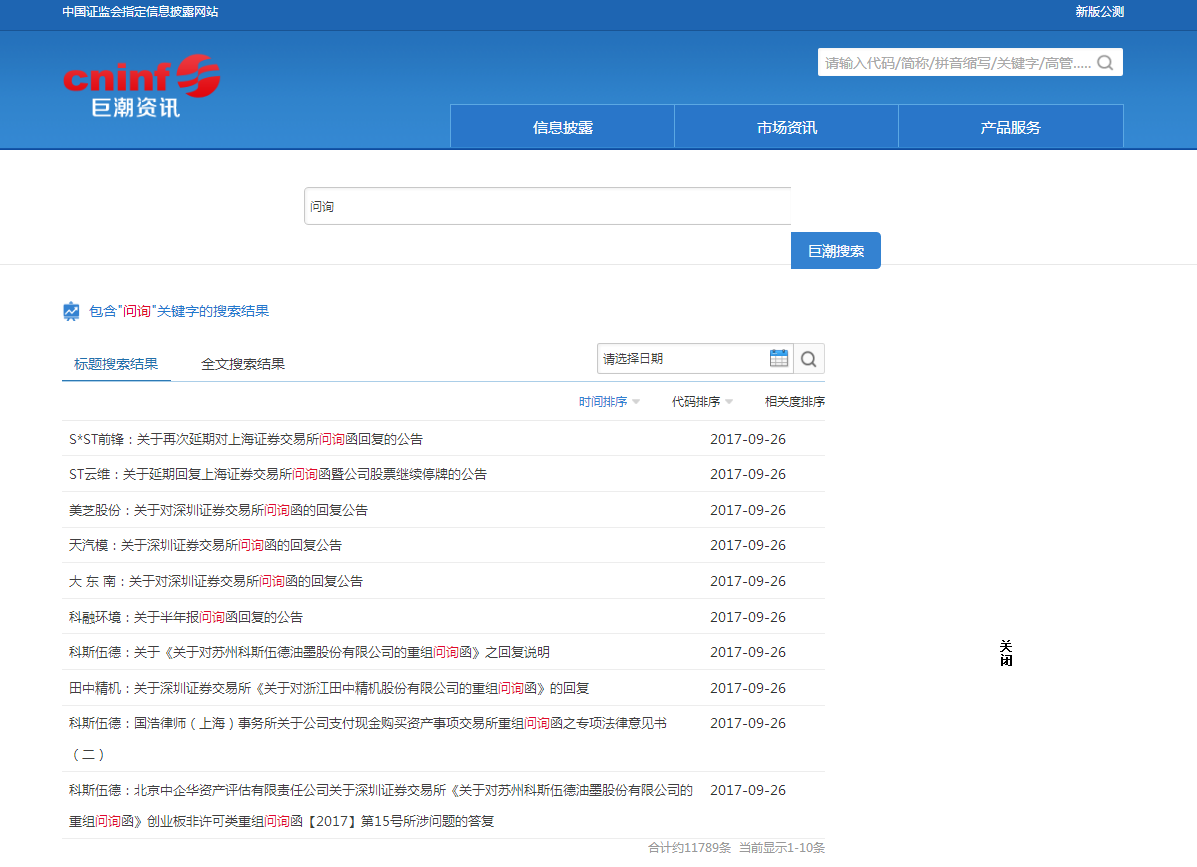
**数据实现1：**A.进入巨潮咨询网（<http://www.cninfo.com.cn/cninfo-new/index>），以“问询”为关键词进行检索，弹出如下界面：



我需要将上述约11789公告全部下载（烦请国庆节下载，这样截止日期为2017-09-30[[1]](#footnote-1)）并保存（文件名为：股票代码+年月日，例如000001\_20170701）按顺序点击任一公告，出现如下界面：

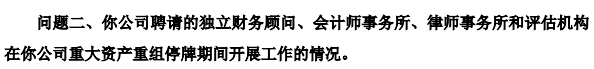


根据上述两个界面，形成如下excel数据格式（证券代码，公告日期，文件大小）。

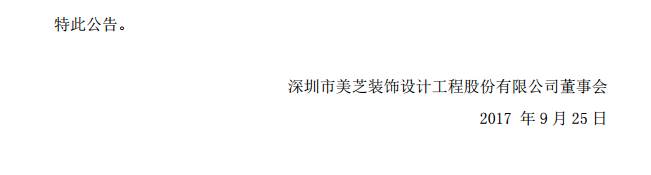
B.针对每一份公告全文进行解析，进一步形成如下数据：证券代码，公告编号，收到问询函时间，公告标题（下图红色下划线书名号内文字），问询问题个数（如实现困难，我可以找人手工阅读），具体相关问题1,2,3,4….。







。。。。。（中间文字部分省略）

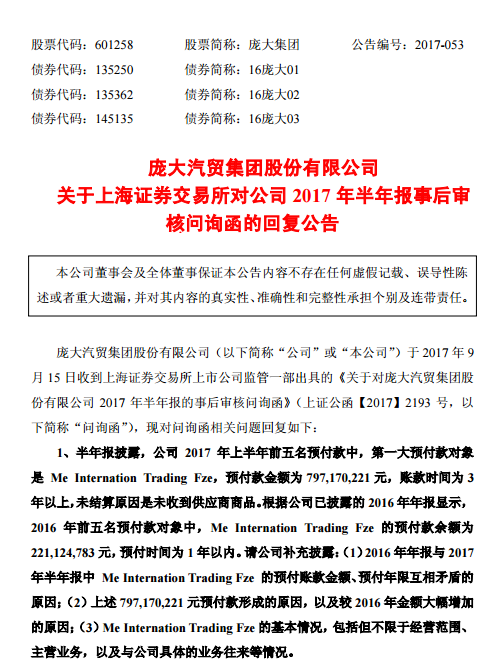


C．以“关注函”为关键词重新进行检索，并重复A.B步骤，并与其结果合并在一起。

以上为独立的第一步，请单独备份为一个文件夹（下同）。

**数据实现2：**A. 进入上海证券交易所官网子网（<http://www.sse.com.cn/> ），以“问询函/关注函”为关键词进行检索，弹出如下界面：



我需要将上述约7366公告全部下载（同样地，烦请国庆节下载，这样截止日期为2017-09-30）并保存（文件名为：股票代码[[2]](#footnote-2)+年月日，例如000001（掌阅科技）\_20170701）按顺序点击任一公告，出现如下界面：

中间部分文字省略。



根据上述界面，针对每一份公告全文进行解析，进一步形成如下数据：证券代码，公告编号，收到问询函时间，公告标题（红色圆圈书名号内文字），问询问题个数（如实现困难，我可以找人手工阅读），具体相关问题1,2,3,4….，公告时间。

**数据实现3：**进入上海证券交易所官网子网（<http://www.szse.cn/> ），路径：信息披露——监管信息公开——问询函件：



按主板/中小板/创业板分别批量下载函件内容及公司回复。函件内容批量下载即可；公司回复请在文档命名添加HF即可（或者回复函单独存放在另一个文件夹）。根据上述界面，针对每一份公告全文进行解析，进一步形成如下数据：证券代码，公告编号，收到问询函时间，公告标题（红色圆圈书名号内文字），问询问题个数（如实现困难，我可以找人手工阅读），具体相关问题1,2,3,4….，公告时间。

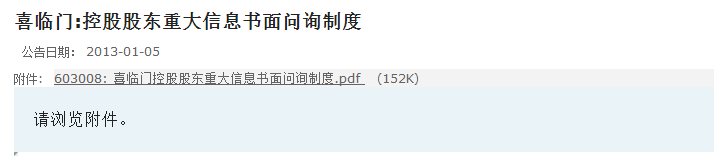
**数据实现4：**如表格“问询公告.xls”所示，点击任意公告标题，



弹出如下界面：

两个步骤：

1. 爬取证券代码，公告时间，公告标题，文件大小。
2. 批量下载相关文件并保存（无需重新命名）。



简而言之，第一，现有各网站披露数据均不全，我因此采用上述四种办法尝试将有问询的公司进行整合，将样本缺失最小化。

第二将问询函内容进行解析，可以将监管部门问题的数量/主要问题（关键词）进行归类。

我个人觉得第一步问题不大，第二步我近期尝试对各类问询人工打样出来。届时再参照我的样板尝试进行解析。

1. 我可能明天五月份会再请你帮我更新数据2017-10-01到2018-05-31.烦请保留程序。 [↑](#footnote-ref-1)
2. 若无股票代码，以公司名称亦可。 [↑](#footnote-ref-2)