```python
#爬取全书网的每个章节的正文内容
import re
import requests
import os
web_url = 'https://www.xs4.cc/0_4/'  # 选择全书网的某一本书的网址
web_url2 = re.findall('cc/(.*?)/', web_url)[0]  # 取得书的网址的一部分的字符串
req = requests.get(web_url)  # 实例化一个requests请求
req.encoding = 'gbk'  # 编码设置为gbk，防止乱码
shuming = re.findall('<h1>(.*?)</h1>', req.text)[0]  # 取得书名
mulu = re.findall('html">(.*?)</a></dd>', req.text, re.S)  # 获取目录，生成一个列表
wangzhi = re.findall(f'<a href="/{web_url2}/(.*?).html', req.text, re.S)  # 获取书名的网址的一部分，也是一个列表
dict1 = {}
for i in range(9, len(mulu)):  # 从9开始循环，因为0-8是最新章节，之后才是第一章
    dict1[mulu[i]] = f'{web_url}{wangzhi[i]}.html'  # 把目录和网址放到字典中
if os.path.exists(f'd:/{shuming}'):
    pass
else:
    os.mkdir(f'd:/{shuming}')
count=0
for k,v in dict1.items():
    if count==10: #如果不想一次爬取过多，这里设置小一些
        break
    else:
        req_zhengwen=requests.get(v)
        req_zhengwen.encoding='gbk'
        neirong=re.findall('<div id="content">(.*?)
</div>',req_zhengwen.text,re.S)[0] #通过正则抓取正文内容
        neirong=neirong.replace(' ','').replace('<br />','') #过滤不需要的字符
        # print(neirong)
    with open(f'd:/{shuming}/{k}.txt','w+') as file1:
        file1.write(neirong)
    count += 1

```