任务名称： 爬虫与解析任务-腾讯证券所有上市企业页面信息

任务编号： TGNB\_PARSE\_10

1. 具体步骤

**沪A深A部分：**

1. 获取附件列表中的股票代码，也就是后面的[dddddd]部分
2. 访问http://stockhtm.finance.qq.com/sstock/quotpage/q/[dddddd].htm#6

点击”公司概况”：



点击“员工组成”：



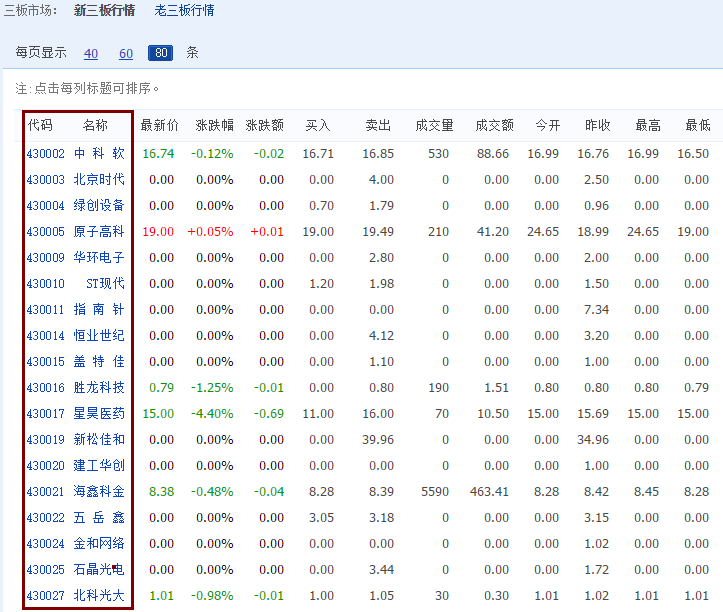
1. 访问http://stockhtm.finance.qq.com/sstock/quotpage/q/dddddd.htm#12



**新三板部分：**

1. 遍历

<http://stockapp.finance.qq.com/mstats/#mod=list&id=three_newplate&module=THREE&type=NEWPLATE&sort=32&page=X&max=80,> X取值[1, 144](含144)



获得代码清单

2.访问http://gu.qq.com/nq[dddddd]/jbzl/gk



3.访问http://gu.qq.com/nq[dddddd]/cwbb



1. 解析要求：

全部页面解析，其中，【所属板块】采用列表方式，【公司简介变动】采取嵌套列表字典方式，即[{},{},{}]。

“员工组成”和“年报摘要”部分只解析最新的部分，见上图红框部分，涉及货币部分全部存为字典并转为人民币元，比如u’经营现金流’: {u’金额’: 1472550000’, u’币种’: u’人民币’}

三、访问要求： 沪深一共3303家，新三板一共11475家，希望一天爬完。

四、整体目标：

沪深一共3303家，新三板一共11475家，希望一天爬完。。