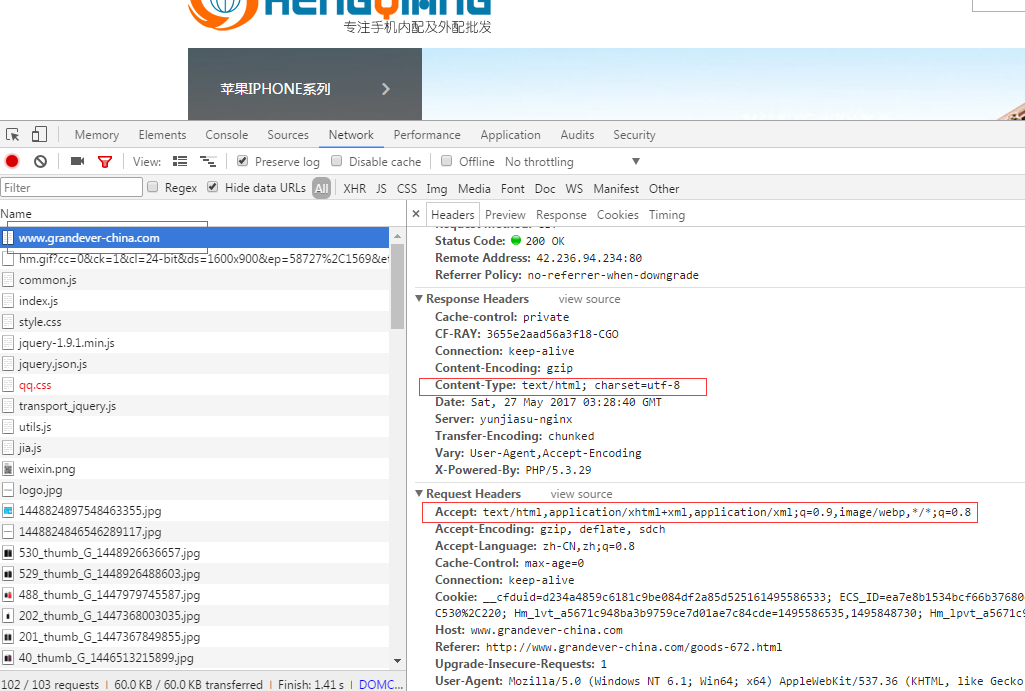
爬取的目的 ： 获取整个网站的图片，以及图片信息，下载到本地文件夹

一：网页传递方法分析



这是个标准的html网页，没有用到post方法，也没有通过ajax返回的json数据。

静态网页 可以用request和beautifulsoup解析网页源码~~~~

二：入口: 首页 <http://www.grandever-china.com/>

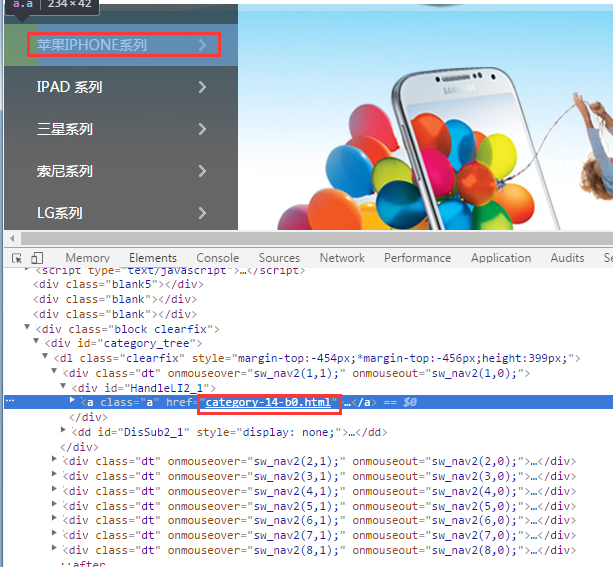
分析首页 ：分两部分 导航栏 + 产品列





我个人感觉 首页的产品列表的连接url无需解析，，因为它 包含在导航栏的连接里产品列表里

第一步获取所有导航栏里系列的链接。



点击 苹果IPHONE系列：



我去解析这个包含所有苹果产品的 全部商品列表页

获取 <a href =”goods-672.html”>这个属性



不过：商品列表页有分页

苹果系列：第二页 分页

<http://www.grandever-china.com/category-14-b0-min0-max0-attr0-2-shop_price-DESC.html>

三星 系列 ：第二页分页

<http://www.grandever-china.com/category-13-b0-min0-max0-attr0-2-shop_price-DESC.html>

构造重要的

newurl：

http://www.grandever-china.com+url+-b0-min0-max0-attr0-{}-shop\_price-DESC.html.format()

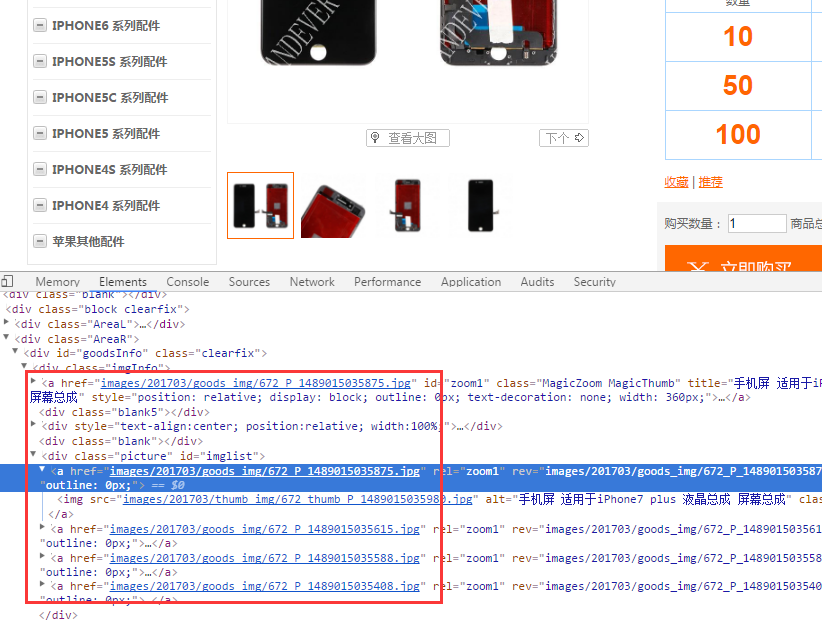
(url是第一步在首页提取的字段)

for url in newurl：遍历提取出url，解析到的所有的图片链接属性：获取 <a href =”goods-672.html”>这个属性，存在这个urllist容器里

貌似需要最后构造一次：imgurl= <http://www.grandever-china.com/>+ [url for url in urlist]

进入单个图片链接页面：

<http://www.grandever-china.com/goods-672.html>，提取下载图片



这个思路里 我需要构造两次，，一次是获取所有分页的商品url，，，，一次是获取所有单页商品的url。