

Bonjour,
Je suis en train d'essayer de coder l'ouverture d'un fichier HTML avec python comme suit :

```
#importation du module urllib depuis la bibliothèque standart python:
import urllib
#from HTMLParser import HTMLParser.      #pour partie récupération

#import re
#from BaseHTMLProcessor import BaseHTMLProcessor

#test d'ouverture du fichier HTML:
#sock = urllib.urlopen("/home/to105478/Comparaison_Rapports/Filled_Hole/v7-1-1") #fonction urlopen du module
urllib
#htmlSource = sock.read().
#sock.close()
#print htmlSource

#test de récupération de données:
sock = urllib.urlopen("/home/to105478/Comparaison_Rapports/Filled_Hole/v7-1-1")

class MyHTMLParser(HTMLParser):
    def handle_starttag(self, tag, attrs):
        print "Start tag:", tag
        for attr in attrs:
            print "    attr:", attr
    def handle_endtag(self, tag):
        print "End tag :", tag
    def handle_data(self, data):
        print "Data    :", data
    def handle_comment(self, data):
        print "Comment :", data
    def handle_entityref(self, name):
        c = unichr(name2codepoint[name])
        print "Named ent:", c
    def handle_charref(self, name):
        if name.startswith('#x'):
            c = unichr(int(name[1:], 16))
        else:
            c = unichr(int(name))
        print "Num ent  :", c
    def handle_decl(self, data):
        print "Decl    :", data

parser = MyHTMLParser().#association au parser des classes créées pour partition et lecture des données de la
feuille html

#parser.feed("sock"). #ne marchera certainement pas, préférer:
parser.feed(<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN"><HTML><HEAD><TITLE>Composite-Filled-Hole
Dossier</TITLE><META http-equiv=Content-Type content="text/html; charset=windows-1252"><LINK href="FH avec
Bush_files/plugin.css" type=text/css rel=stylesheet><SCRIPT language=Javascript src="FH avec Bush_files/plugin.js"
type=text/javascript></SCRIPT>)

#à ce niveau là, doit normalement afficher les valeurs par classe

#recherche des valeurs de W/D et e/D dans les data:|
```

Line: 50 Col: 52 | INS | NORM | essai code ouverture HTML.py

Mon fichier s'ouvre bien grâce au début de ma procédure, passée en commentaires dans la fenêtre ci-dessus. Cependant, après avoir défini mes classes et mon parser.feed, je ne parviens pas à stocker les valeurs 0.425105 et 6.69456 respectivement sous T/D et W/D comme sur la version HTML qui suit:

Ratio results

***W/D**

6.69456

***Tab of Ratio**

ElemId	T/D
1	0.425105

Ce qui correspond aux lignes du HTML suivantes:

```
<div align=left><H1 id="html"><a name="resultpart">RESULTS</a></H3></div>
<BLOCKQUOTE>
<div align=left><H3 id="html">Ratio results</H3></div>
<BLOCKQUOTE>
<div align=left><B>*W/D</B></div>
<BLOCKQUOTE>
6.69456<BR>
</BLOCKQUOTE>
<div align=left><B>*Tab of Ratio</B></div>
<TABLE class="table" border=1 cellpadding=3>
<TR><TH>ElemId</TH><TH>T/D</TH></TR>
<TR><TD>1</TD><TD>0.425105</TD></TR></TABLE><BR>
</BLOCKQUOTE>
<div align=left><H3 id="html">Bearing Bypass intermediary results</H3></div>
<BLOCKQUOTE>
<div align=left><B>*Bearing Bypass</B></div>
<TABLE class="table" border=1 cellpadding=3>
<TR><TH>ElemId</TH><TH>T/D</TH><TH>Material</TH><TH>Compos</TH><TH>TH Com</TH><TH>TH
```

Lors de l'exécution, j'ai également une erreur au niveau de la définition de mon passer.feed:

- **En-tête: à placer dans le parser.feed(...)**

- Depuis éditeur de texte Windows:

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN">
<HTML><HEAD><TITLE>Composite-Filled-Hole Dossier</TITLE>
<META http-equiv=Content-Type content="text/html; charset=windows-
1252"><LINK
href="FH avec Bush_files/plugin.css" type=text/css rel=stylesheet>
<SCRIPT language=Javascript src="FH avec Bush_files/plugin.js"
type=text/javascript></SCRIPT>
```

- Depuis ouverture sur python depuis linux :

```
[to105478@caefr0p039:~]$ python /home/to105478/essai_code_ouverture_HTML.py
sys:1: DeprecationWarning: Non-ASCII character '\xc3' in file /home/to105478/essai_code_ouverture_HTML.py on line 2, but no encoding declared; see http://www.python.org/peps/pep-0263.html for details
<HTML>
<HEAD>
  <TITLE> Composite-Filled-Hole Dossier </TITLE>
  <LINK REL="stylesheet" TYPE="text/css" HREF="plugin.css" > <SCRIPT
  SRC="plugin.js" TYPE="text/javascript" LANGUAGE="Javascript"></SCRIPT>
</HEAD>
<BODY >
<div align="center"><TABLE border="5" BORDERCOLOR="#C3E3FA" CELLPADDING=25><TR><
TD><font size="5"> Report for Composite-Filled-Hole Analysis</font></TD></TR></T
ABLE></div><BR>
  <div align="right">ISAMI Analyst Vv7.1.1_a350 </div>
  <BR><BR><BR>
  <H3 id="html"> EXECUTION DATA </H3>
  <BLOCKQUOTE>
  DATE OF EXECUTION : 2012-09-14 08:14:07 <BR>
  ANALYSIS TYPE: filledhole <BR>
```

Je ne parviens pas à savoir ce qu'il faut que je mette dans le `passer.feed(?)`.

Y aurait-il par hasard quelqu'un qui pourrait me renseigner ou me venir en aide?

Merci à vous et bonne soirée.