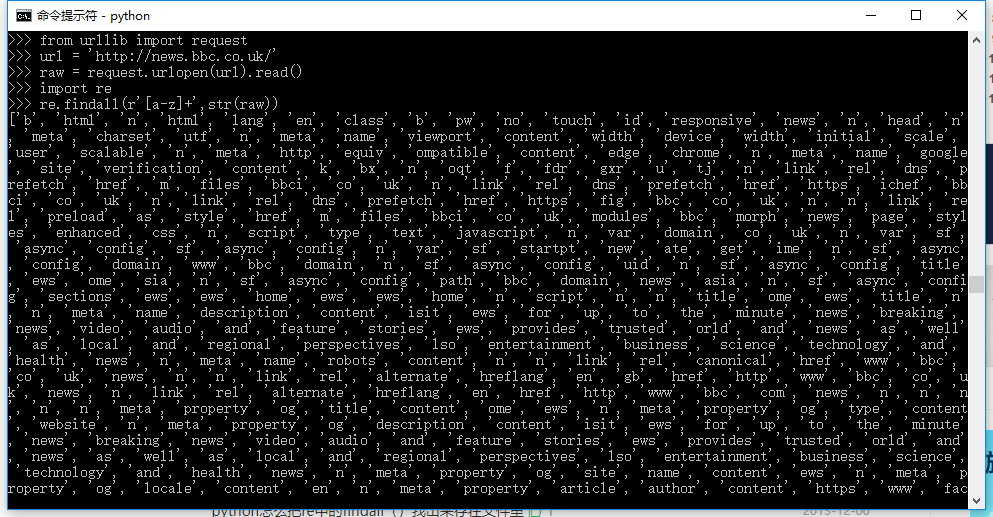
03-22. ◑使用上面建议的正则表达式处理网址http://news.bbc.co.uk/，检查处理结果。你会看到那里仍然有相当数量的非文本数据， 特别是 JavaScript命令。 你可能还会发现句子分割没有被妥善保留。定义更深入的正则表达式，改善此网页文本的提取。

答：





**改进**：添加去掉网页标签的正则表达式，再把提取的结果保存到txt文件

代码：

