The htmllib module

This module contains a tag-driven HTML parser, which sends data to a formatting object. For more examples on how to parse HTML files using this module, see the descriptions of the **formatter** module.

Example: Using the htmllib module

```
# File: htmllib-example-1.py

import htmllib
import formatter
import string

class Parser(htmllib.HTMLParser):
    # return a dictionary mapping anchor texts to lists
    # of associated hyperlinks

    def __init__(self, verbose=0):
        self.anchors = {}
        f = formatter.NullFormatter()
        htmllib.HTMLParser.__init__(self, f, verbose)

    def anchor_bgn(self, href, name, type):
        self.save_bgn()
        self.anchor = href

    def anchor_end(self):
        text = string.strip(self.save_end())
        if self.anchor and text:
            self.anchors[text] = self.anchors.get(text, []) + [self.anchor]

file = open("samples/sample.htm")
html = file.read()
file.close()

p = Parser()
p.feed(html)
p.close()

for k, v in p.anchors.items():
    print k, "=>", v

print


link => ['http://www.python.org']
```

If you're only out to parse an HTML file, and not render it to an output device, it's usually easier to use the **sgmllib** module instead.