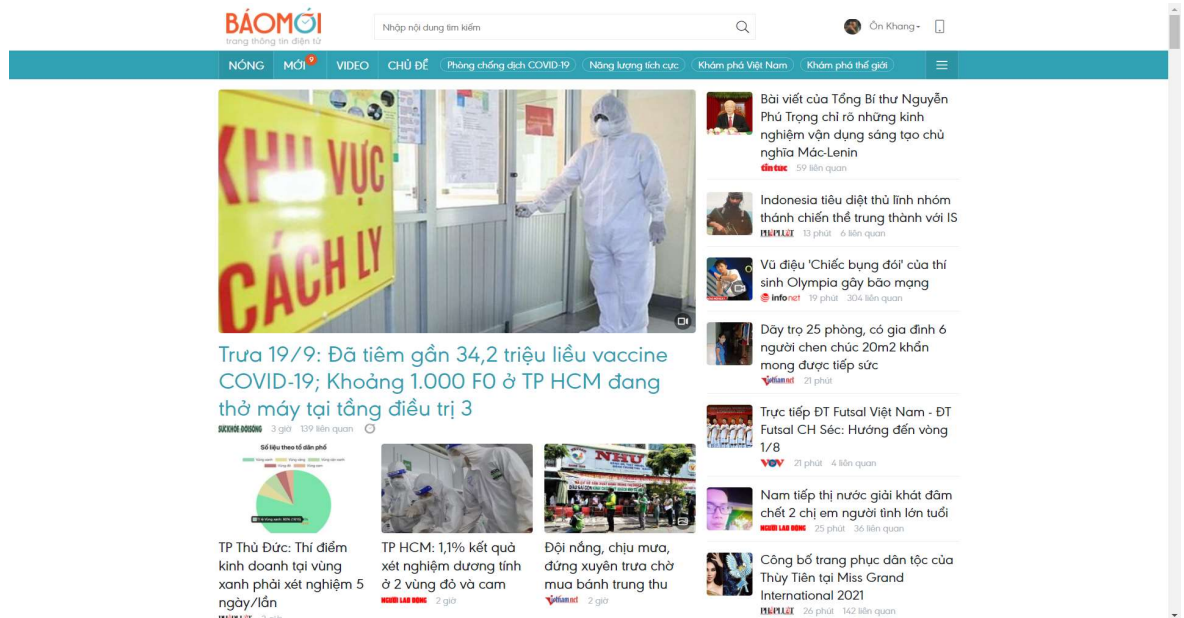
Khi xưa để cập nhật tình hình xã hội, hoặc theo dõi một sự kiện nào đó chúng ta thường phải mất nhiều ngày thậm chí là nhiều tuần để báo đài xuất bản về vấn đề đó, nhưng hiện nay chưa mất tới một vài giờ là chúng ta có thể cập nhật được nội dung như vậy, đó là nhờ các website đưa tin giúp chúng ta có thể tiếp cận thông tin một cách nhanh chóng, điển hình như CNN, BBC..., còn trong nước chúng ta có các trang đưa tin lớn như VnExpress, Dân Trí, Vietnamnet.



Hình 1: Trang chủ VnExpress

Do các trang đưa tin hiện nay đa phần cung cấp cho chúng ta thông tin trên nhiều lĩnh vực khác nhau, nên nếu chúng ta muốn quan tâm một số vấn đề theo sở thích cá nhân sẽ khó tìm được nguồn phù hợp với sở thích, vì thế các kỹ thuật crawl dữ liệu là phương án tốt nhất để chúng ta có thể tùy chỉnh nội dung cần lấy phù hợp với yêu cầu riêng của chính ta.

Diễn hình trang Baomoi.com sử dụng kỹ thuật crawl RSS để tổng hợp các bài đăng nổi bật từ các trang tin tức khác về hiển thị, sau khi người dùng truy cập vào các nội dung đó, thì trang sẽ chuyển hướng đến trang chính của bài báo.



Hình 2: Trang chủ Baomoi.com

Tóm lại kỹ thuật crawler là kỹ thuật thu thập dữ liệu từ các đường links cho trước trên các website trên mạng, từ đó những dữ liệu khi crawl có thể được lưu trữ trong cơ sở dữ liệu của bạn để phục vụ việc phân tích hoặc sử dụng với mục đích khác nhau. Hoặc có thể show trực tiếp ra trang web như các web tin tức, dự báo thời tiết,...

### 3. Mục tiêu đề tài

Sử dụng công nghệ Wordpress để xây dựng một website, trong đó nội dung chính của website là về thông tin các tin tức mới trong ngày, được cập nhật một cách tự động, kèm theo cơ chế giọng đọc giúp tạo ra trang báo nói.

Các chức năng chính:

- Admin có quyền đăng, sửa, xóa bài viết.
- Tự động cập nhật các thông tin nổi bật từ các nguồn tin khác.
- Tự động cập nhật thông tin các , chương trình từ các nguồn có sẵn.
- Tự động đọc trang báo khi người dùng kích hoạt.

#### 4. Đối tượng và phạm vi nghiên cứu

Cách thức hoạt động của một website.

Nghiên cứu cách hoạt động của hệ quản trị nội dung Wordpress.

Nghiên cứu về UI/UX khi xây dựng một website tin tức.

Các ngôn ngữ lập trình web:

- Front-end: HTML, CSS, Javascripts.
- Back-end: Php.
- Hệ quản trị cơ sở dữ liệu MySQL.

Một số kỹ thuật crawl dữ liệu, plugin hỗ trợ.

Phạm vi nghiên cứu: xây dựng một website tin tức.

#### 5. Phương pháp nghiên cứu

Tìm hiểu cách thức hoạt động của Wordpress:

- Cài đặt Wordpress.
- Tạo một website demo.
- Cài đặt theme giao diện phù hợp với chủ đề.
- Tìm hiểu các plugin hỗ trợ việc xây dựng trang tin tức.
- Tinh chỉnh css.

Nghiên cứu về UI/UX để tinh chỉnh giao diện.

Nghiên cứu PHP để chỉnh sửa các theme cho phù hợp với yêu cầu.

Tìm hiểu các kỹ thuật crawl dữ liệu hiện nay, sử dụng một số plugin để tạo một tính năng tự động cập nhật tin tức.

#### 6. Kết quả đạt được

Xây dựng được một website tin tức bằng WordPress. Sử dụng các Plugin trong Wordpress thực hiện cơ chế lấy tin tự động từ nhiều nguồn Website khác, sử dụng api từ ResponsiveVoice để xây dựng tính năng giọng đọc máy và sử dụng theme giao diện để xây dựng một giao diện thân thiện và dễ sử dụng cho người dùng.

## **7. Bố cục**

### **Phần giới thiệu:**

Giới thiệu tổng quát về đề tài.

### **Phần nội dung:**

**Chương 1:** Mô tả hệ thống

**Chương 2:** Thiết kế và cài đặt

**Chương 3:** Kiểm thử và đánh giá

**Chương 4:** Kết quả

### **Phần kết luận:**

Trình bày kết quả đạt được và hướng phát triển website.

### **Phần tài liệu tham khảo:**

Trình bày các tài liệu tham khảo.

## PHẦN NỘI DUNG

### CHƯƠNG 1: MÔ TẢ HỆ THỐNG

#### 1. Tổng quan về bài toán

##### 1.1. Khái niệm về tin tức

Tin tức là thông tin về các sự kiện hiện tại. Điều này có thể được cung cấp thông qua nhiều phương tiện khác nhau: truyền miệng, in ấn, hệ thống bưu chính, phát thanh truyền hình, thông tin điện tử ,...

##### 1.2. Khái niệm về website tin tức

Đáp ứng nhu cầu nắm bắt thông tin của mọi người, đặc biệt các tin tức mới nhất của trong nước và thế giới. Website tin tức kết hợp giữa truyền đạt thông tin tin tức và kết hợp sự đa dạng thông tin từ nhiều nguồn website khác nhau để xây dựng hệ thống website.

##### 1.3. Mô tả hệ thống website

Việc xây dựng website tin tức kết hợp thông tin từ nhiều trang website tin tức khác nhau tạo nên sự đa dạng thông tin và nội dung chính thống từ các trang website lớn. Sử dụng công cụ Wordpress là một phần mềm mã nguồn mở miễn phí được nhiều cá nhân và doanh nghiệp sử dụng để xây dựng một trang web, blog hoặc tin tức cho chính cá nhân và doanh nghiệp đấy.

Website tin tức trong đề tài với cơ chế lấy tin tự động được xây dựng bằng phần mềm WordPress với các Plugin miễn phí giúp tạo nên một trang website tin tức đa dạng thông tin và các tính năng của một trang website phù hợp cho mọi người sử dụng.

## 2. Vấn đề và giải pháp liên quan đến website

### 2.1. Xampp

XAMPP là một loại phần mềm khá phổ biến và thường được các lập trình viên sử dụng để xây dựng và phát triển các dự án website theo ngôn ngữ PHP. XAMPP thường được dùng để tạo máy chủ web được tích hợp sẵn PHP, MySQL, Apache, FTP Server, Mail Server và các công cụ như phpMyAdmin.

Đặc biệt thì Xampp có giao diện quản lý khá tiện lợi cho phép người dùng chủ động bật tắt hoặc khởi động lại các dịch vụ máy chủ bất kỳ lúc nào. Phần mềm này cũng được thiết lập và phát triển dựa trên mã nguồn mở.

Nhìn chung thì XAMPP được xem là một bộ công cụ hoàn chỉnh dành cho lập trình viên PHP trong việc thiết lập và phát triển các website, chúng tích hợp các thành phần quan trọng và tương thích lẫn nhau như:

- **Apache:** Web Server mã nguồn mở Apache là máy chủ được sử dụng rộng rãi nhất trên toàn thế giới để phân phối nội dung Web. Ứng dụng được cung cấp dưới dạng phần mềm miễn phí bởi Apache Software Foundation.
- **MySQL / MariaDB:** Trong MySQL, XAMPP chứa một trong những hệ quản trị cơ sở dữ liệu quan hệ phổ biến nhất trên thế giới. Kết hợp với Web Server Apache và ngôn ngữ lập trình PHP, MySQL cung cấp khả năng lưu trữ dữ liệu cho các dịch vụ Web. Các phiên bản XAMPP hiện tại đã thay thế MySQL bằng MariaDB (một nhánh của dự án MySQL do cộng đồng phát triển, được thực hiện bởi các nhà phát triển ban đầu).
- **PHP:** Ngôn ngữ lập trình phía máy chủ PHP cho phép người dùng tạo các trang Web hoặc ứng dụng động. PHP có thể được cài đặt trên tất cả các nền tảng và hỗ trợ một số hệ thống cơ sở dữ liệu đa dạng.
- **Perl:** ngôn ngữ kịch bản Perl được sử dụng trong quản trị hệ thống, phát triển Web và lập trình mạng. Giống như PHP, Perl cũng cho phép người dùng lập trình các ứng dụng Web động.

#### Ưu điểm:

- XAMPP có thể chạy được trên tất cả các hệ điều hành: Từ Cross-platform, Window, MacOS và Linux.
- XAMPP có cấu hình đơn giản cũng như nhiều chức năng hữu ích cho người dùng. Tiêu biểu gồm: giả lập Server, giả lập Mail Server, hỗ trợ SSL trên Localhost.



- XAMPP tích hợp nhiều thành phần với các tính năng:
- Apache
- PHP (tạo môi trường chạy các tập tin script \*.php);
- MySQL: thay vì phải cài đặt từng thành phần trên, giờ đây các bạn chỉ cần cài XAMPP là chúng ta có 1 web server hoàn chỉnh.
- **Mã nguồn mở:** Không như **Appserv**, **XAMPP** có giao diện quản lý khá tiện lợi. Nhờ đó, người dùng có thể chủ động bật tắt hoặc khởi động lại các dịch vụ máy chủ bất kỳ lúc nào.

**Nhược điểm:**

- Không hỗ trợ cấu hình Module.
- Không có Version MySQL.
- Dung lượng của XAMPP cũng tương đối nặng, dung lượng file cài đặt của XAMPP là 141Mb, nặng hơn nhiều so với WAMP chỉ 41Mb

**2.2. Hệ quản trị nội dung Wordpress**

WordPress là một mã nguồn mở bằng ngôn ngữ PHP để hỗ trợ tạo blog cá nhân, và nó được rất nhiều người sử dụng ủng hộ về tính dễ sử dụng, nhiều tính năng hữu ích. Qua thời gian, số lượng người sử dụng càng ngày càng tăng lên, những lập trình viên cũng tham gia đông đảo để phát triển mã nguồn WordPress để xây dựng, phát triển thêm những tính năng vô cùng tuyệt vời.

Và cho đến thời điểm hiện tại, WP đã được xem như là một hệ quản trị nội dung (CMS – Content Management System) vượt trội để hỗ trợ người dùng tạo ra nhiều thể loại website khác nhau như blog cá nhân, website tin tức/tạp chí, web quảng bá doanh nghiệp, web bán hàng – thương mại điện tử, thậm chí với các loại website có độ phức tạp cao như đặt phòng khách sạn, thuê xe, đăng dự án bất động sản,... Hầu như mọi hình thức website với quy mô nhỏ và vừa đều có thể triển khai trên nền tảng WordPress.

Tuy nhiên, như thế không đồng nghĩa với việc WordPress chỉ thích hợp với các dự án quy mô nhỏ mà vừa, hiện nay, có tới khoảng 28% website trong danh sách 100 website lớn nhất thế giới chọn sử dụng mã nguồn wordpress này. Như tạp chí

TechCrunch, MTV News, Bata, Quartz, CNN, BBC America, Sony Music, ... và rất nhiều website khác.

**b2/cafeblog** thường được biết đến với cái tên đơn giản hơn là b2 hay cafeblog là tiền thân của WordPress. Nó cũng được viết bằng ngôn ngữ lập trình PHP để dùng với MySQL bởi Michel Valdrighi, người đã trở thành nhà phát triển chính của WordPress hiện nay.

### **Ưu điểm:**

- Dễ sử dụng: Dùng WordPress khi thiết kế trang web là dễ sử dụng. Thao tác đơn giản rất đơn giản, dễ hiểu và dễ vận hành nên người sử dụng không cần biết kiến thức lập trình nâng cao.
- Dễ quản lý: Hệ thống quản trị rất dễ dàng, tất cả các mục như bài đăng, giao diện, cài đặt, ... được sắp xếp dễ hiểu, khoa học và hợp lý cho người mới bắt đầu sử dụng.
- Nguồn mở miễn phí: Thiết kế trang web miễn phí.
- Tối ưu hóa SEO: Có các công cụ mặc định để giúp SEO trang web dễ dàng hơn và nhanh hơn.
- Thiết kế trang web Mã nguồn mở hỗ trợ người dùng thiết kế trang web nhanh không cần phải biết quá nhiều kiến thức lập trình.
- Hỗ trợ nhiều ngôn ngữ WordPress hỗ trợ 52 ngôn ngữ trong đó có tiếng Việt. Vì vậy người dùng có thể dễ dàng làm việc và thao tác trên đó.

### **Nhược điểm:**

- Bảo mật không tốt: Sự phổ biến của CMS WordPress khiến các trang web dễ dàng bị tấn công bởi những kẻ phá hoại.
- Cài đặt template và plugin: Cài đặt không đơn giản nếu hoạt động không đúng cách có thể dẫn đến nhiều xung đột xảy ra khi sử dụng.
- Phù hợp với doanh nghiệp nhỏ: Chỉ phù hợp với doanh nghiệp vừa và nhỏ. Đây không phải là giải pháp tối ưu đối với các công ty lớn.

### 2.3. Ngôn ngữ hỗ trợ

- **MySQL** là hệ quản trị cơ sở dữ liệu tự do nguồn mở phổ biến nhất thế giới và được các nhà phát triển rất ưa chuộng trong quá trình phát triển ứng dụng. Vì MySQL là hệ quản trị cơ sở dữ liệu tốc độ cao, ổn định và dễ sử dụng, có tính khả chuyên, hoạt động trên nhiều hệ điều hành cung cấp một hệ thống lớn các hàm tiện ích rất mạnh. Với tốc độ và tính bảo mật cao, MySQL rất thích hợp cho các ứng dụng có truy cập CSDL trên internet. Người dùng có thể tải về MySQL miễn phí từ trang chủ. MySQL có nhiều phiên bản cho các hệ điều hành khác nhau: phiên bản Win32 cho các hệ điều hành dòng Windows, Linux, Mac OS X, Unix, FreeBSD, NetBSD, Novell NetWare, SGI Irix, Solaris, SunOS,...

- **PHP: Hypertext Preprocessor**, thường được viết tắt thành **PHP** là một ngôn ngữ lập trình kịch bản hay một loại mã lệnh chủ yếu được dùng để phát triển các ứng dụng viết cho máy chủ, mã nguồn mở, dùng cho mục đích tổng quát. Nó rất thích hợp với web và có thể dễ dàng nhúng vào trang HTML. Do được tối ưu hóa cho các ứng dụng web, tốc độ nhanh, nhỏ gọn, cú pháp giống C và Java, dễ học và thời gian xây dựng sản phẩm tương đối ngắn hơn so với các ngôn ngữ khác nên PHP đã nhanh chóng trở thành một ngôn ngữ lập trình web phổ biến nhất thế giới.

#### - Bộ ba ngôn ngữ lập trình web

+ **HTML** là viết tắt của cụm từ **Hypertext Markup Language** (tạm dịch là Ngôn ngữ đánh dấu siêu văn bản). **HTML** được sử dụng để tạo và cấu trúc các thành phần trong trang web hoặc ứng dụng, phân chia các đoạn văn, heading, titles, blockquotes... và **HTML** không phải là ngôn ngữ lập trình.

+ **CSS** là chữ viết tắt của Cascading Style Sheets, nó là một ngôn ngữ được sử dụng để tìm và định dạng lại các phần tử được tạo ra bởi các ngôn ngữ đánh dấu (HTML). Nói ngắn gọn hơn là ngôn ngữ tạo phong cách cho trang web. Bạn có thể hiểu đơn giản rằng, nếu HTML đóng vai trò định dạng các phần tử trên website như việc tạo ra các đoạn văn bản, các tiêu đề, bảng,...thì CSS sẽ giúp chúng ta có thể thêm style vào các phần tử HTML đó như đổi bố cục, màu sắc trang, đổi màu chữ, font chữ, thay đổi cấu trúc...

+ **JavaScript**, theo phiên bản hiện hành, là một ngôn ngữ lập trình thông dịch được phát triển từ các ý niệm nguyên mẫu. Ngôn ngữ này được dùng rộng rãi cho các trang web (phía người dùng) cũng như phía máy chủ (với Nodejs). Nó vốn được phát triển bởi Brendan Eich tại Hãng truyền thông Netscape với cái tên đầu tiên Mocha, rồi sau đó đổi tên thành LiveScript, và cuối cùng thành JavaScript. Giống Java, JavaScript có cú pháp tương tự C, nhưng nó gần với Self hơn Java. .js là phần mở rộng thường được dùng cho tập tin mã nguồn JavaScript.

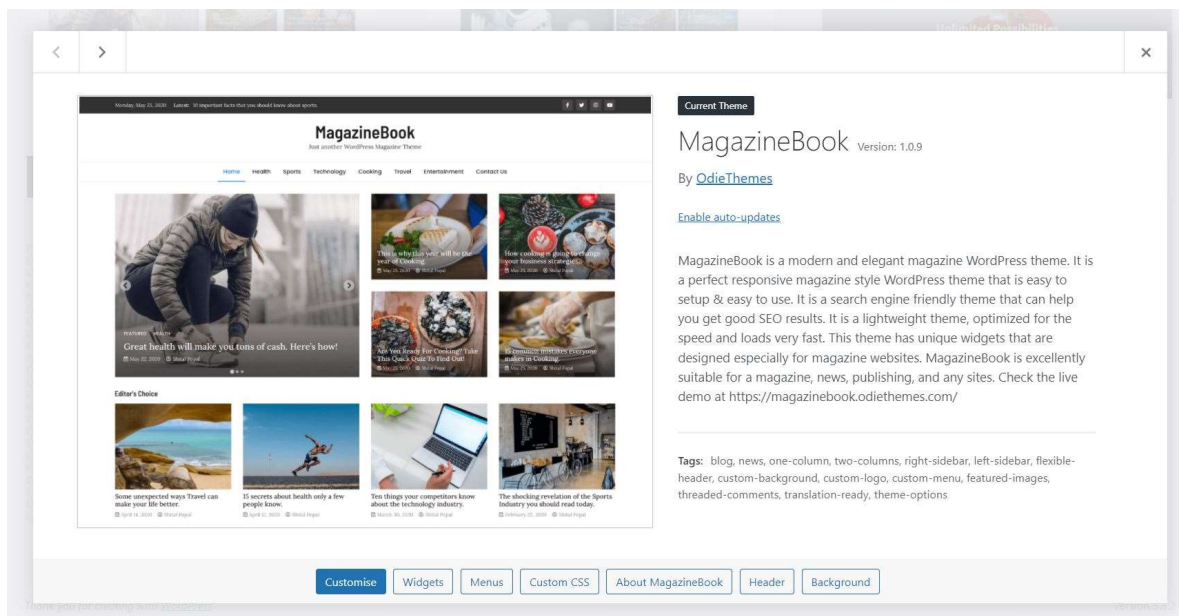
## CHƯƠNG 2: THIẾT KẾ VÀ CÀI ĐẶT

### 1. Thiết kế hệ thống

#### 1.1. Xây dựng chủ đề

##### 1.1.1. Theme MagazineBook

**MagazineBook** là một chủ đề WordPress tạp chí hiện đại và thanh lịch. Đây là một chủ đề WordPress theo phong cách tạp chí đáp ứng hoàn hảo, dễ cài đặt và dễ sử dụng. Đây là một chủ đề thân thiện với công cụ tìm kiếm có thể giúp bạn có được kết quả SEO tốt. Đây là một chủ đề nhẹ, được tối ưu hóa cho tốc độ và tải rất nhanh. Chủ đề này có các widget độc đáo được thiết kế đặc biệt cho các trang web tạp chí.



Hình 3: Theme MagazineBook

##### 1.1.2. Đặc điểm

###### Phiên bản miễn phí:

- Không tối ưu SEO
- Không có dịch vụ chăm sóc khách hàng
- Bị giới hạn tính năng

###### Phiên bản trả phí:

- Thao tác cài đặt nhanh chóng đơn giản
- Tối ưu hóa SEO

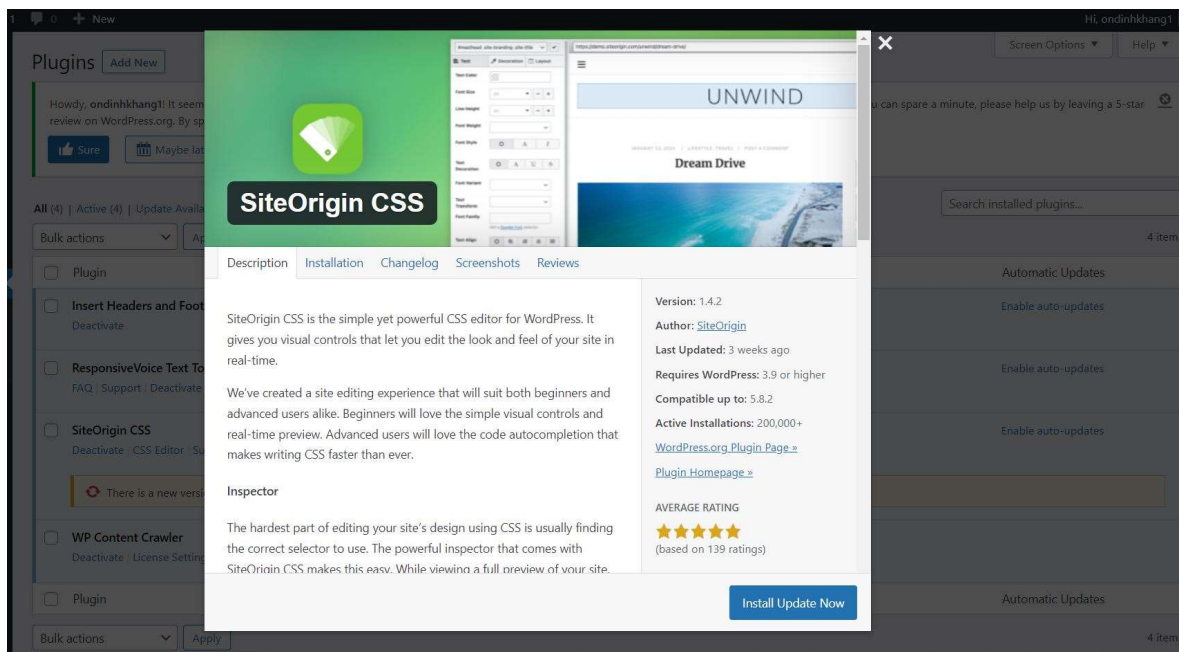
- Phát triển bảo trì tiện lợi
- Cấu trúc rõ ràng, dễ quản lý.

## 1.2. Tùy chỉnh giao diện

### 1.2.1. Plugin SiteOrigin Css

SiteOrigin CSS là trình chỉnh sửa CSS đơn giản nhưng mạnh mẽ cho WordPress. Nó cung cấp cho bạn các điều khiển trực quan cho phép bạn chỉnh sửa giao diện trang web của mình trong thời gian thực.

Plugin hỗ trợ chỉnh sửa trang web phù hợp với cả người mới bắt đầu và người dùng nâng cao. Người mới bắt đầu sẽ thích các điều khiển trực quan đơn giản và xem trước thời gian thực. Người dùng nâng cao sẽ thích tính năng tự động hoàn thành mã giúp viết CSS nhanh hơn.



Hình 4: Plugin SiteOrigin CSS

Ưu điểm:

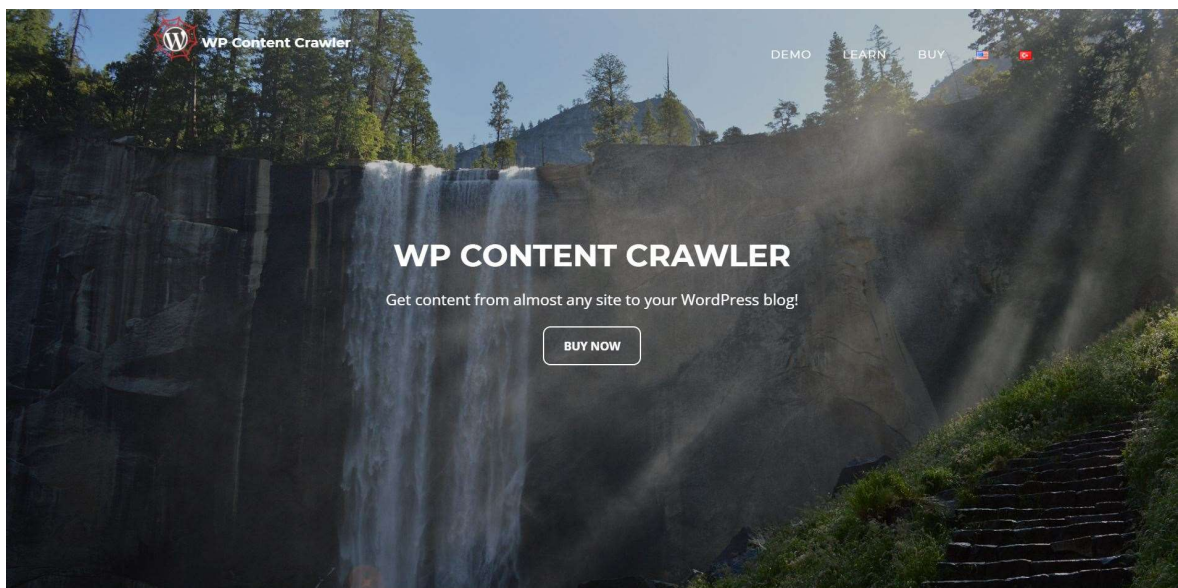
- Plugin miễn phí.
- Tạo Css nhanh.
- Không cần nhiều kiến thức về Css để sử dụng.
- Có cơ chế chỉnh sửa css bằng giao diện bằng UI.

- Có cơ chế chỉnh sửa giao diện bằng code.

### 1.3. Cơ chế lấy tin

#### 1.3.1. Plugin WP Content Crawler

WordPress Content Crawler là một plugin WordPress có thể thu thập thông tin (quét, lấy, truy xuất) nội dung từ hầu hết mọi trang web hiện có trên Internet. Nó sử dụng các bộ chọn CSS để định vị và truy xuất nội dung trong mã nguồn của trang web đích. Các bộ chọn CSS rất dễ học và rất hữu ích để định vị thông tin trong trang web đích. Plugin cũng đi kèm với Trình kiểm tra trực quan mà bạn có thể sử dụng để nhấp vào một phần tử để tìm bộ chọn CSS của nó. Công cụ này cũng có khả năng tìm một bộ chọn CSS để tìm các mục tương tự như mục bạn nhấp vào. Tính năng này khá hữu ích nếu bạn muốn truy xuất



Hình 5: Trang chủ WP Content Crawler

#### Các chức năng chính của plugin:

- **Lưu mọi chi tiết bài đăng:** Tiêu đề, đoạn trích, nội dung, thẻ, danh mục, ngày tháng, meta tùy chỉnh, phân loại, từ khóa meta, mô tả meta, hình ảnh nổi bật, hình ảnh bài đăng, trạng thái... Tất cả mọi thứ.
- **Thu thập thông tin (quét, lấy, lưu) bài đăng:** Sau khi cài đặt được định cấu hình, plugin sẽ tìm URL của các bài đăng và tự động thu thập dữ liệu chúng trong nền.

- **Xóa bài viết:** Bạn muốn xóa các bài đăng cũ đã được thu thập thông tin? Plugin có thể xóa chúng tự động.

- **Lưu danh mục:** Danh mục mục tiêu không tồn tại trong trang web của bạn? Không vấn đề gì. Plugin có thể tạo các danh mục mục tiêu cho bạn. Chỉ cần xác định các bộ chọn CSS để tìm tên danh mục. Chúng thậm chí có thể được tạo dưới dạng danh mục con.

### 1.3.2. Đặc điểm

#### Ưu điểm:

- Tài liệu hướng dẫn chi tiết..
- Dễ dàng sử dụng
- Tự động crawl dữ liệu theo lịch cài đặt
- Không giới hạn nội dung
- Lưu được nội dung đã crawl trên server

#### Nhược điểm:

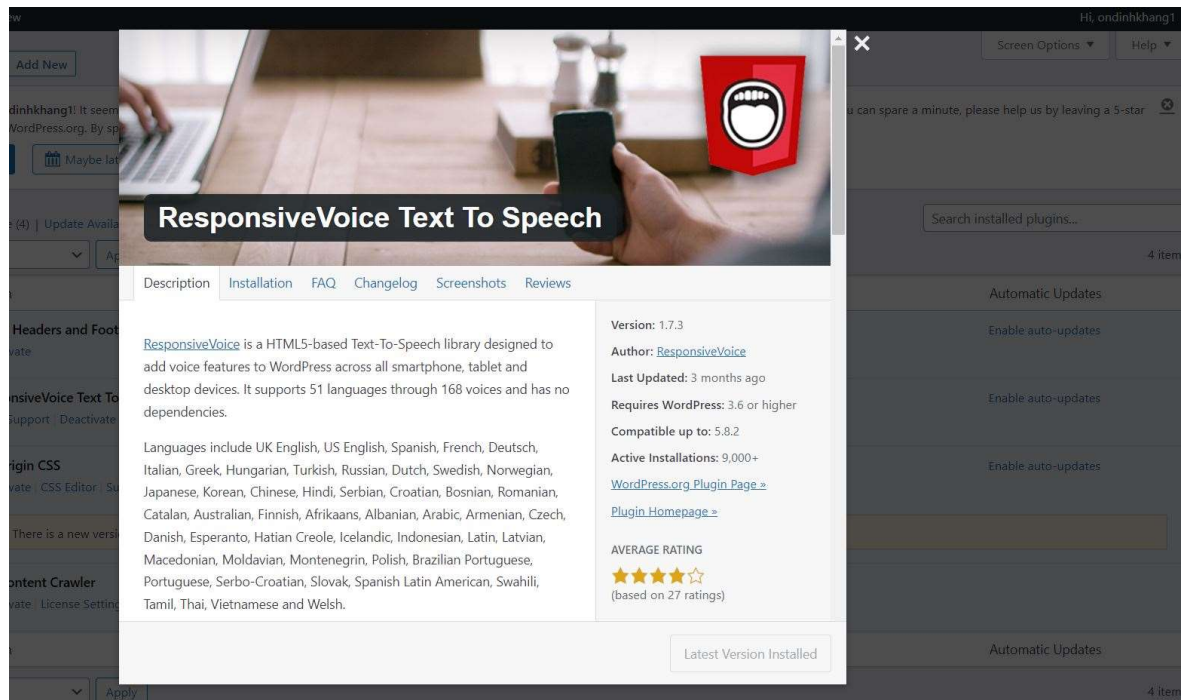
- Một số trang không thể crawl được hình ảnh
- Lỗi hiển thị bố cục ở một số theme
- Không tương thích với một số theme

## 1.4. Cơ chế chuyển văn bản thành giọng nói

### 1.4.1. Plugin ResponsiveVoice Text To Speech

ResponsiveVoice là một thư viện Text-To-Speech dựa trên HTML5 được thiết kế để thêm các tính năng giọng nói vào WordPress trên tất cả các thiết bị điện thoại thông minh, máy tính bảng và máy tính để bàn. Nó hỗ trợ 51 ngôn ngữ thông qua 168 giọng nói.





Hình 6: Plugin ResponsiveVoice Text To Speech

#### 1.4.2. Đặc điểm

##### Ưu điểm:

- Hoàn toàn miễn phí.
- Sử dụng đơn giản.
- Tài liệu hướng dẫn nhiều.
- Hỗ trợ nhiều loại ngôn ngữ.

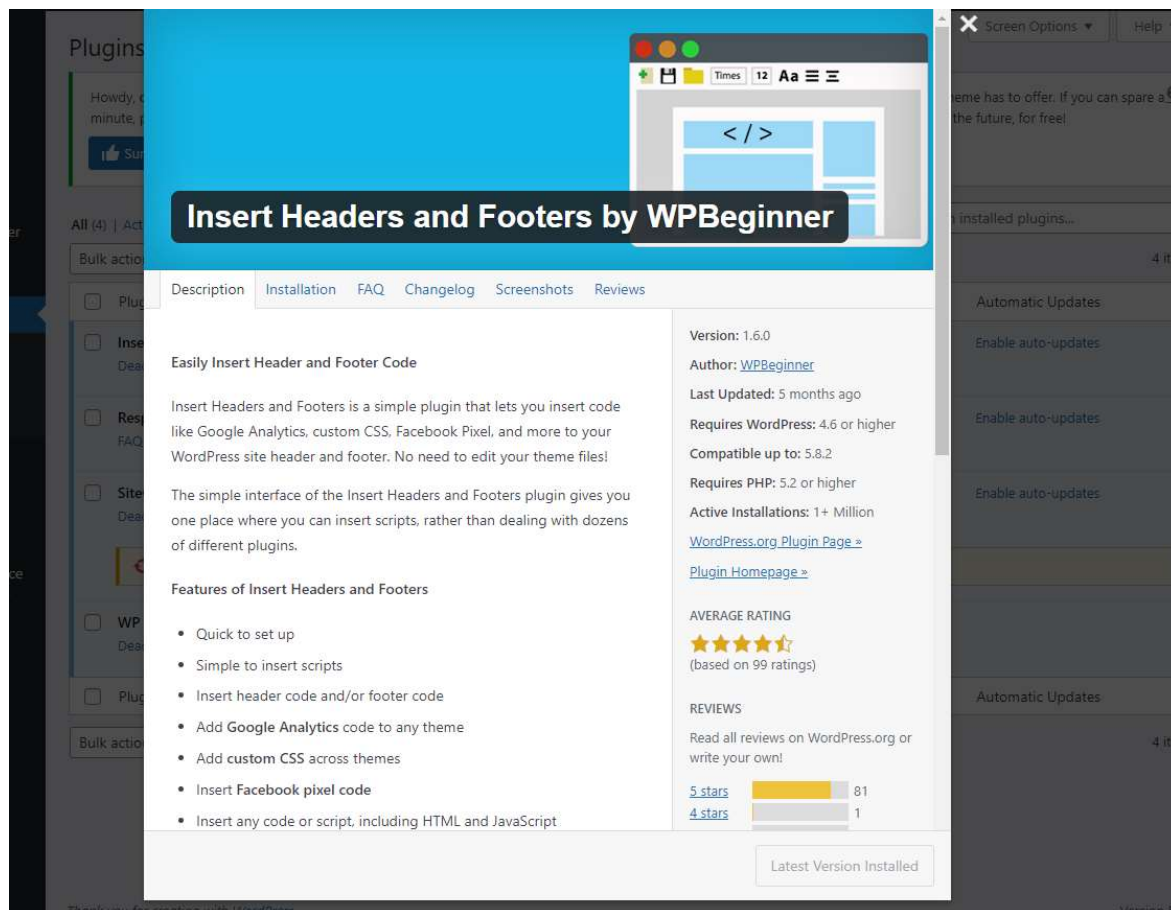
##### Nhược điểm:

- Giọng đọc chưa được tự nhiên.
- Riêng ngôn ngữ tiếng việt chưa hỗ trợ giọng đọc là nữ.

#### 1.5. Plugin Insert Headers and Footers

Insert Headers and Footers là một plugin đơn giản cho phép chèn mã như Google Analytics, CSS tùy chỉnh, Facebook Pixel, v.v. vào đầu trang và chân trang của trang web WordPress. Không cần phải chỉnh sửa các tệp chủ đề. Giao diện đơn

giản của plugin Insert Headers and Footers cung cấp một nơi để có thể chèn tập lệnh, thay vì xử lý hàng chục plugin khác nhau.



Hình 7: Plugin Insert Headers and Footers

Các tính năng của Chèn đầu trang và chân trang:

- Nhanh chóng để thiết lập
- Đơn giản để chèn tập lệnh
- Chèn mã đầu trang và / hoặc mã chân trang
- Thêm mã Google Analytics vào bất kỳ chủ đề nào
- Thêm CSS tùy chỉnh trên các chủ đề
- Chèn mã pixel Facebook
- Chèn bất kỳ mã hoặc tập lệnh nào, bao gồm HTML và JavaScript

## 2. Cài đặt hệ thống

### 2.1. Cài đặt Wordpress

- Tiến hành tải và cài đặt Wordpress từ trang <https://vi.wordpress.org/download/>
- Tiến hành tải và cài đặt web server Xampp <https://www.apachefriends.org/download.html>
- Tiến hành giải nén file wordpress vừa tải về, sau đó copy toàn bộ mã nguồn vào thư mục htdocs của Xampp.

Ví dụ: C:\xampp\htdocs\testwordpress

- Sau đó bật Xampp lên vào trình duyệt truy cập vào phpmyadmin để tiến hành tạo cơ sở dữ liệu
- Sau khi tạo xong cơ sở dữ liệu tiến hành truy cập localhost để tiến hành cài đặt wordpress

Ví dụ: <http://localhost/testwordpress>

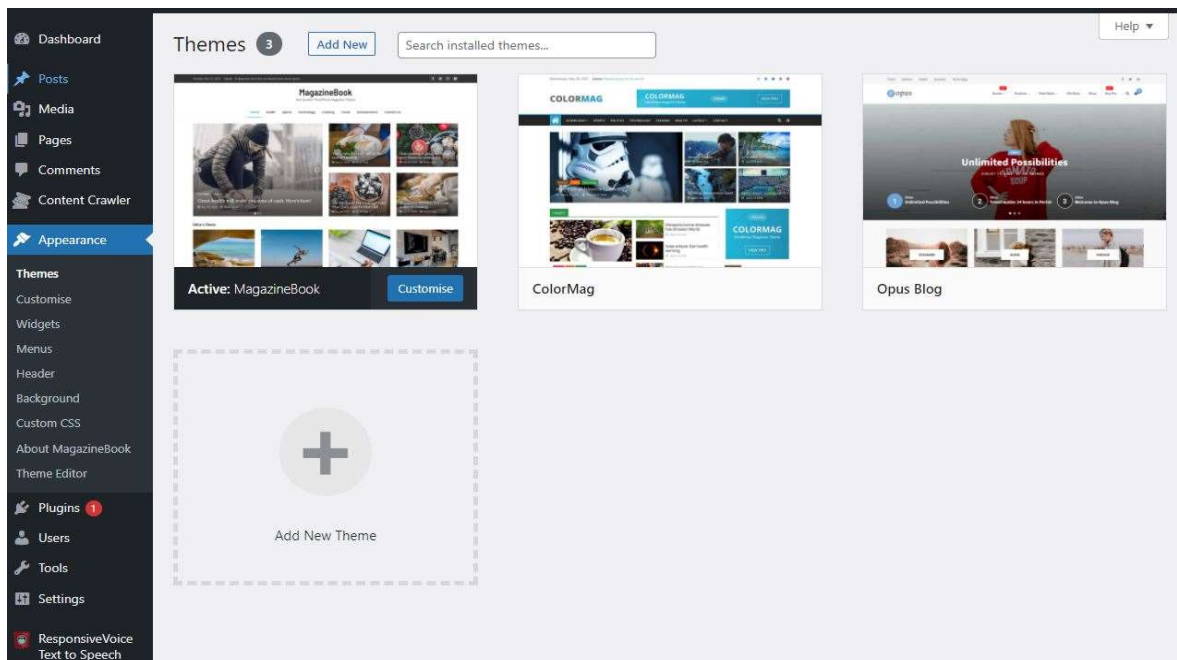
- Làm theo chỉ dẫn của wordpress để tiến hành cài đặt.



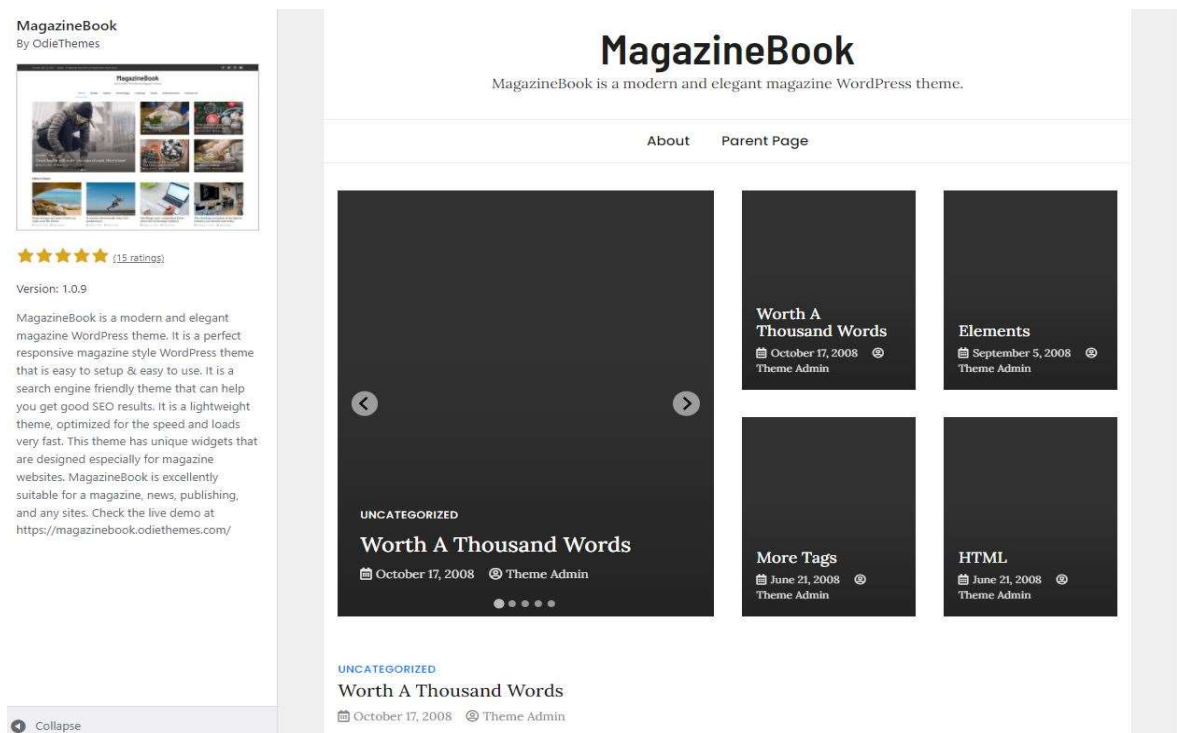
Hình 8: Cài đặt WordPress

## 2.2. Cài đặt theme

- Tiến hành tải và kích hoạt theme MagazineBook



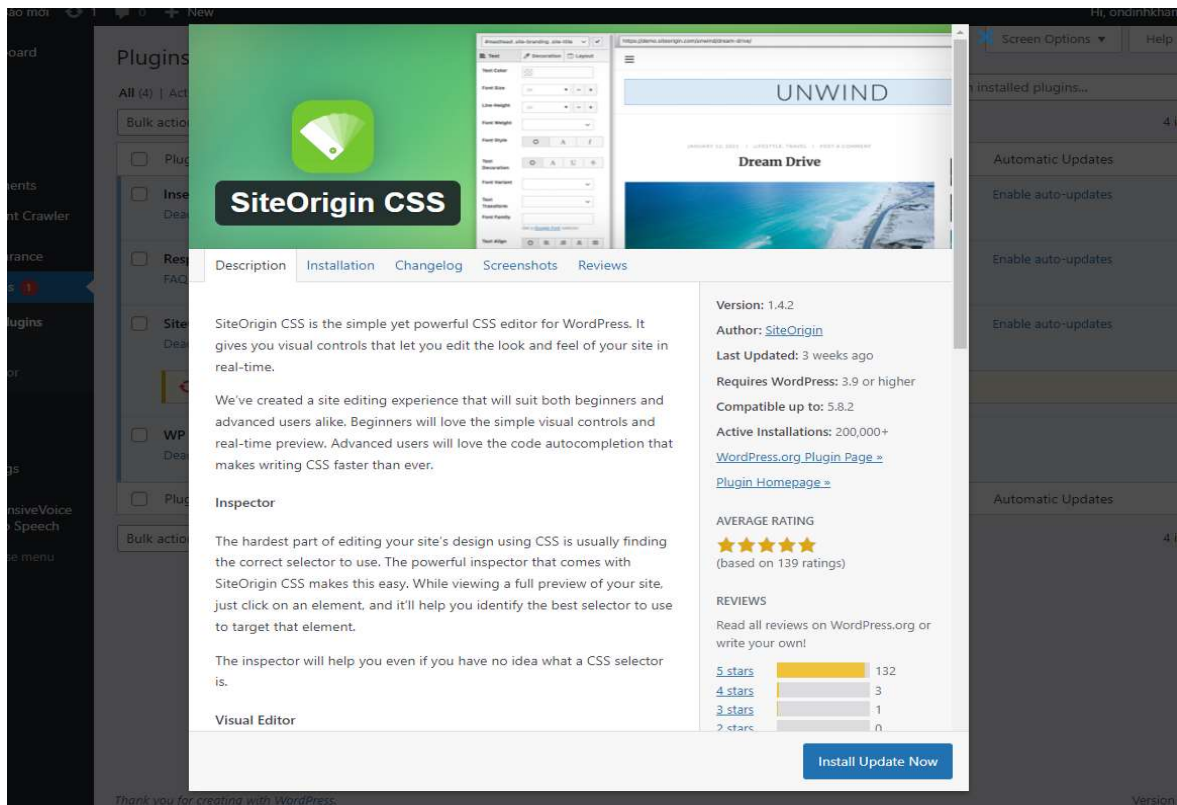
Hình 9: Kích hoạt theme MagazineBook



Hình 10: Bố cục của them Magazine Book

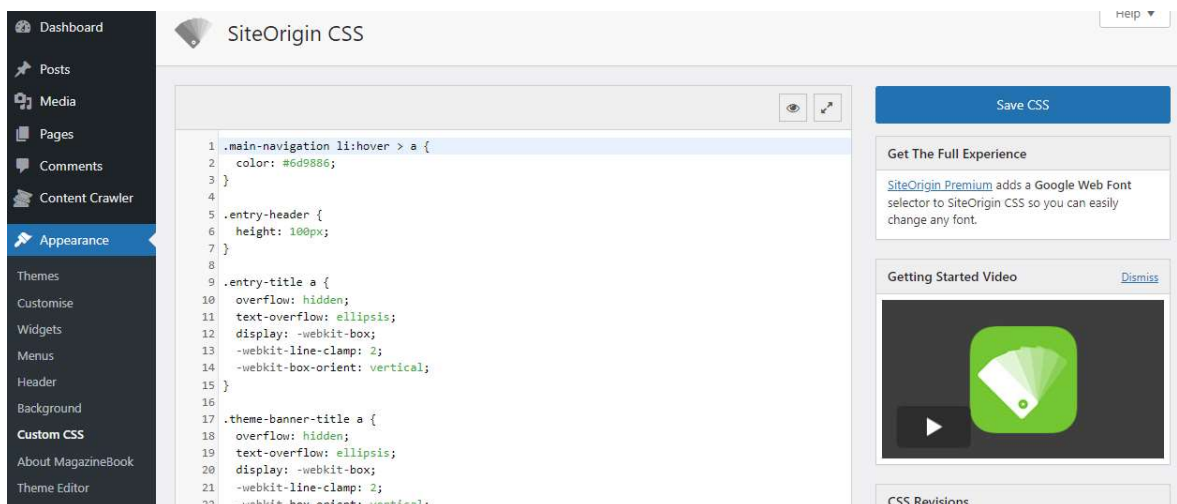
## 2.3. Tinh chỉnh giao diện

- Cài đặt plugin **SiteOrigin CSS** để tiến hành css.



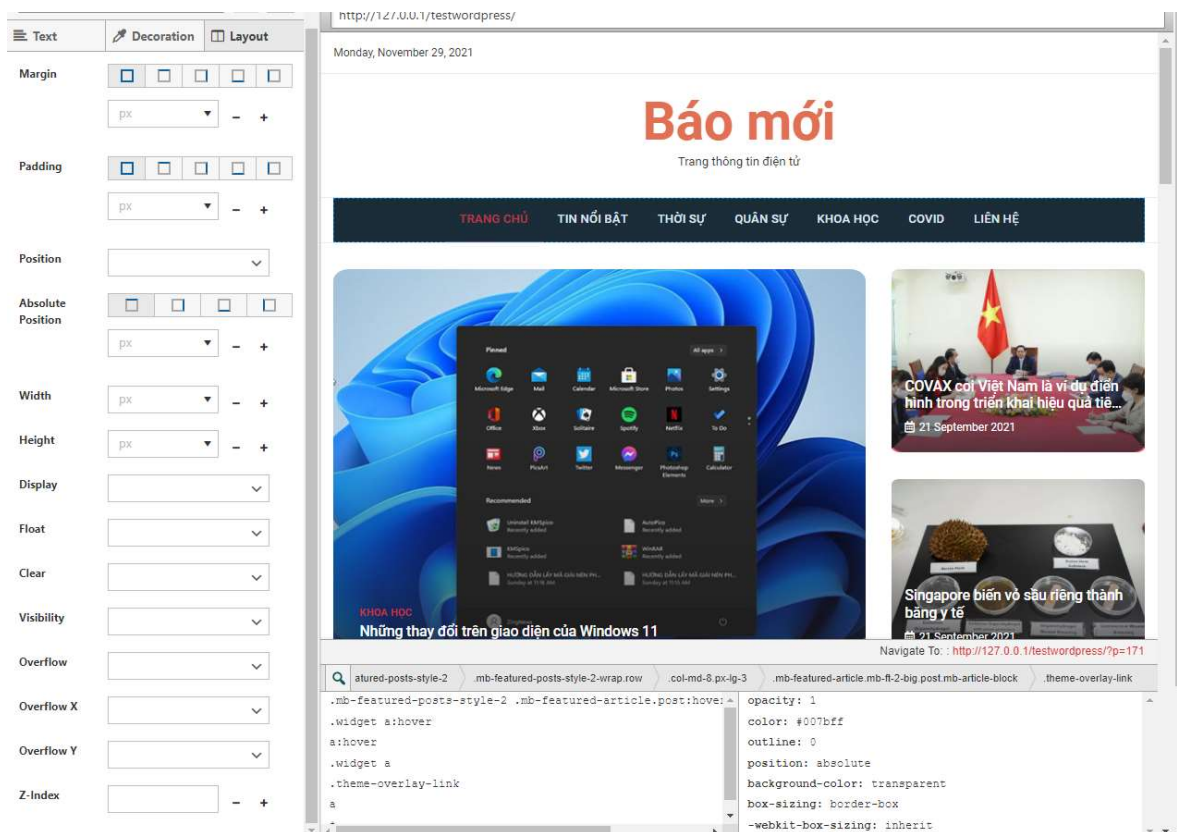
Hình 11: Cài đặt SiteOrigin CSS

- Vào **Appearance -> Custom CSS** để tiến hành CSS cho trang web.



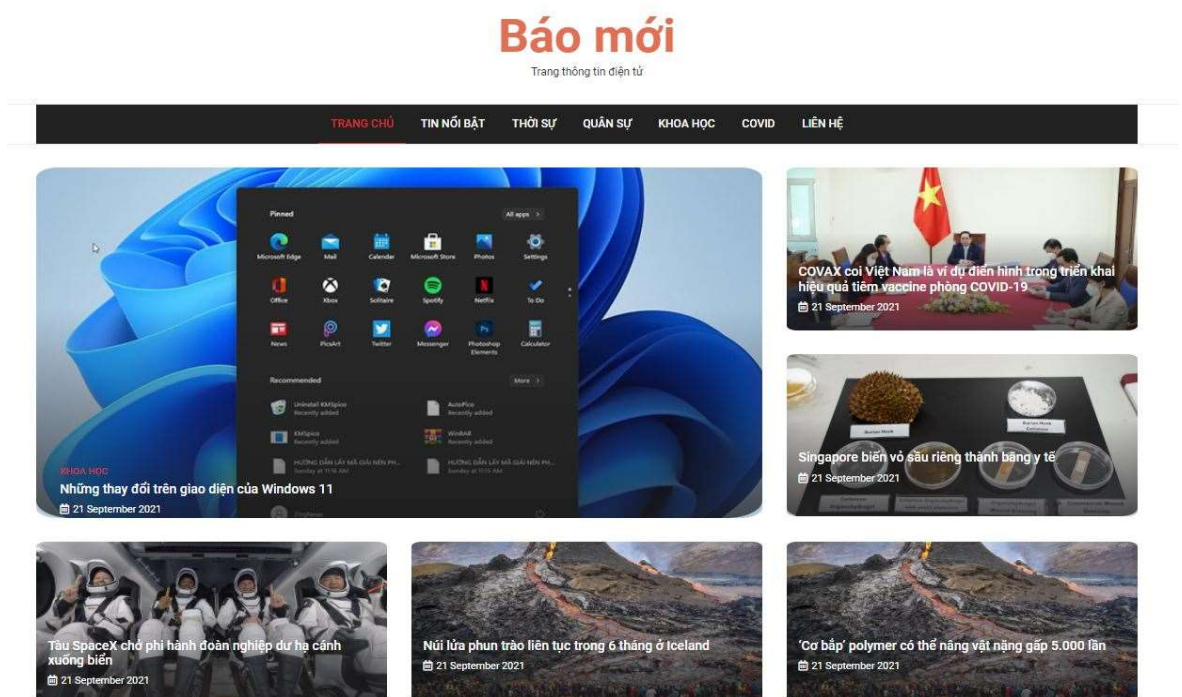
Hình 12: Custom Css bằng giao diện code



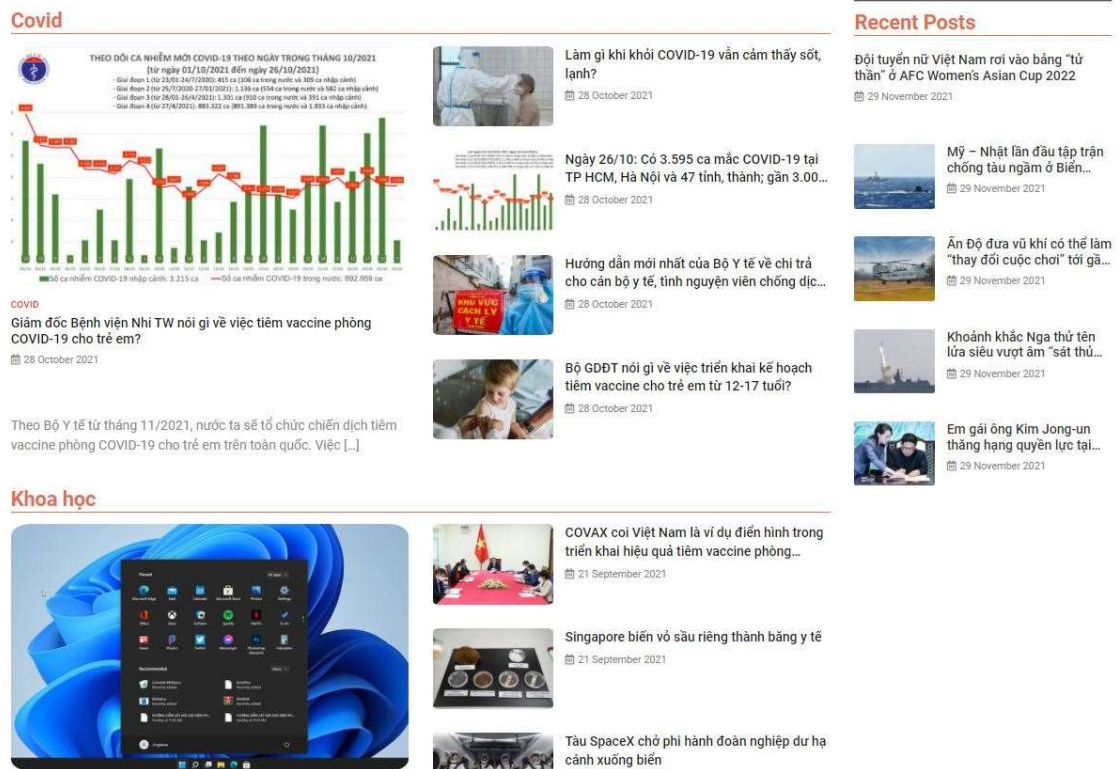


Hình 13: Custom Css bằng giao diện UI

## 2.4. Giao diện của Website



Hình 14: Trang chủ Website khi hoàn thành

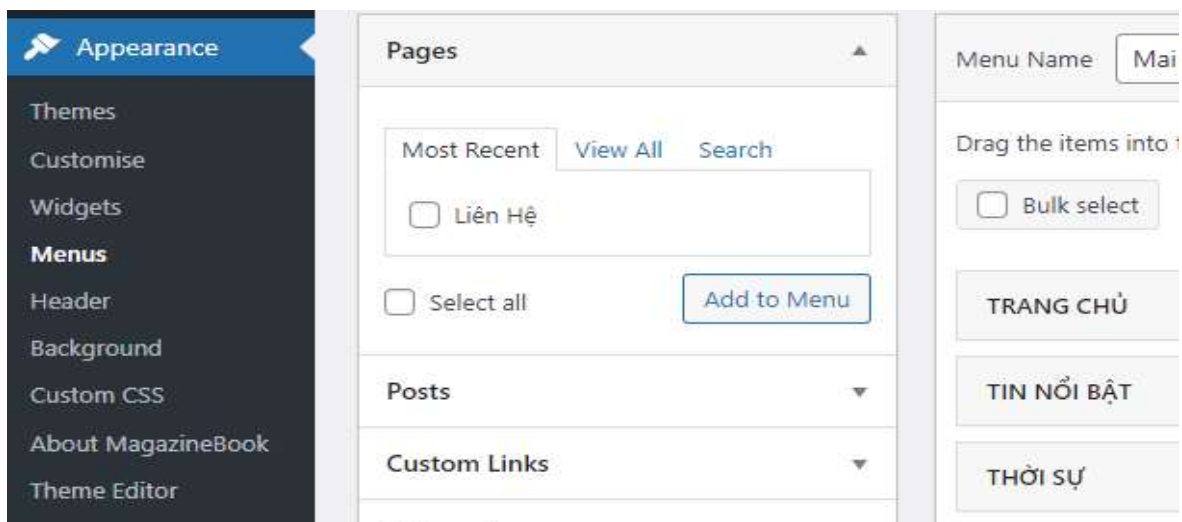


Hình 15: Trang chủ Website

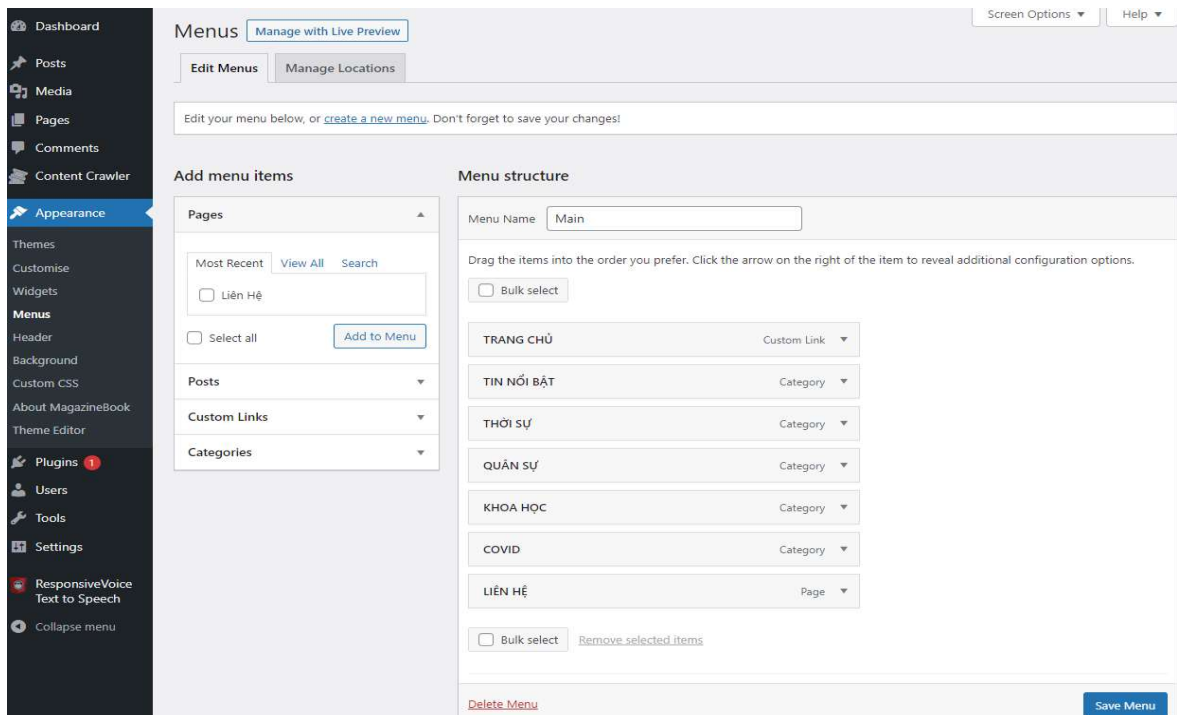
## 2.5. Xây dựng Menu

Menu giúp cho người xem cũng như người quản trị có thể phân mục nội dung một cách dễ dàng.

- Vào **Appearance -> Menu** để tạo menu cho trang web



Hình 16: Appearance -> Menu

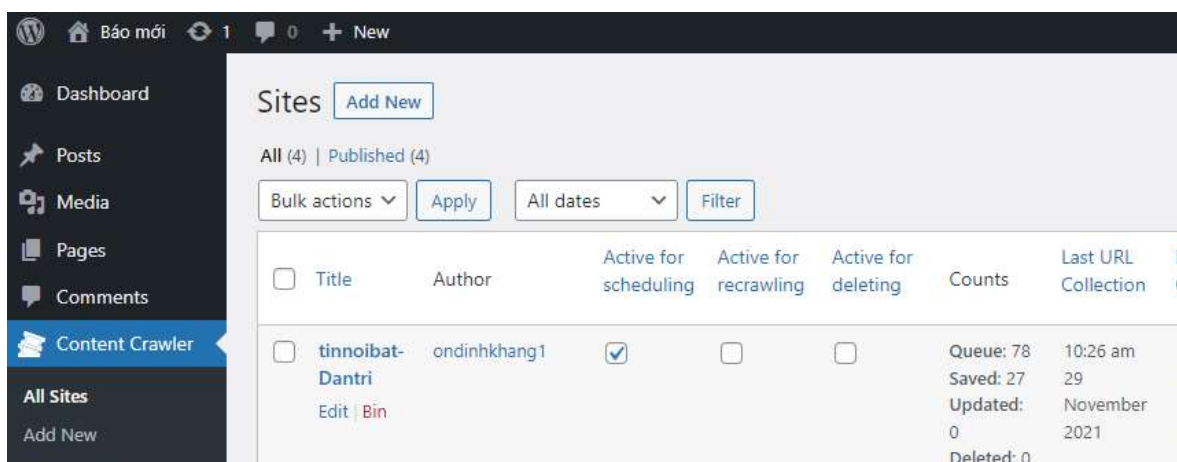


Hình 17: Tạo Menu cho trang

## 2.6. Cài đặt cơ chế lấy tin tự động

Tiến hành cài đặt WP Content Crawler -> tiến hành Active plugin sau khi đã cài đặt xong.

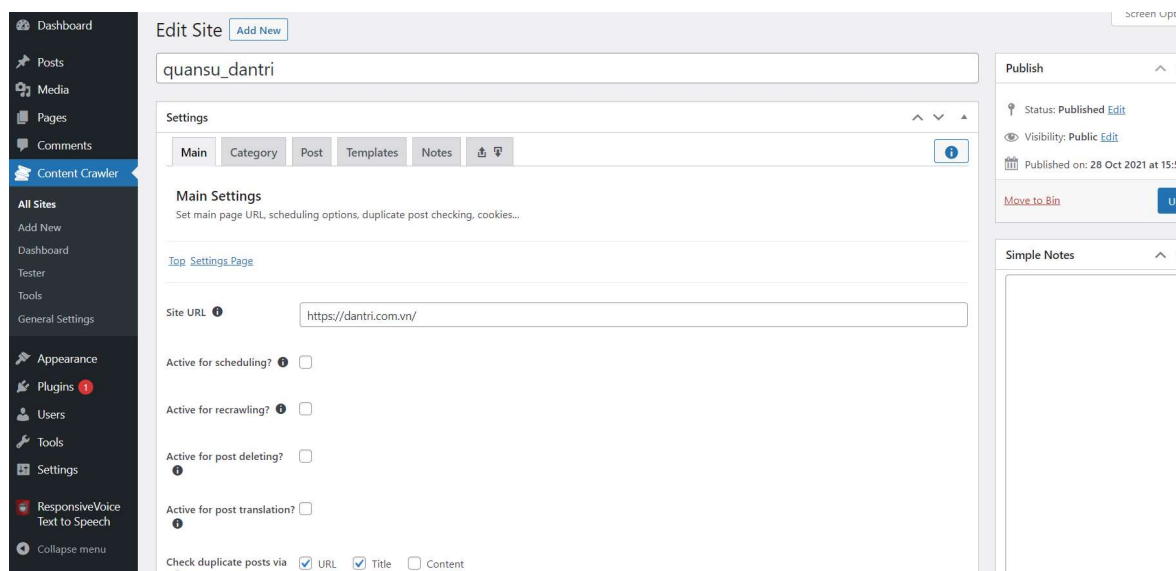
Sau khi đã cài đặt tiến hành vào **Content Crawler** -> **chọn Add New** để tiến hành cấu hình nội dung cần crawl



Hình 18: Add New Content Crawler



Tại mục Add new site nhập tên của chiến dịch crawl, tiếp đến mục Site URL nhập địa chỉ của trang web cần crawl. Trong ví dụ này tôi tiến hành crawl trang dantri.com



Hình 19: Add new Site

**Tại tab Category**, mục Category List Page URL điền địa chỉ web cần crawl

Mục Category Map chọn danh mục cần lưu nội dung vào menu trong trang web của mình (ví dụ mục Quân sự trong menu). Ô kế bên điền địa chỉ trang chứa danh mục cần crawl

Mục Category Post URL Selectors, xác định bộ chọn CSS sẽ tìm các phần tử lưu trữ URL danh mục trong mã nguồn

Mục Test Category URL điền URL trang web để tiến hành kiểm tra.

Các mục còn lại giữ mặc định

Settings

Main Category Post Templates Notes

**Category Page Settings**  
Category mapping and other settings to be used when crawling category pages

[Top](#) [Next Page](#) [Featured Images](#) [Manipulate HTML](#) [Unnecessary Elements](#) [Notifications](#)

Category List Page URL

Category List URL Selectors

Category Map

Test Category URL

Category Post URL Selectors

Collect URLs in reverse order? ☐

Hình 20: Tab Category

Tại tab Post, mục Test Post URL điền địa chỉ trang chứa nội dung bài báo

Main Category Post Templates Notes

**Post Page Settings**  
Selectors and options to be used when crawling post pages

[Top](#) [Category](#) [Date](#) [Meta](#) [Featured Image](#) [Images](#) [Custom Short Codes](#) [List Type Posts](#) [Pagination](#) [Post Meta](#) [Taxonomies](#) [Manipulate HTML](#) [Unnecessary Elements](#) [Notifications](#)

Test Post URL

Post Title Selectors

Post Excerpt Selectors

Post Content Selectors

Post Tag Selectors

Post Slug (Permalink) Selectors

Hình 21: Tab Post

Mục Post Title Selectors chứa selector CSS của mục tiêu đề bài báo

Mục Post Excerpt Selectors chứa selector CSS của phần tóm tắt bài báo, có thể có hoặc không

Mục Post Content Selectors chứa selector CSS của phần nội dung chính bài báo

Mục Post Tag Selectors chứa selector CSS của phần thẻ tag.

Tại phần Featured Image tích chọn mục save, phần này cho phép bài báo dùng ảnh bên trong phần nội dung làm ảnh đại diện bài báo.

### Featured Image

---

Save featured image, if it is ☒  
not found in category  
page?  



*Hình 22:Featured Image*

Phần Image chọn save all để cho phép lưu hình ảnh về phía server.

### Images

---

Save all images in the post ☒  
content?  


Save images as media?  ☒

*Hình 23: Save Image*

Phần Custom Short Code tại một biến conten với selector là nội dung chính của bài báo



Hình 24: Custom Short Code

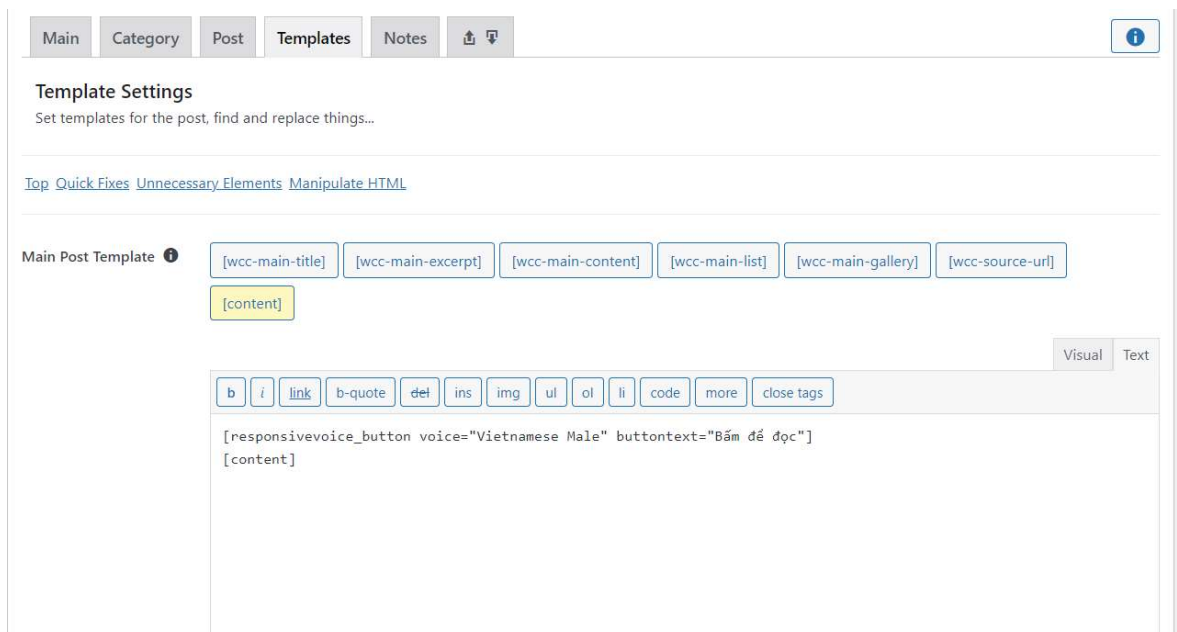
Các thành phần còn lại giữ mặc định

Tại tab Templates, mục Main Post Template điền nội dung:

`[responsivevoice_button voice="Vietnamese Male" buttontext="Bấm để đọc"]`

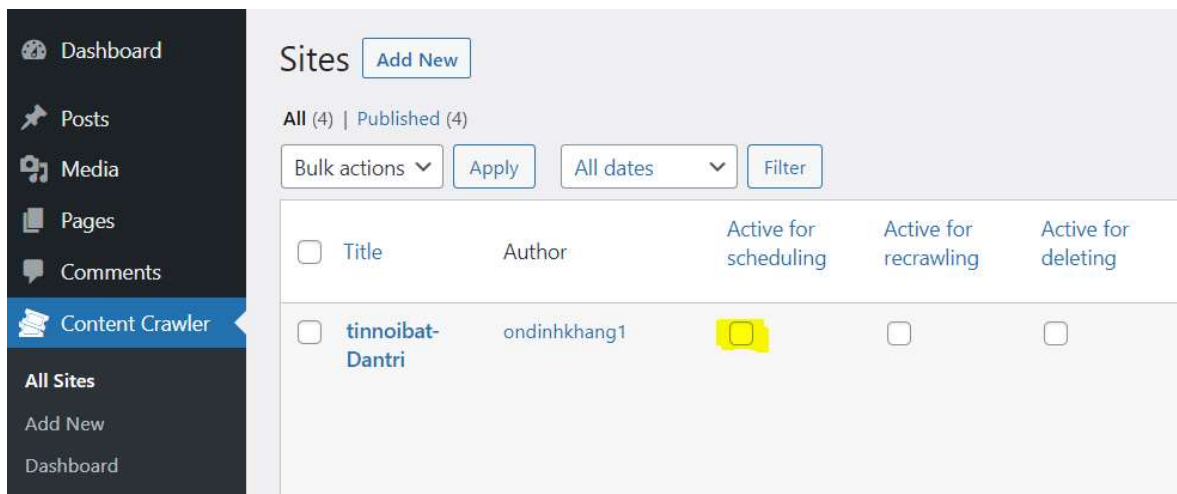
`[content]`

Phần này giúp trang web có thể nhúng plugin tạo giọng nói vào trong lúc crawl dữ liệu



Hình 25: Templates

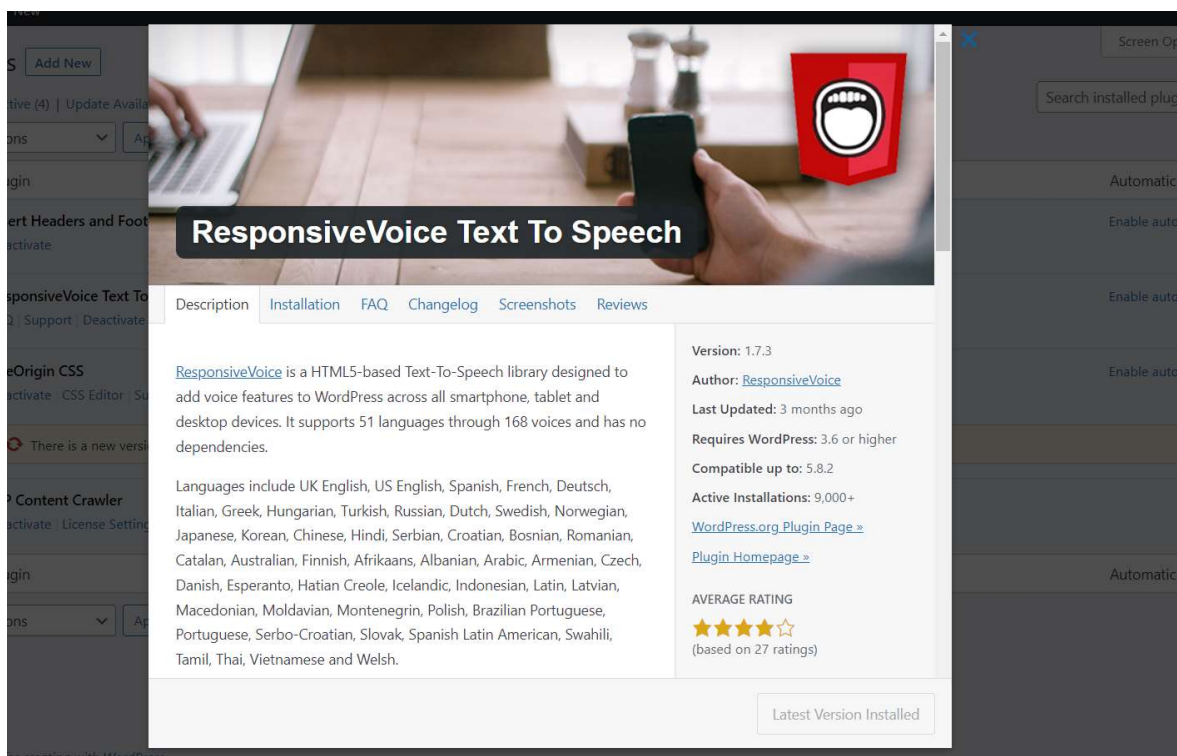
Sau khi đã cấu hình tiến hành ấn save để lưu cấu hình lại, sau đó về mục **All Site** tại dòng chứa title đã đặt lúc đặt tên chiến dịch crawl, tick vào mục Active for scheduling để tiến hành kích hoạt việc crawl



Hình 26: Active for scheduling

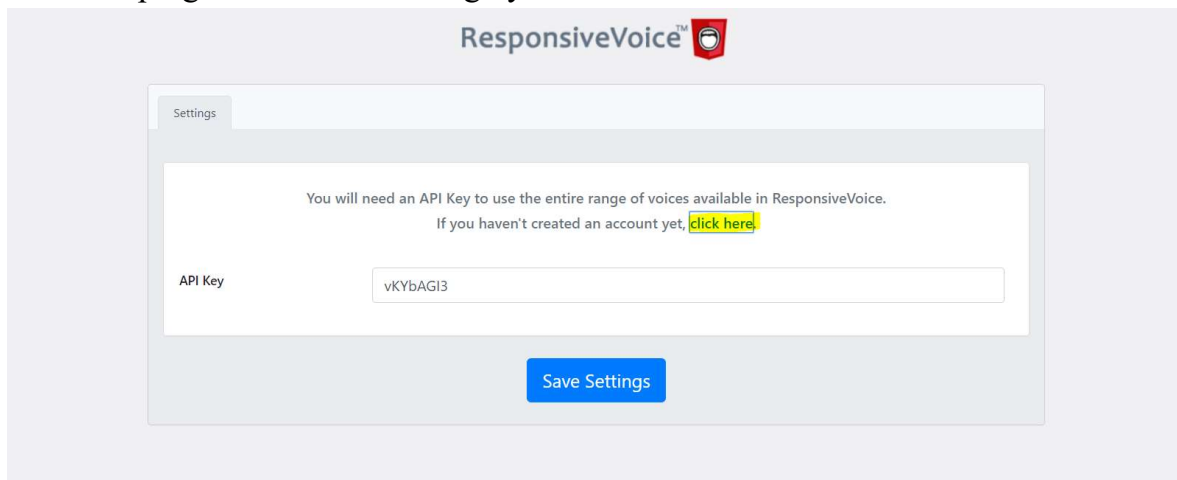
## 2.7. Cài đặt cơ chế giọng đọc

Tiến hành cài đặt và kích hoạt plugin ResponsiveVoice Text To Speech

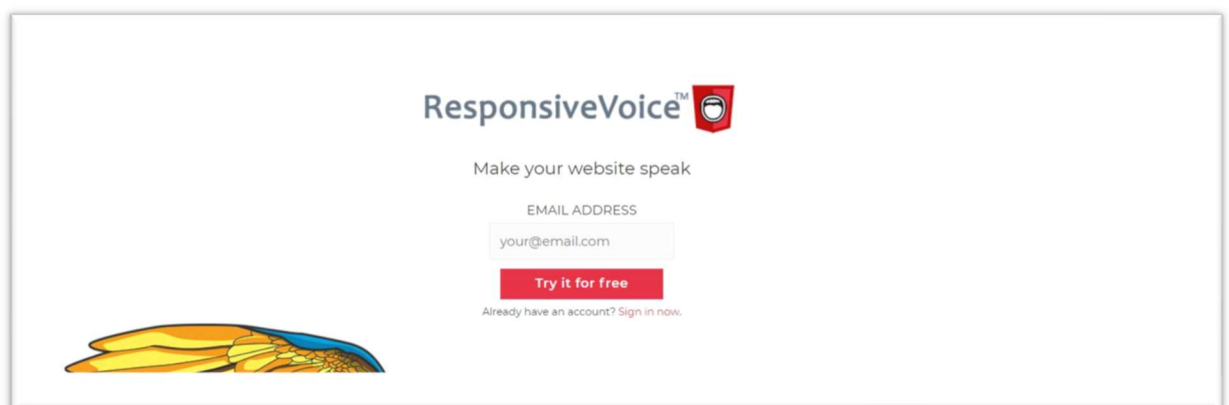


Hình 27: ResponsiveVoice Text To Speech

Bật plugin và tiến hành đăng ký tài khoản:

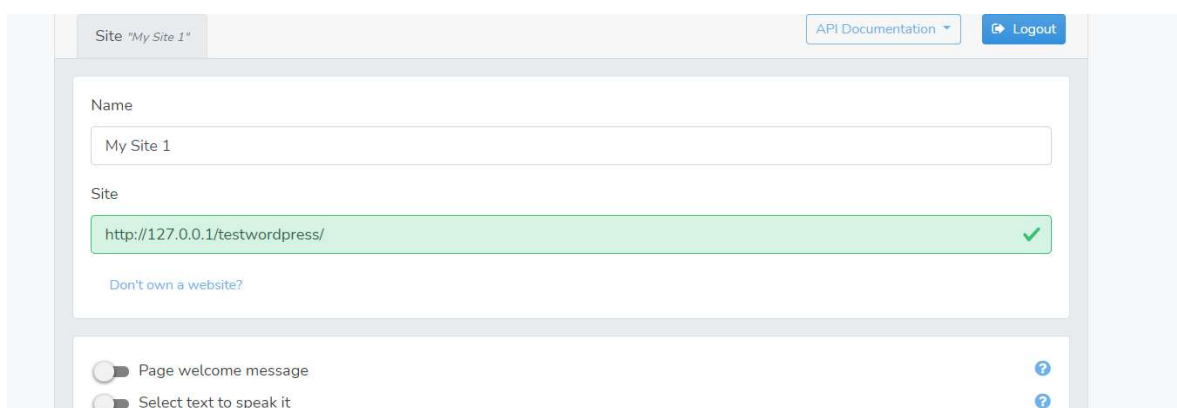


Hình 28: Key API



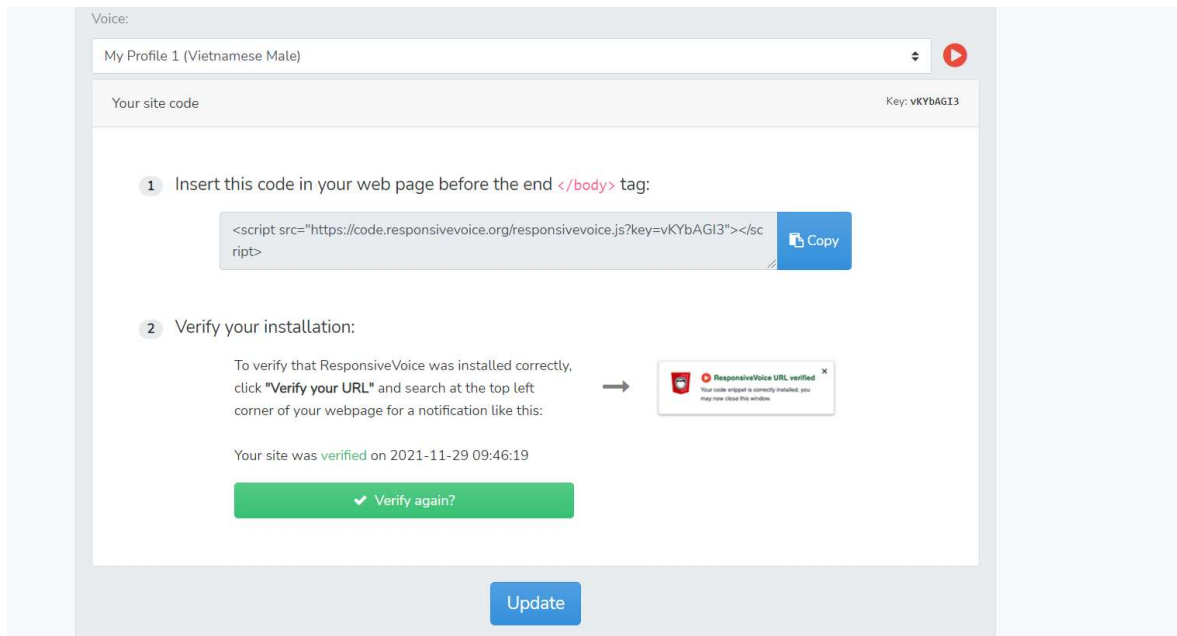
Hình 29: Login ResponsiveVoice

Tiến hành điền các cấu hình và xác minh trang web:



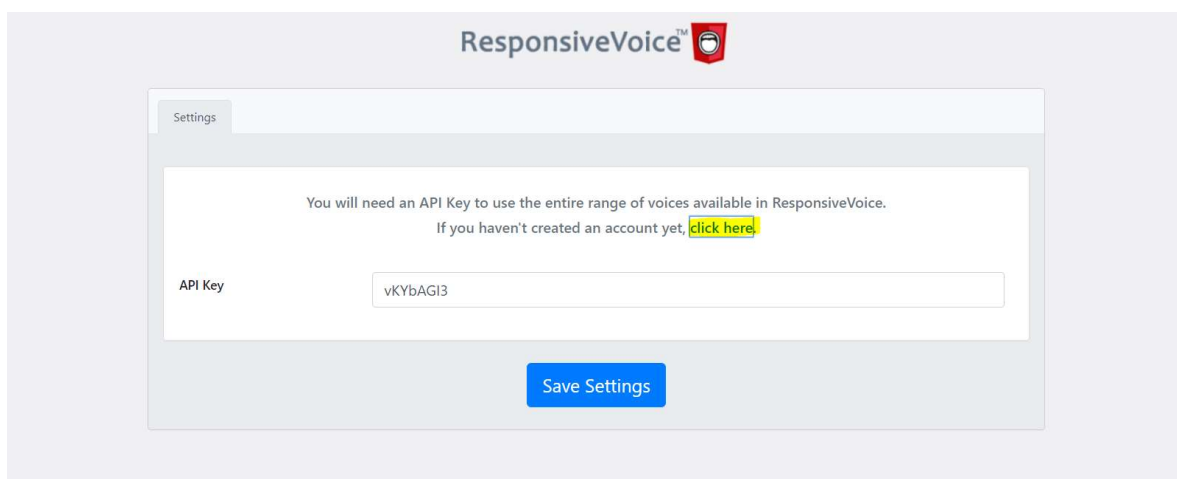
Hình 30 Cấu hình trang

Sau khi đã xác minh trang web xong tiến hành lưu trữ lại key đã cấp, và đoạn script vừa tạo.



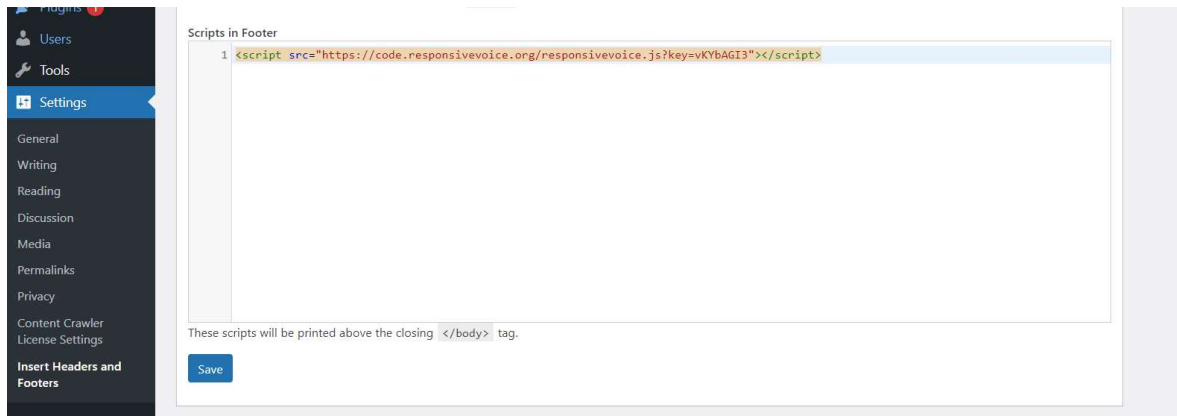
Hình 31: Script ResponsiveVoice

Quay lại plugin ResponsiveVoice Text To Speech và dán key đã lưu lúc nãy



Hình 32: Nhập API Key

Cài đặt plugin Insert Header and Footers và tiến hành dán đoạn script đã được tạo từ bước trước đó vào footer, sau đó nhấn save.



Hình 33: Insert Headers and Footers

### Kết quả:

Sau mỗi bài báo được crawl thì đoạn script của cơ chế giọng đọc sẽ được tự động thêm vào bài. Trên đầu mỗi bài báo sẽ xuất hiện một nút bấm, khi người dùng click vào API sẽ được gọi và sẽ tiến hành đọc nội dung của bài báo.



Hình 34: Button giọng đọc



## CHƯƠNG 3: KIỂM THỬ VÀ ĐÁNH GIÁ

### 1. Kiểm thử

#### 1.1. Cơ chế lấy tin

Trường hợp 1: kiểm tra thời gian lấy bài viết.

	Thời gian lấy	Kết quả
Bài viết nguồn	Dưới 2 phút	Thành công

Trường hợp 2: kiểm tra độ chính xác của bài viết

Bài viết nguồn	Độ chính xác
Tiêu đề bài viết	100%
Nội dung bài viết	95%
Hình ảnh	90%

#### 1.2. Cơ chế giọng đọc

Bài viết	Giọng đọc	Độ chính xác	Nội dung	Tốc độ
Bài viết A	Nam	80%	100%	Hơi chậm

### 2. Đánh giá các cơ chế

#### 2.1. Cơ chế lấy tin

- Cơ chế lấy tin hoạt động ổn định, thời gian lấy tin được cập nhật sau khoảng 2 phút.
- **Nhược điểm:** một số bài báo không hiển thị Featured image.
- **Cách khắc phục:** truy cập và chỉnh sửa thủ công bài viết, tìm hình ảnh tương tự để thay thế

## 2.2. Cơ chế giọng đọc

- Cơ chế đọc không giới hạn số từ, độ chính xác tương đối cao
- **Nhược điểm:** giọng đọc chưa bỏ qua các kí hiệu đặc biệt trong bài, giọng đọc còn chậm và ngắt nhịp không được tự nhiên.
- **Cách khắc phục:** tăng tốc độ giọng đọc.

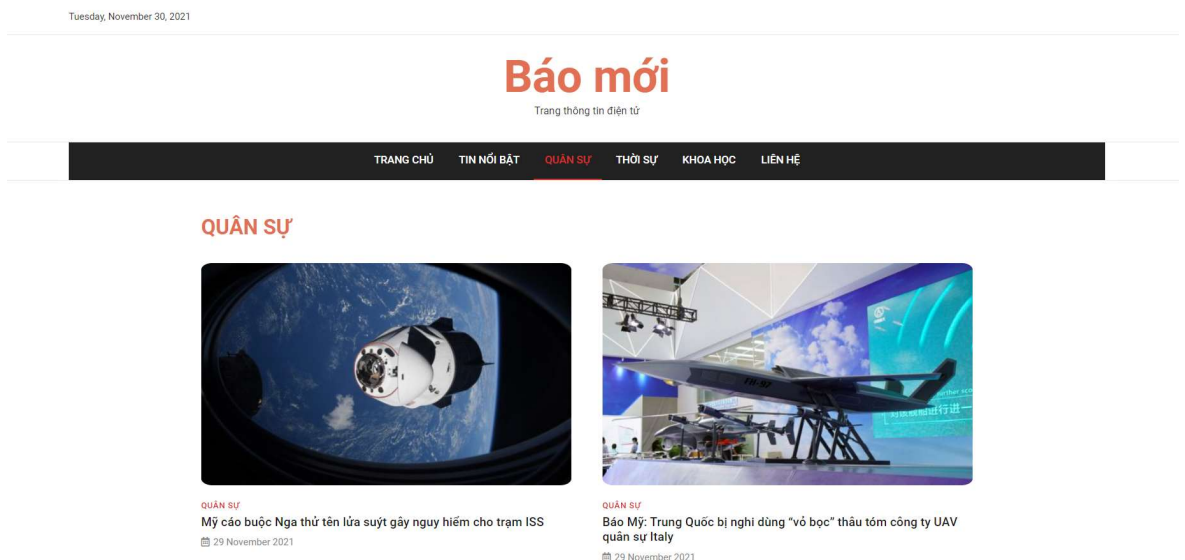
## CHƯƠNG 4: KẾT QUẢ

### 1. Giao diện

- Website có thiết kế giao diện dễ xem, thông tin bài viết hiển thị rõ ràng.  
Website tin tức hoạt động tốt trên nhiều trình duyệt web: Google Chrome, Internet Explorer,... Tuy nhiên chưa tối ưu tốt trên các thiết bị smart phone, tablet.



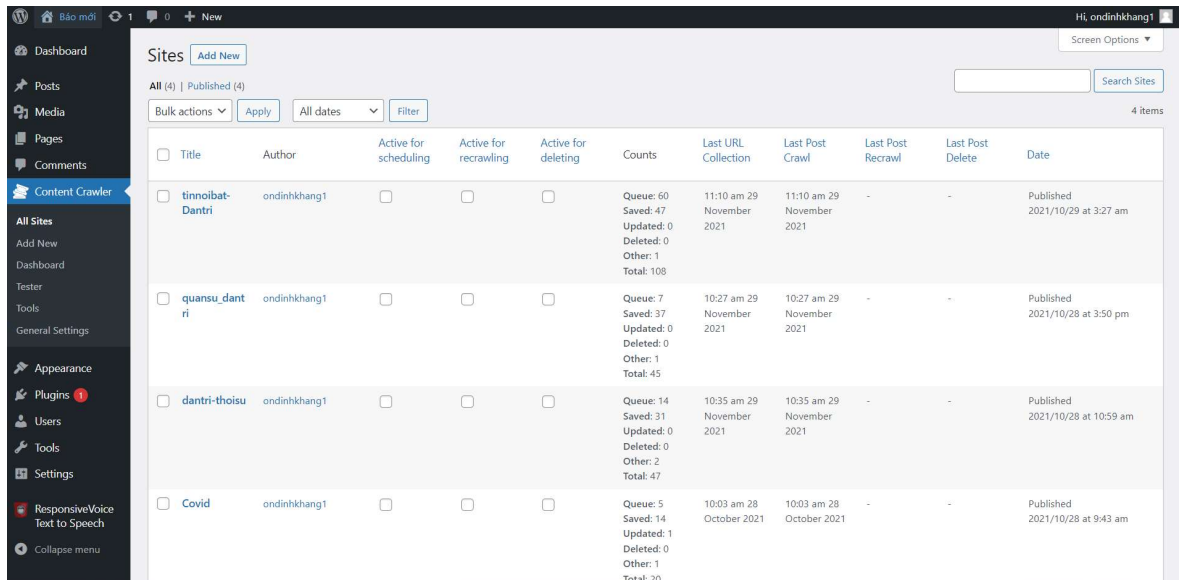
Hình 35: Single Page



Hình 36: Danh mục quân sự

## 2. Cơ chế lấy tin

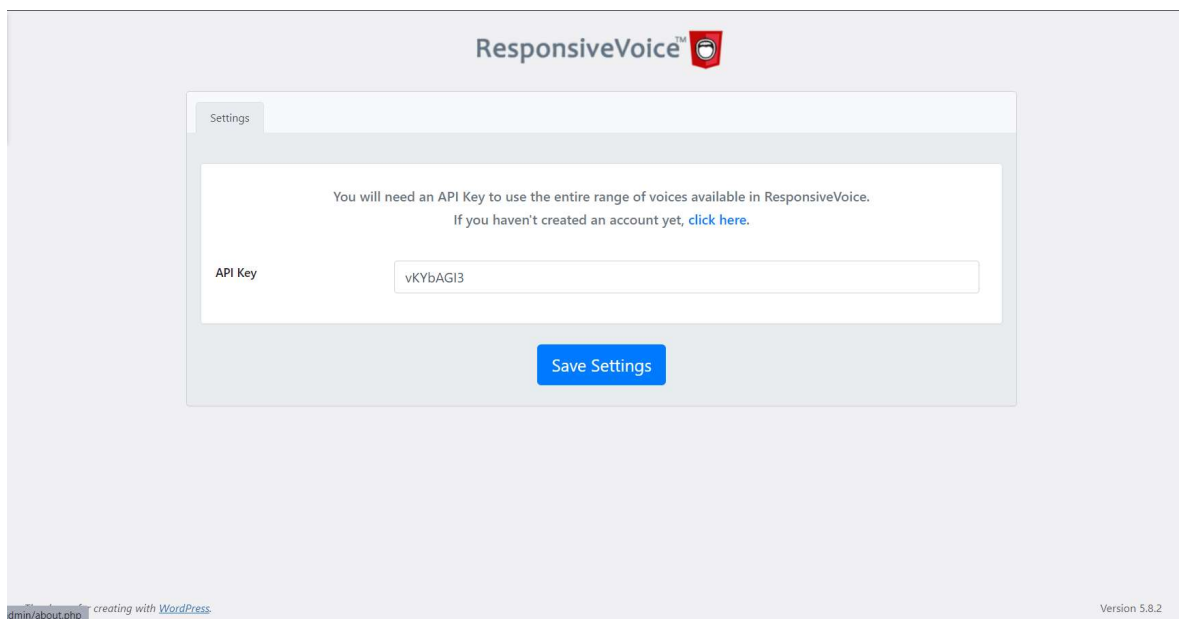
- Cơ chế lấy tin hoạt động ổn định, trung bình mỗi bài báo sẽ được cập nhật dưới 2 phút, nội dung bài báo hiển thị đầy đủ, hình ảnh rõ ràng.



Hình 37: Plugin Content Crawler

## 3. Cơ chế giọng đọc

- Giọng đọc hoạt động tương đối ổn định, nội dung đọc đầy đủ, tuy nhiên giọng đọc chưa được tự nhiên.



Hình 38: Plugin ResponsiveVoice

## PHẦN KẾT LUẬN

### 1. Kết quả đạt được

#### 1.1. Lý Thuyết

Sau khoảng thời gian thực hiện đề tài:

- Ứng dụng các kiến thức đã học được như: thu thập yêu cầu, phân tích các yêu cầu, thiết kế phần mềm, kiểm thử phần mềm ở môn Nhập môn công nghệ phần mềm.
- Học hỏi được thêm một hệ quản trị nội dung mới là WordPress.
- Ứng dụng thực tế được ngôn ngữ PHP, SQL,...
- Hoàn thiện được bản thân về quản lý thời gian và kế hoạch học tập hợp lý, rèn luyện được kỹ năng tự nghiên cứu

#### 1.2. Thực Tế

- Xây dựng thành công website tin tức chạy được trên nhiều trình duyệt.
- Website tin tức được xây dựng dựa vào các trang website tin tức lớn chính thống.
- Nội dung bài viết được lấy từ bài viết gốc hoàn thiện hơn 95% .
- Website hoạt động ổn định, có khả năng lấy tin tự động từ rất nhiều trang website khác.

### 2. Hướng phát triển

- Chuyển đổi ngôn ngữ tiếng viết sang nhiều ngôn ngữ khác cho các người đọc nước ngoài
- Tối ưu giao diện điện thoại.
- Tối ưu hóa nội dung bài viết so với bài viết gốc và nâng chất lượng hình ảnh của bài viết lên.
- Website cần được nâng cấp các Plugin bản quyền có tính phí như: **Theme, plugin giọng đọc, plugin lấy tin** giúp các hoạt động trên website được tốt hơn.
- Cài đặt thêm một số tính năng feedback để nhận những đóng góp từ người dùng.

## PHẦN TÀI LIỆU THAM KHẢO

- **Document Plugin WP Content Crawlers**  
<https://docs.wpcontentcrawler.com/>
- **Document Plugin ResponsiveVoice Text To Speech**  
<https://responsivevoice.org/api/>
- **Document Plugin SiteOrigin CSS**  
<https://siteorigin.com/css/getting-started/>