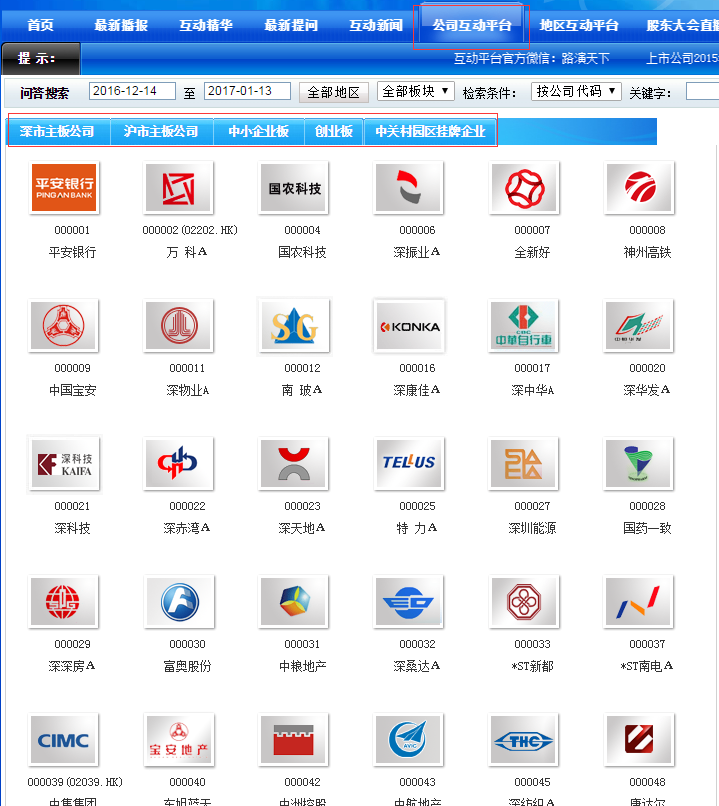
# Ircs爬取分析

主页：<http://irm.p5w.net/gszz/>



## 获取所有的公司名称 index 及股票代码

### 查看index



深市主板公司 395

沪市主板公司 529

中小企业板 823

创业板 574

中关村园区挂牌企业 73

### 查看scrapy request 返回的htm Js 里面



### 把js 改写成 java 保存入 mongoDB

## 分析公司主页company Index



互动精华url

<http://ircs.p5w.net/ircs/interaction/allQuestionForSsgs.do?condition.type=1&condition.stockcode=000001&condition.stocktype=S>

最新提问url

<http://ircs.p5w.net/ircs/interaction/allQuestionForSsgs.do?condition.type=2&condition.stockcode=000001&condition.stocktype=S>

Condition.stockcode 表示股票代码

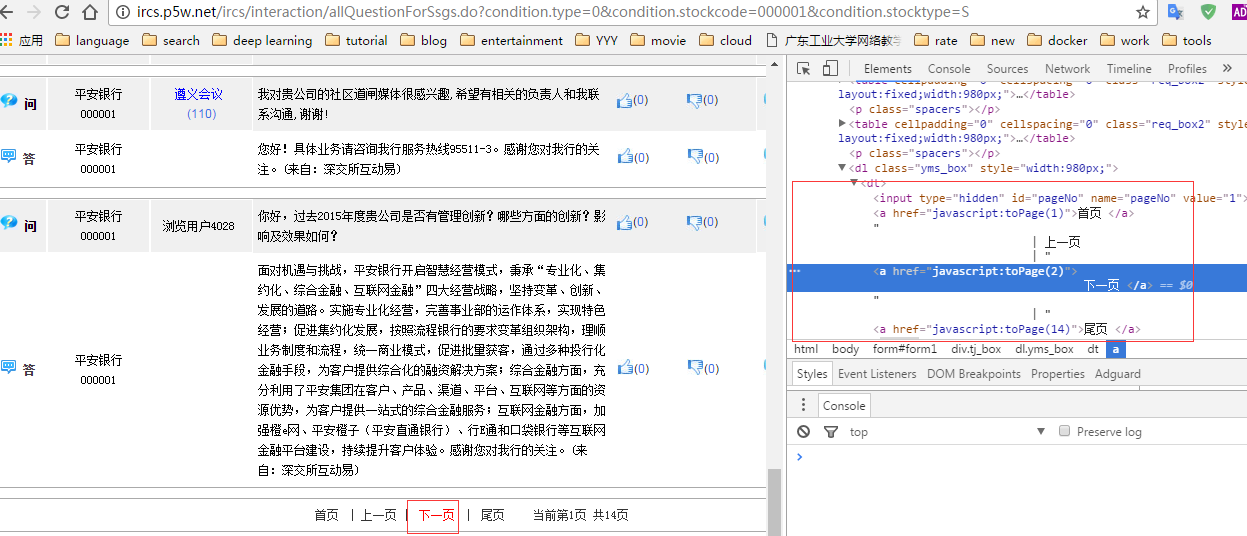
Condition.type = 1 表示 问与答

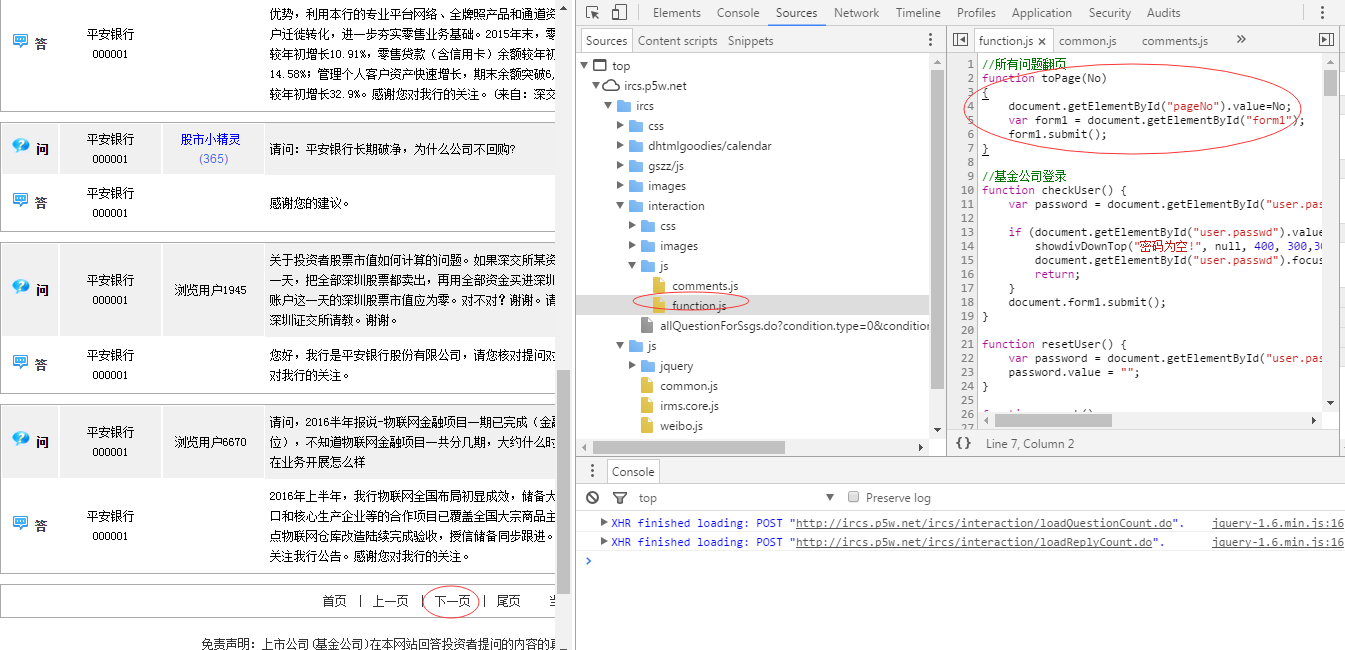
Condition.type = 2 表示 问与没回答

尝试Condition.type = 0 发现为所有的question 包括上面两者

## 在线问答页分析







翻页方式 : 表单上传 post 请求， 参数 pageNO

ircsSpider 保存在线提问数据

## 主题交流活动url分析

**分析 hd\_url index 21637**

db.huodong.find({"hd\_url":/topic/})

db.huodong.find({"hd\_url":/ma/})

db.huodong.find({"hd\_url":/bbs.asp/})

db.huodong.find({"hd\_url":/roadshow/})

db.huodong.find({"hd\_url":/t.rsc/})

db.huodong.find({})

**db.huodong.find({"hd\_url":/topic/})**

<http://ircs.p5w.net/ircs/topicInteraction/bbs.do?rid=17805>

9002 个

db.huodong.find({"hd\_url":/topicInteraction.\*bbs.do/})

6178

db.huodong.find({"hd\_url":/topicInteraction.\*video.do/})

2813

db.huodong.find({"hd\_url":/topicInteraction.\*bbsFor/})

8

8999

**db.huodong.find({"hd\_url":/ma/})**

<http://irm.p5w.net/ma/2016/sdqd/01/index.htm> 接待日介绍 无BBS讨论

4085 个

**db.huodong.find({"hd\_url":/bbs.asp/})**

1965

分两种

db.huodong.find({"hd\_url":/newzspt.\*bbs.asp/})

db.huodong.find({"hd\_url":/zsptbs.\*bbs.asp/})

<http://newzspt.p5w.net/bbs/bbs.asp?boardid=3945> 1100

<http://zsptbs.p5w.net/bbs/chatbbs/bbs.asp?code=601872&web=&selcode=679> 865

**db.huodong.find({"hd\_url":/roadshow/})**

<http://roadshow2008.p5w.net/2009/lpyl09/yh.asp>

5067

要跳转到 网上交流 <http://roadshow2008.p5w.net/2009/lpyl09/bbs/bbs.asp>

624 .asp

分两种

Yh.asp bbs/bbs.asp 460

Yhcq.asp 加bbs.htm 151

**db.huodong.find({"hd\_url":/t.rsc/})**

1420

找到rid

yhcq.htm 140

<http://ircs.p5w.net/ircs/topicInteraction/bbsForRoadshow.do?rid=15466>

yh.htm 471

<http://ircs.p5w.net/ircs/topicInteraction/bbsForRoadshow.do?rid=4692>

index.htm 808

<http://ircs.p5w.net/ircs/topicInteraction/bbsForRoadshow.do?rid=7022>

other

102

db.huodong.find({"hd\_url":/bbsForRoadshow/})

<http://zsptbs.p5w.net/bbs/chatbbs/left.asp?boardid=1153&pageNo=2>



qr\_topic/rsc 542.9k json

qr\_bbs\_new 70.7k xml

qr\_bbs \_zs 146.1k html

qr\_road 140.1k xml

### hd\_url 种类(总**21637条**)

A类：

<http://ircs.p5w.net/ircs/topicInteraction/bbs.do?rid=15971> 9002

<http://irm.p5w.net/ma/2016/sdqd/01/index.htm> 4085

<http://irm.p5w.net/rsc/2016/300032/02/index.htm?rid=17212> 1420

B类：

<http://zsptbs.p5w.net/bbs/chatbbs/bbs.asp?code=601872&web=&selcode=679> 865

C类：

<http://newzspt.p5w.net/bbs/bbs.asp?boardid=3945> 1100

D类：

<http://roadshow2008.p5w.net/2009/lpyl09/yh.asp> 5067

A类：

eg: <http://ircs.p5w.net/ircs/topicInteraction/questionPage.do>

POST { pageNo, rid}

返回JSON 数据

B类：

eg:<http://zsptbs.p5w.net/bbs/chatbbs/left.asp?boardid=1153&pageNo=2>

GET { pageNo, boardid}

返回HTML数据

C类:

<http://newzspt.p5w.net/bbs/question_page.asp?boardid=3763&bbs=1&pageNo=1>

POST or GET

返回XML数据

D类:

hd\_url + ‘question\_page.asp?’

POST {pageNO}

返回XML数据