

- 1、将要抽取的网页 **HTML** 文件置于 **html** 文件夹下；
- 2、TextExtract_to_TXT.exe 将抽取的正文结果保存为文本文件；
- 3、TextExtract_to_XML.exe 将抽取的正文结果保存为 **XML** 文件；
- 4、clear.exe 将删除所有抽取到的正文结果文件。