

Subject:

Date:

tst.html

↓ loose web page



tagsoup

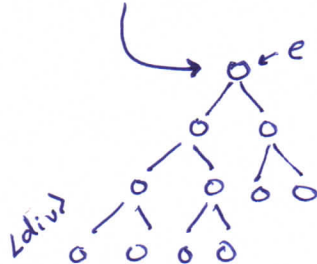


tst.xhtml

dom = parse ()

e.get Elements By Tag Name
('div')

import libxml2



w3schools.com
XML Dom tutorial

use firebug

C

```
sum = 0
for (i=1; i<11; i++)
    sum = sum + i;
```



Python

```
sum = 0
for i in range (1,11,1)
    sum = sum + i
```

```
def sum(n):
    if n==1:
        return 1
    else:
        return n + sum(n-1)
```

```
get_text(e):
    lst = [ ]
    if e.nodeType in (3,+):
        e.data lst.append(e.data)
    else:
```

```
def main(hfn):
    xml = ''
    long_xhtml = html_to_xhtml
    global dom
    dom = parse(long_xhtml)
```

```
        lst = e.childNodes
        for x in lst:
            lst = lst + get_text(x)
    return lst
```

lst = get Elements By Tag Name

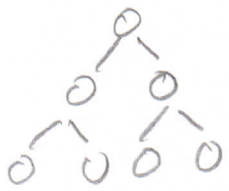
Subject:

Date:

get Ele by Name ('book')

lst = [, , ,]

for x in lst;



e.childnode

lst = e.childnode

lst = get Ele by tag name ('year')