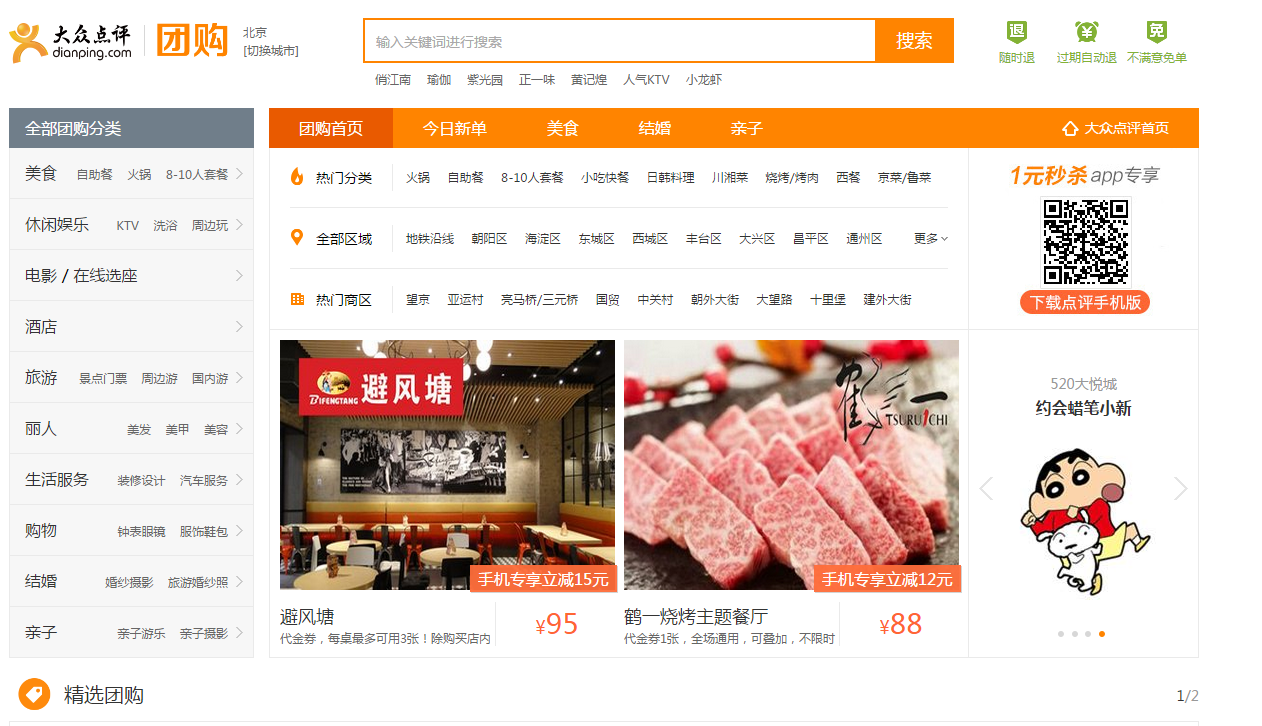
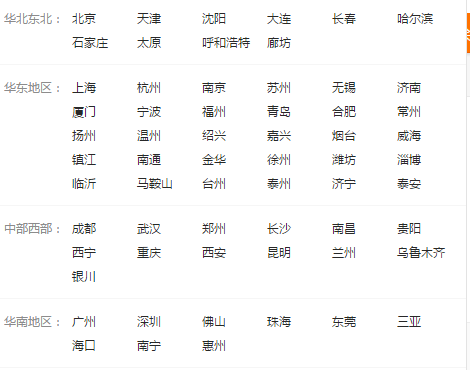
团购食品图片数据采集说明

1. 需要采集数据的网站：http://t.dianping.com/

进入后页面如下

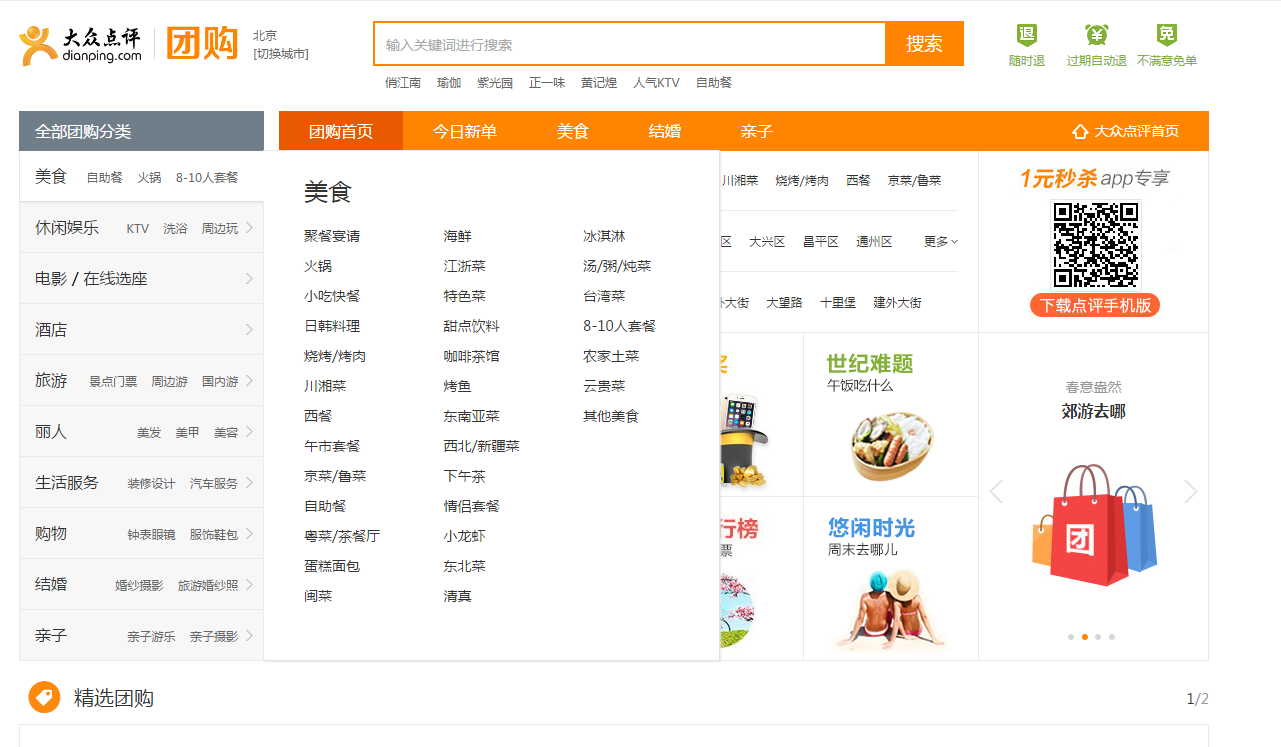


1. 我们需要按照城市进行采集，**城市列表**如下，采用之前大众点评所用的城市。



***切换城市可以在域名后直接加城市拼音，例如http://t.dianping.com/beijing***

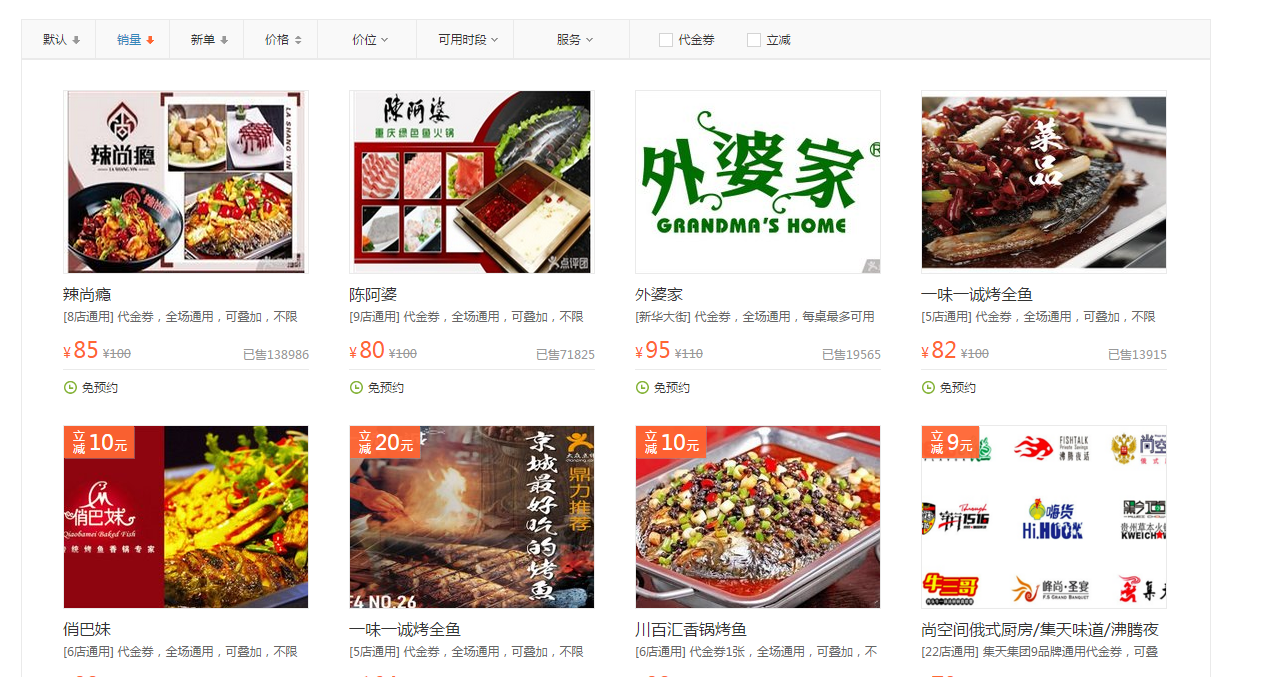
1. 选择“北京”，在“美食”标签下面出现菜系分类



1. 每个菜系都需要采集, 除了“其他美食”。点击如“烤鱼”



1. 我们选择按照“销量”排序 (**下载前100个团购的图片**）。（有小部分团购会同时属于几个菜系，例如同时属于火锅和聚餐宴请，这个先不管。菜系下有的就下）



1. 点击其中一个结果：“ **辣尚瘾“**



需要采集内容如下：1）团购ID（例如网址http://t.dianping.com/deal/10686119）； 2）评分信息栏；3）团购详情（各个团购详情格式不太相同，目前先把文字全部抓取下来， 放在一个文本文件，按照原来的分词和分段）4）产品介绍（文字和图片，文字单独放在一个文本文件，按照原来的分词和分段。图片按顺序命名为1,2,3,4…）



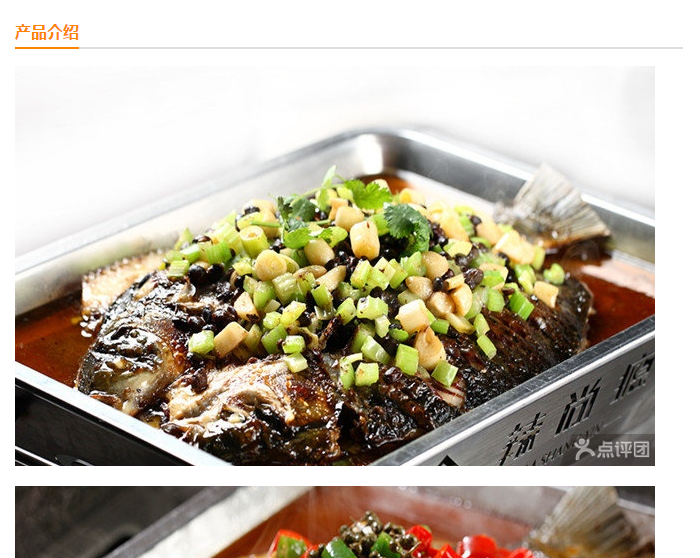
**1**



**2**



**3**



**4**

生成的目录结构可以如下：

--城市1

--菜系1

---团购1

---产品介绍

---团购详情

---团购2

---团购3

--菜系2

--菜系3

--城市2

--城市3

…

…