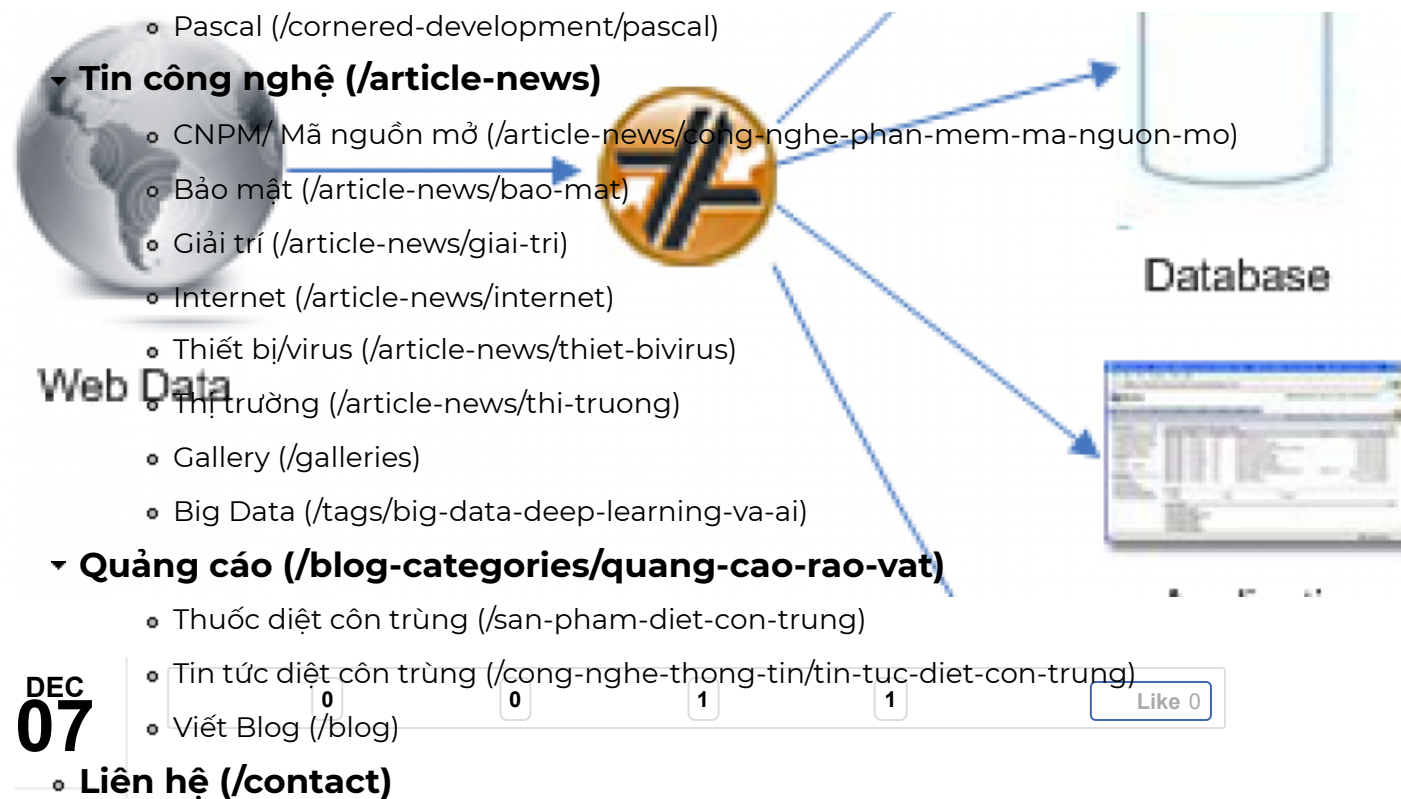




# Rút trích nội dung từ một link website bất kỳ trong ASP.NET

- Máy chủ (/cornered-development/may-chu)
- Mail server (/cornered-development/mail-server)
- C/C++ (/cornered-development/cc)
- Mac iOS (/cornered-development/iphone-mac-ios)
- HTML/CSS (/cornered-development/htmlcss)
- JavaScript (/cornered-development/javascript)
- Lan - Wan (/cornered-development/lan-wan)
- MS access (/cornered-development/microsoft-office-access)



Posted by: admin

Bình luận (/development/1451/rut-trich-noi-dung-tu-mot-link-website-bat-ky-trong-aspnet#disqus\_thread)

## Rút trích nội dung từ một link website bất kỳ trong ASP.NET (/development/1451/rut-trich-noi-dung-tu-mot-link-website-bat-ky-trong-aspnet)

Bạn thiết kế một trang site tin tức đòi hỏi công việc cập nhật tin tức tự động một cách thường xuyên? Nhưng bạn không có đủ nhân lực và thời gian để làm công việc này một cách thủ công? WEBXAULA sẽ hướng dẫn các bạn cách lấy nội dung một website một cách tự động.



Để lấy được nội dung website thông qua một URL bất kì. Việc đầu tiên ta phải lấy về được toàn bộ html của link đó, sau đó phân tích chuỗi html đó, dựa vào các id hoặc name của div, table, span, v.v.... mà ta sẽ lấy được nội dung cần lấy. Để làm được việc này các bạn cần có một chút kiến thức về RegularExpressions.

Do mỗi website có cấu trúc khác nhau nên cách trích lọc và phân tích html cũng khác nhau. Ở đây tôi sẽ ví dụ cho các bạn lấy nội dung tin tức từ website vnexpress.net. Cụ thể tôi sẽ lấy nội dung của link sau: **<http://vnexpress.net/gl/xa-hoi/2012/08/phu-nu-chet-loa-the-trong-khach-san/>**

Đầu tiên, các bạn add thêm 4 thư viện sau:

```
using System.Text.RegularExpressions;
using System.Text;
using System.Net;
using System.IO;
```

Hàm lấy html của link bất kì:

```
public string GetWebContent(string strLink)
{
    string strContent = "";
    try
    {
        WebRequest objWebRequest = WebRequest.Create(strLink);
        objWebRequest.Credentials = CredentialCache.DefaultCredentials;
        WebResponse objWebResponse = objWebRequest.GetResponse();
        Stream receiveStream = objWebResponse.GetResponseStream();
        StreamReader readStream = new StreamReader(receiveStream, System.Text.Encoding.UTF8);
        strContent = readStream.ReadToEnd();
        objWebResponse.Close();
        readStream.Close();
    }
    catch (Exception ex)
    {
        return ex.Message;
    }
}
```

```
}  
    return strContent;  
}
```

Bây giờ công việc tiếp theo là lấy ra từng thành phần mà bạn muốn lấy.

**Phân tích lấy phần tiêu đề:** Các bạn viết hàm sau:

```
public string LayTieuDe(string Content)  
{  
    string pattern = "<H1 class=Title>[^<]+";  
    Regex Title = new Regex(pattern);  
    Match m = Title.Match(Content);  
    if (m.Success)  
        return m.Value.Substring(16, m.Value.Length - 16);  
    return "";  
}
```

**Phân tích lời phần mô tả:** Hàm được viết tương tự như sau:

```
public string LayMoTa(string Content)  
{
```

```
string pattern = "<H2 class=Lead>[^\<]+";  
Regex Title = new Regex(pattern);  
Match m = Title.Match(Content);  
if (m.Success)  
    return m.Value.Substring(15, m.Value.Length - 15);  
return "";  
}
```

**Phân tích lấy phần nội dung.** Các bạn tạo hàm sau:

```
public string LayNoiDung(string Content)  
{  
    string pattern = "<P class=Normal>[^\<]+";  
    Regex Title = new Regex(pattern);  
    Match m = Title.Match(Content);  
    if (m.Success)  
        return m.Value.Substring(16, m.Value.Length -  
16).Replace("/Files", "http://vnexpress.net/Files").Replace("/gl", "http://vnexpress.net/gl");  
    return "";  
}
```

Khi nào cần sử dụng thì bạn chỉ cần gọi các hàm này ra. VD:

```
lblThongBao.Text = LayTieuDe(GetWebContent("http://vnexpress.net/gl/xa-hoi/2012/08/phu-nu-chet-loa-the-trong-khach-san/"))
    + "<br/><br/>"
    + LayMoTa(GetWebContent("http://vnexpress.net/gl/xa-hoi/2012/08/phu-nu-chet-loa-the-trong-khach-san/"))
    + "<br/><br/>"
    + LayNoiDung(GetWebContent("http://vnexpress.net/gl/xa-hoi/2012/08/phu-nu-chet-loa-the-trong-khach-san/"))
    + "";
```

Đây là bước đầu căn bản để lấy nội dung một link bất kỳ, các bạn có thể dựa vào để phát triển thêm lấy các phần nâng cao hơn. Chúc các bạn thành công

---

## Bài cùng chuyên mục

---

- › Giải pháp SESSION TIMEOUT trong ASP.NET (/development/3123/giai-phap-session-timeout-trong-aspnet)
- › Hướng dẫn đưa FCK Editor vào web ASP.NET (/development/1558/huong-dan-dua-fck-editor-vao-web-aspnet)

- › Hướng dẫn cấu hình ứng dụng web ASP.NET (/development/1557/huong-dan-cau-hinh-ung-dung-web-aspnet)
- › Xây dựng website bán quần áo phụ nữ trực tuyến (/development/1522/xay-dung-website-ban-quan-ao-phu-nu-truc-tuyen)
- › Tấn công kiểu SQL Injection và các phòng chống trong ASP.NET (/development/1512/tan-cong-kieu-sql-injection-va-cac-phong-chong-trong-aspnet)
- › Hướng dẫn sử dụng lồng các control trong ASP.Net (/development/1466/huong-dan-su-dung-long-cac-control-trong-aspnet)
- › Xây dựng form liên hệ cho website trong ASP.Net (/development/1465/xay-dung-form-lien-he-cho-website-trong-aspnet)
- › Cách nhúng reCaptcha vào ASP.NET website (/development/1464/cach-nhung-recaptcha-vao-aspnet-website)
- › Cách export dữ liệu ra file excel trong asp.net (/development/1463/cach-export-du-lieu-ra-file-excel-trong-aspnet)
- › Hướng dẫn Rewrite URL trong asp.net (/development/1462/huong-dan-rewrite-url-trong-aspnet)

## Bình luận

0 Comments

Express Magazine

 Login ▾

 Recommend

 Tweet

 Share

Sort by Best ▾



Start the discussion...

LOG IN WITH

OR SIGN UP WITH DISQUS 

Name



☐ I agree to Disqus' [Terms of Service](#)☐ I agree to Disqus' processing of email and IP address, and the use of cookies, to facilitate my authentication and posting of comments, explained further in the [Privacy Policy](#)

Be the first to comment.

## ALSO ON EXPRESS MAGAZINE

**Nha Trang với 10 Địa Điểm du lịch Nha Trang nổi tiếng**

2 comments • 2 years ago



**Jim Collins** — villa nha trang gia rẻ cho thuê giá rẻ 02871060258 để tư vấn , khách sạn nha trang giá rẻ

**UNG THƯ GAN GIAI ĐOẠN CUỐI, GIẢI PHÁP TỐI ƯU: LÁ ĐU ĐỦ VÀ SẢ**

2 comments • 5 years ago



**Hoa** — không biết ngoài việc chữa bệnh còn có tác dụng giải độc cơ thể không? có gây hại cho cơ thể khi

**Có tiền, có con phụ nữ không cần chồng**

5 comments • 5 years ago



**cuong ho** — Đúng là thời hiện đại, khi mà có mọi thứ trong tay thì phụ nữ không cần người đàn ông nữa

**5 cách để copywriter cần làm đối với Google**




3 comments • 5 years ago



**Fan Zheng** — cảm ơn vì bài viết hay :)

CUỐI TIẾP NHƯ ...

---

 [Subscribe](#)  [Add Disqus to your site](#)[Add Disqus](#)  [Disqus' Privacy Policy](#)[Privacy Policy](#)[Privacy Policy](#)[Privacy Policy](#)

## ADVERTISEMENT

---

**Jobsora**

(<https://jobsora.com/>)

## QUẢNG CÁO BÀI VIẾT

---



(/posts/view/2189/chung-ta-co-kiem-tien-tu-drupal-appstore)

Chúng ta có thể kiếm tiền từ Drupal Appstore (/posts/view/2189/chung-ta-co-kiem-tien-tu-drupal-appstore)

Presently the drupal community is in an uproar, triggered by Robert Douglass' announcement of his session at the Drupal Dev Days in Brussels





(/posts/view/1942/database-backups-cua-drupal-site-vao-dropbox)

Database backups của Drupal site vào Dropbox (/posts/view/1942/database-backups-cua-drupal-site-vao-dropbox)

*You can lose all your hard earned user generated content in a flash if you don't regularly create backups of your Drupal database*



**HINO VIỆT NHẬT**  
Số 319, Hùng Vương, Hải Phòng

## **XE TẢI HINO VIỆT NHẬT**

**Đại lý 3S xe tải Hino Hải Phòng**



**Xe tải thùng  
Hino**



**Xe tải Hino  
môi trường**



**Xe tải  
chuyên**

**HOTLINE : 0931.579.900**  
**website : <http://hinohaiphong.vn>**

en

TRIL

(/posts/view/2999/cap-

nhat-cac-key-workflow-pho-bien-trong-open-atrium-2)

Cập nhật các key Workflow phổ biến trong Open Atrium 2 (/posts/view/2999/cap-nhat-cac-key-workflow-pho-bien-trong-open-atrium-2)

*The Open Atrium Workbench module, along with several dependent modules, was a collaboration between Phase2 (sjosh and myself) and the Community (dsnopek).*

## MONTHLY ARCHIVE

---

October 2011 (/archive/201110) (1)

November 2011 (/archive/201111) (296)

December 2011 (/archive/201112) (241)

January 2012 (/archive/201201) (84)

February 2012 (/archive/201202) (63)

March 2012 (/archive/201203) (55)

April 2012 (/archive/201204) (24)

May 2012 (/archive/201205) (26)

June 2012 (/archive/201206) (37)

July 2012 (/archive/201207) (128)

**1 of 8**

sau › (/development/1451/rut-trich-noi-dung-tu-mot-link-website-bat-ky-trong-aspnet?page=1)



---

Copyright © 2011 - 2017 expressmagazine. All rights reserved.



(<http://localhost:5555/development/1451/rut-trich-noi-dung-tu-mot-link-website-bat-ky-trong-aspnet>)