

عنوان: حذف تمامی تگ‌ها منهای چند تگ خاص از HTML دریافتی

نویسنده: وحید نصیری

تاریخ: ۱۳۸۸/۰۶/۰۵ ۲۱:۱۲:۰۰

آدرس: [www.dotnettips.info](http://www.dotnettips.info)

برچسب‌ها: Regular expressions

در ادامه مطلب " [عبارات باقاعده‌ای در مورد کار با تگ‌ها](#) " ، عبارت باقاعده مربوطه به حذف تمامی تگ‌ها برای فرمت زدایی یک متن بسیار جالب است اما مشکلی را که به وجود خواهد آورد، از بین بردن سطرهای موجود است. به عبارت دیگر با استفاده از این عبارت با قاعده، کل متن در امتداد یک سطر قرار می‌گیرد. اکنون می‌خواهیم تمامی تگ‌ها منهای دو تگ مربوط به p و br حذف شوند. چه باید کرد؟

```
private static readonly Regex _pbrRegex = new Regex(@"(<?!br|/br|p|/p).+?>",
    RegexOptions.Compiled | RegexOptions.IgnoreCase);
/// <summary>
/// حذف تمامی تگ‌ها منهای دو تگ ذکر شده
/// </summary>
/// <param name="html"></param>
/// <returns></returns>
public static string CleanTagsExceptPbr(string html)
{
    return _pbrRegex.Replace(html, string.Empty);
}
```

و اگر بخواهیم یک سری تست برای آن بنویسیم به موارد زیر می‌توان اشاره کرد:

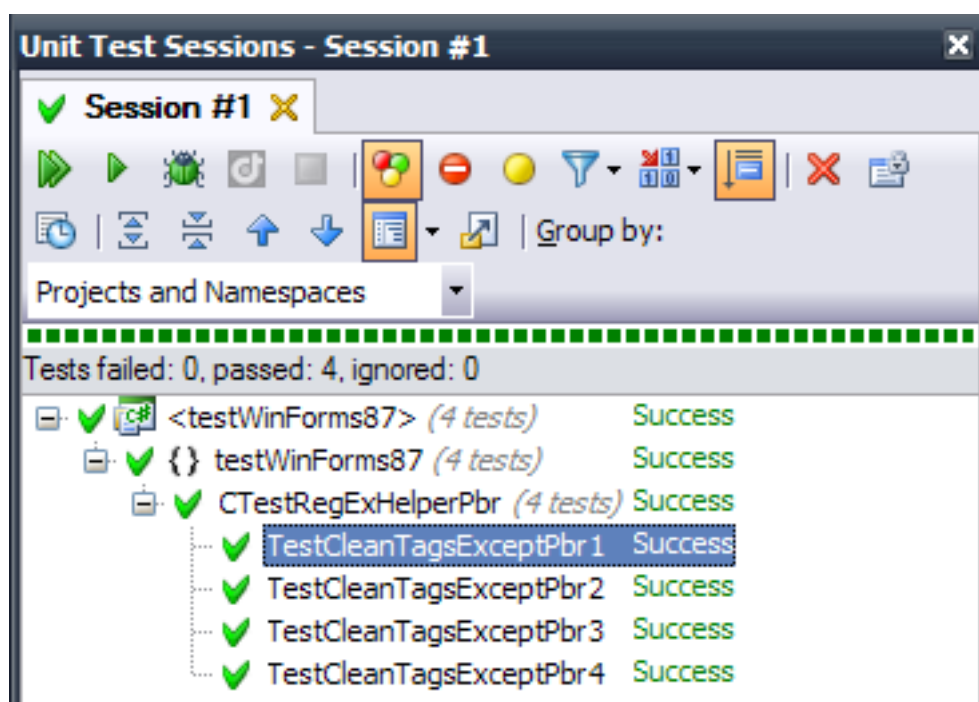
```
using NUnit.Framework;

namespace testWinForms87
{
    [TestFixture]
    public class CTestRegexHelperPbr
    {
        [Test]
        public void TestCleanTagsExceptPbr1()
        {
            Assert.AreEqual(
                CRegexHelper.CleanTagsExceptPbr("<b>data1</b><br/>data2"),
                "data1<br/>data2");
        }

        [Test]
        public void TestCleanTagsExceptPbr2()
        {
            Assert.AreEqual(
                CRegexHelper.CleanTagsExceptPbr("<b>data1</b><br>data2"),
                "data1<br>data2");
        }

        [Test]
        public void TestCleanTagsExceptPbr3()
        {
            Assert.AreEqual(
                CRegexHelper.CleanTagsExceptPbr("<p><b>data1</b><br/>data2</p>"),
                "<p>data1<br/>data2</p>");
        }

        [Test]
        public void TestCleanTagsExceptPbr4()
        {
            Assert.AreEqual(
                CRegexHelper.CleanTagsExceptPbr("<b>data1</b><p>data2<br />"),
                "data1<p>data2<br />");
        }
    }
}
```



## نظرات خوانندگان

نویسنده: mdpdotnet  
تاریخ: ۱۳۸۸/۰۶/۰۵ ۲۳:۲۷:۴۰

در مورد Regex ای که نوشتید یه توضیحی میدید ؟  
من زیاد با رجکس ها جور نیستم. تازه دارم یاد میگیرم.  
(:

نویسنده: وحید نصیری  
تاریخ: ۱۳۸۸/۰۶/۰۶ ۰۰:۲۰:۳۵

سلام  
یک debug visualizer برای VS.Net هست به نام Regular Expression Visualizer. با VS2005 و 2008 سازگار است.  
آنها از آدرس زیر دریافت کنید:  
<http://weblogs.asp.net/roshero/archive/2005/11/26/AnnouncingRegexKit10.aspx>  
سپس فایل‌های dll آنرا در یکی از مسیرهای زیر بسته به نگارش VS.Net خودتون کپی کنید:  
My Documents\Visual Studio 2008\Visualizers  
یا  
My Documents\Visual Studio 2005\Visualizers  
اکنون در VS.Net روی سطر return \_pbrRegex یک breakpoint بگذارید و نتیجه را مشاهده کنید.  
یک مثال عملی:  
<http://professionalaspnet.com/archive/2008/06/18/Regular-Expression-Visualizer.aspx>  
برای نمونه خروجی عبارت باقاعده مثال جاری به صورت زیر است که کمک شایانی است در درک عبارت فوق و امثال آن:

```
>  
zero-width negative lookahead  
br  
or  
br/  
or  
p  
or  
p/  
End Capture  
(any character) .  
(one or more times) (non-greedy) +  
<
```

نویسنده: DotNetCoders  
تاریخ: ۱۳۸۸/۰۶/۰۶ ۰۱:۳۶:۴۹

ممنون جناب نصیری فقط من با این بخش آخر رجکس مشکل دارم.

این بخش منظوره :

<?+.

معنیش چیه؟

اگه کتاب کامل و مفید یا منبع خوبی بهم معرفی کنید خودم پیدا میکنم وقت شما رو هم نمیگیرم.

نویسنده: وحید نصیری  
تاریخ: ۱۳۸۸/۰۶/۰۶ ۰۲:۴۳:۴۰

- لطفا سه چهار سطر آخر پاسخ ارسال شده بنده را مرور بفرمائید.
  - کتاب خوب هم، همه این‌را توصیه می‌کنند:
- Mastering Regular Expressions از O'Reilly Media که تا به حال سه edition از آن منتشر شده.