

Forum:[Hack Forums](#)

Thread: [Finding a certain string in a webpage](#)

TomTheBeast - May 5, 2012, 3:19 pm

Hey, Well I've come across a problem, I'm creating a program to find a certain string on a webpage, and the problem is that i can't find out how to do this, can anyone show me anywhere that i could easily find this or steer me in the right direction?

Eternal Five - May 5, 2012, 3:25 pm

```
string source = new WebClient().DownloadString("website");
if (source.Contains("derp")) MessageBox.Show("Source contains derp.");
```

TomTheBeast - May 5, 2012, 3:55 pm

That's not what I'm asking, I'm not sure if i spcified this in my main post, but i need it to output the certain string that it contains, lets say it says "derp" then i need it to grab everything after that until it says "herp", which i have that, but now i need to find how to remove all of the HTML formatting. This is what i have to remove everything except what i need.

```
Code private void getText(string url)
{
    WebClient wc = new WebClient();
    string[] derp = { "derp" };//where the grabbing begins
    string[] herp = { "herp" };//where it ends
    string webpage = getPageSource(url);
    string[] webpage1 = webpage.Split(derp, StringSplitOptions.RemoveEmptyEntries);
    string[] webpage2 = webpage1[1].Split(herp, StringSplitOptions.RemoveEmptyEntries);
    richTextBox1.Text = webpage2[0];
}
```

I just need to remove the HTML tags and output it as simple text.

Eternal Five - May 5, 2012, 4:05 pm

Use regex: `Regex.Replace(source, "<.*?>", string.Empty)`

TomTheBeast - May 5, 2012, 4:42 pm

Thanks, but now I have another problem, and its only with this website, it has the code im trying to grab repeated about 50 times...

Heres some more info on what my project is: basically is goes to rswiki, grabs the NPC's weakness depending on the input, and then inputs the weakness part out in a RTB, but now my problem is, is that the part im grabbing is:

```
Code <tr>
<th nowrap="nowrap">Weakness
</th><td><span style="display:none">[[[]], <a href="/wiki/Silverlight"
title="Silverlight">silverlight</a>/<a href="/wiki/Darklight" title="Darklight">darklight</a>, <a
href="/wiki/Abyssalbane_bolts" title="Abyssalbane bolts" class="mw-redirect">abyssalbane
bolts</a>/<a href="/wiki/Abyssalbane_arrows" title="Abyssalbane arrows"
class="mw-redirect">arrows</a>, <a href="/wiki/Slash" title="Slash"
class="mw-redirect">slash</a><a href="/wiki/Magic" title="Magic">magic</a>, <a
href="/wiki/Silverlight" title="Silverlight">silverlight</a>/<a href="/wiki/Darklight"
title="Darklight">darklight</a>, <a href="/wiki/Abyssalbane_bolts" title="Abyssalbane bolts"
class="mw-redirect">abyssalbane bolts</a>/<a href="/wiki/Abyssalbane_arrows" title="Abyssalbane
```

[illegible]

[illegible]

class="mw-redirect">crush attacks, slash attacks and magic attacksDarklight, Silverlight, Holy water, Crush attacks, Slash attacks and Magic attacks</td></tr>