

# Regular Expressions

**Source:** <http://coding.derkeiler.com/Archive/PHP/alt.php/2004-03/0385.html>

---

**From:** Andrew Bullock ([trullock\\_at\\_yahoo.com](mailto:trullock_at_yahoo.com))

**Date:** 03/10/04

Date: Wed, 10 Mar 2004 14:32:01 -0000

Hi

Im REALLY stuck with regular expressions!

Can someone please give me a pointer with this...

Im trying to extract a load of images from a site, which are in hrefs.  
ie. i want to extract the url and image path

typical hrefs would be:

```
<a href="test.html"></a>  
<a target=_top href="test.html"></a>  
<a href='test.html'> <img src=clickme.jpg></a>
```

as you can see, there may be double quotes, single quotes or no quotes surrounding the image and url path, so i need to be able to account for this.

Also i only want to return info, where there is a jpeg as an image to click, and the url is a html.

Does that make sense?

```
$matchstr =  
'<a\s+.*?href=[\"'\s]?(.*)\"'+src=[\"'\s]?(.*)\"(.*)></a>/i';
```

That almost works, but it returns this :

```
test.html"><img  
thumbs/16.jpg>
```

how can i fix this?

Many thanks in advance

Andrew