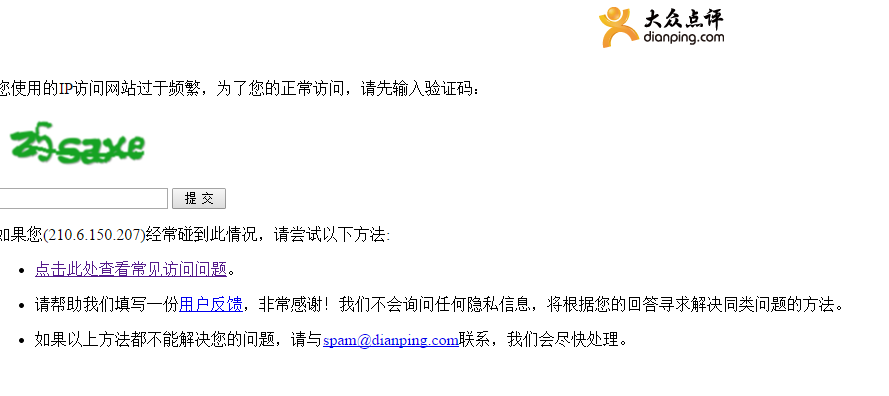
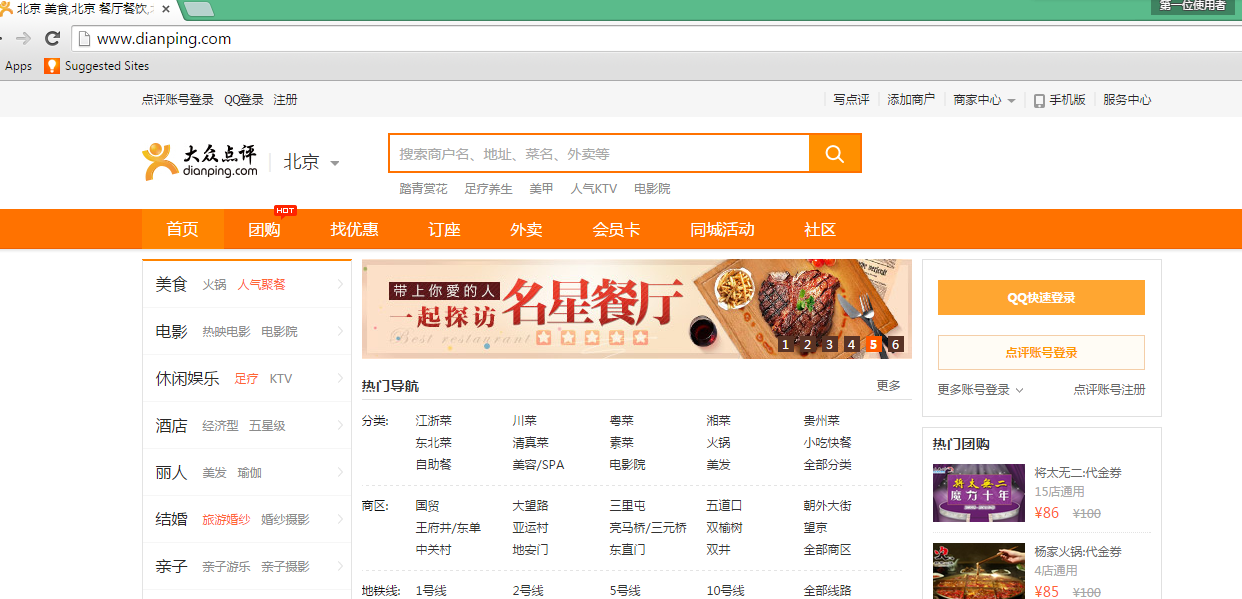
食品图片数据采集说明

1. 需要采集数据的网站：<http://www.dianping.com/>

(\*发现该网站如果访问频繁会要求输入验证码)



进入后页面如下



1. 我们需要按照城市进行采集，选择城市



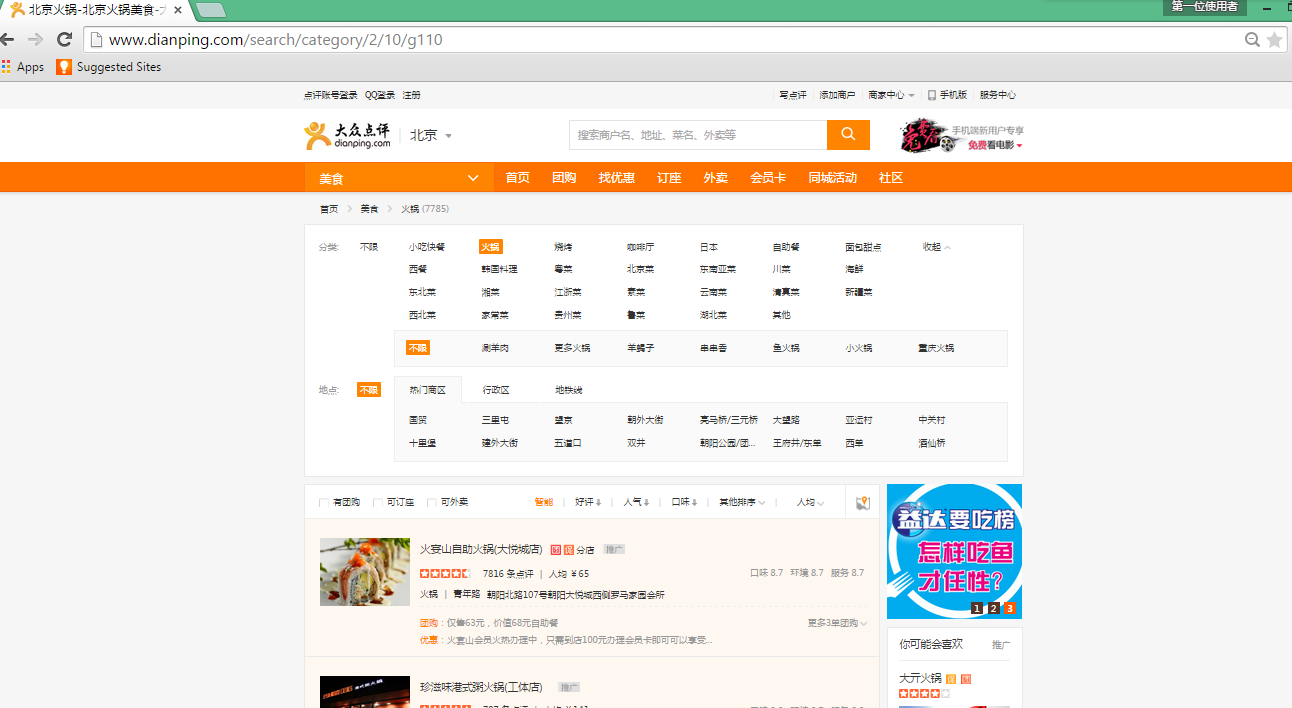
*主要采集大陆城市（大概62个），港澳和国外地区由于数据质量参差不齐，暂时不考虑*

**我浏览了多数大陆城市，发现网页结构相同，因此下面以北京为例，其他城市也是同样采集**

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”。点击如“火锅”



1. 我们选择按照“人气”排序 (**下载前100的店的图片**），大概估计总共有150-200万图片。



1. 点击第一个结果：“**正锅老灶重庆老灶火锅“**

需要采集内容如下：1）位置信息栏； 2）评分信息栏；3） 推荐菜的菜名、图片、和属性标签（有些有，没有则忽略）；4）环境图片和图片名称



**2**

**1**



**3**

下载全部6张图片，命名为 ”九宫格麻辣锅1，2，3，…,6”



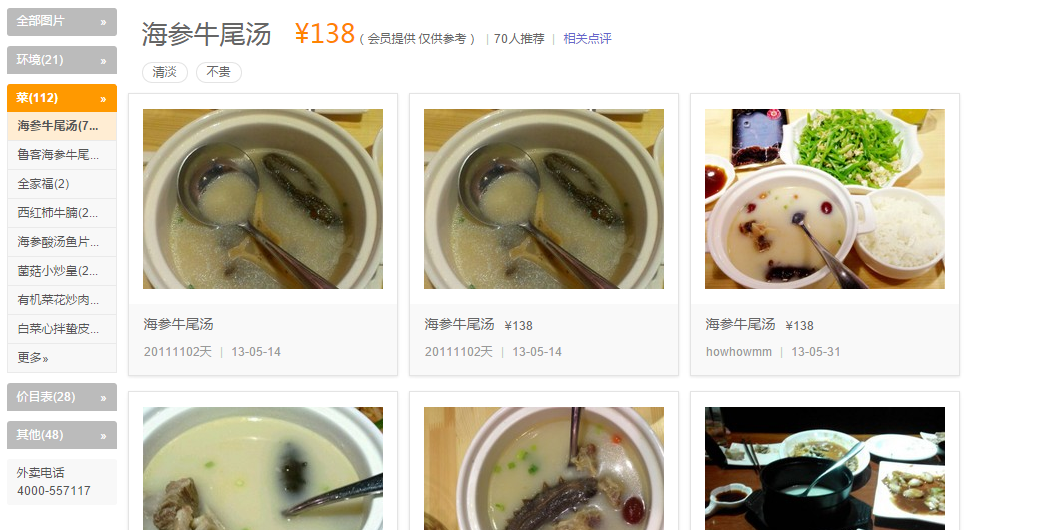
**3**

另外举一个属性的例子。属性标签示例如下：

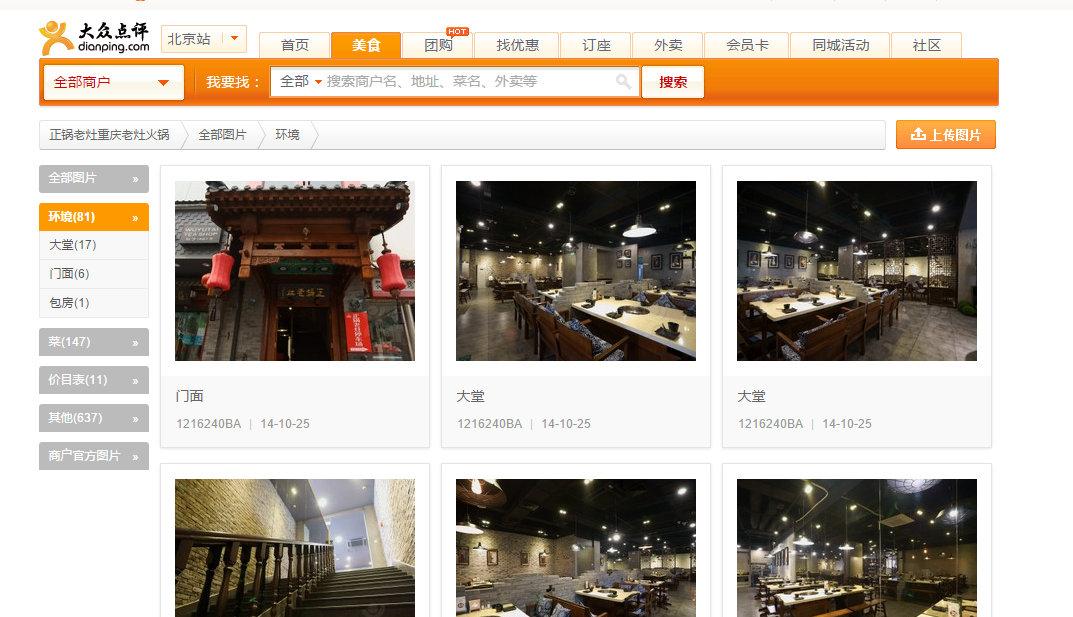


**3**

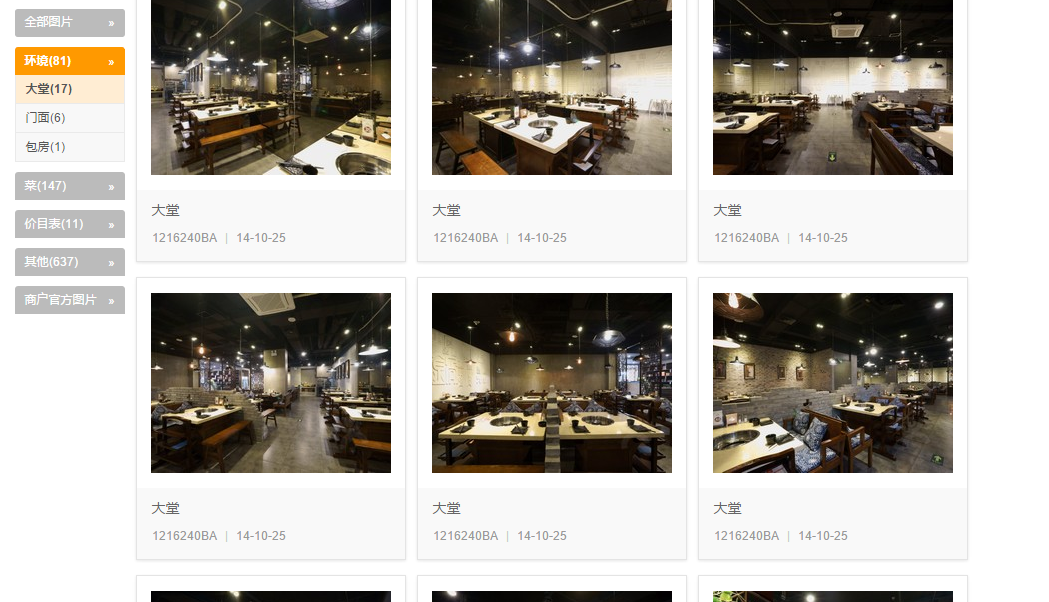
也可以通过该店的相册访问：



环境图片（仅下载有标签的图片，如大堂，门面，包房。不需所有图片都下载）



**4**



**4**