

Re: cut strings and parse for images

Source: <http://coding.derkeiler.com/Archive/Python/comp.lang.python/2004-12/0809.html>

From: Paul McGuire (ptmcg_at_austin.rr._bogus_.com)

Date: 12/06/04

Date: Mon, 06 Dec 2004 20:36:36 GMT

"Andreas Volz" <usenet-spam-trap@brachttal.net> wrote in message
news:20041206203456.251c6d85@frodo.mittelerde...

> *Hi,*

>

> *I used SGMLParser to parse all href's in a html file. Now I need to cut
> some strings. For example:*

>

> <http://www.example.com/dir/example.html>

>

> *Now I like to cut the string, so that only domain and directory is
> left over. Expected result:*

>

> <http://www.example.com/dir/>

>

> *I know how to do this in bash programming, but not in python. How could
> this be done?*

>

> *The next problem is not only to extract href's, but also images. A href
> is easy:*

>

> *Install*

>

> *But a image is a little harder:*

>

> ``

>

Check out the urlparse module (in std distribution). For images, you can
provide a default addressing scheme, so you can expand "images/marine.jpg"
relative to the current location.

— Paul