

如何“攻破”大众点评的文字加密防线？

zuobangbang CSDN 2018-12-20



点击“上方蓝字”关注CSDN



© 视觉中国

作者 | 左伊雅

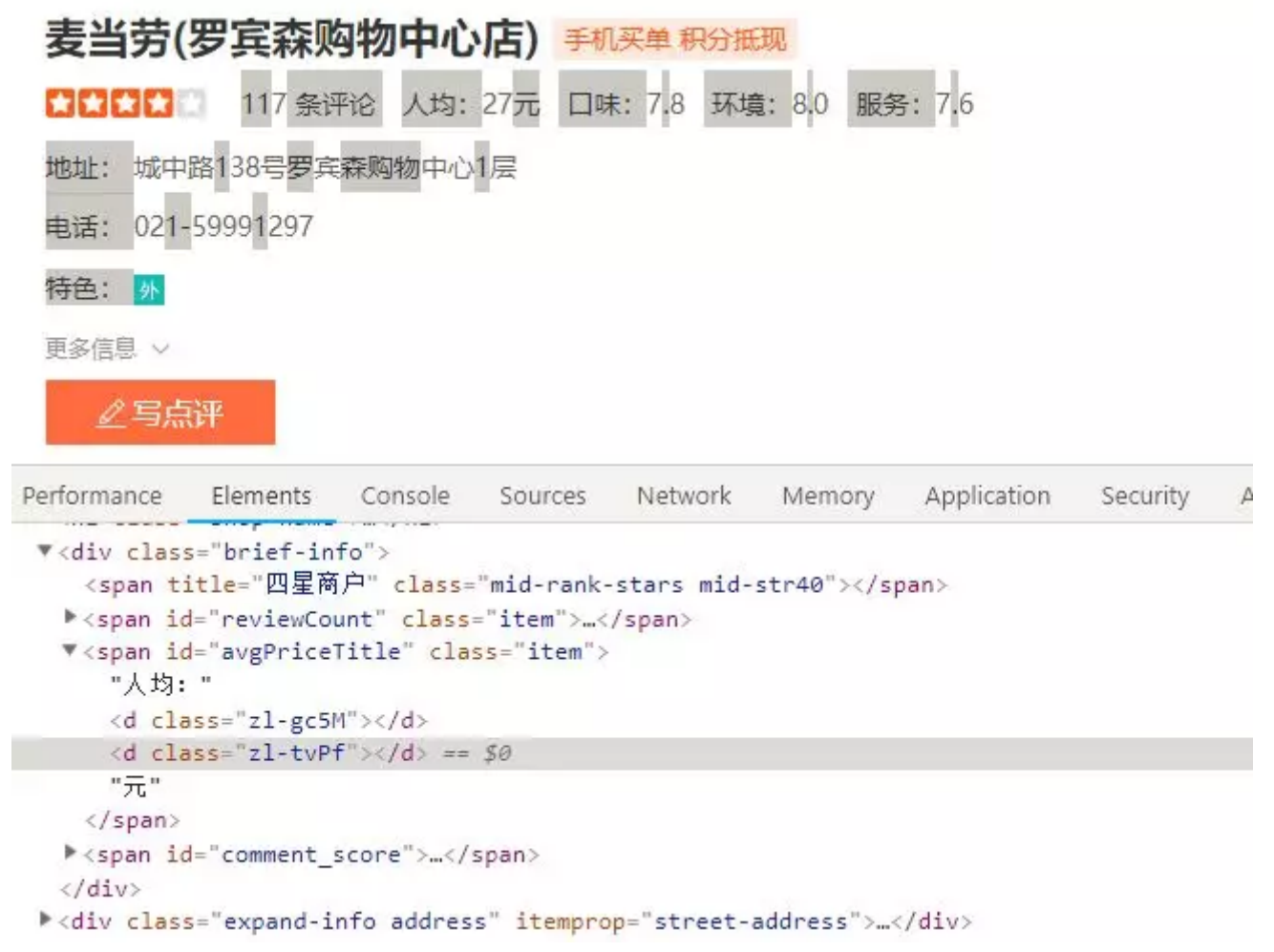
责编 | 郭芮

相信在座的各位一定有爬过大众点评的数据，但是却因为它的各种加密力不从心。这里必须告诉你，看完这篇文章，大众点评加密你就能轻松搞定。

现以麦当劳（罗宾森购物中心店）为例进行说明：
<http://www.dianping.com/shop/77335766>。



从上图就可以看到，7，2，8，0，6，城，中，路等等都无法被选择，再打开开发者工具定位，它的人均27元中的7对应.zl-tvPf；2对应.zl-gc5M。这就是大众点评的加密，将部分文字用标签代替。



文字加密的方法有两种：一种是映射，一种是图片。而大众点评用的正是后者——这是怎么看出来的呢？

大家仔细看一看下面的两幅图，笔者分别选择了.zl-gc5M；.zl-tvPf。观察一下两幅图的右侧（也就是两个标签对应的style），它们的右上是d[class^="zl-"]；右下分别是.zl-gc5M 和.zl-tvPf。

```

"人均: "
<d class="zl-gc5M"></d> == 50
<d class="zl-tvPf"></d>
"元"
</span>
span id="comment_score"></span>
</div>
.v class="expand-info address" itemprop="street-address"></div>
<class="expand-info tel"></p>
.v class="promosearch-wrapper"></div>
<class="J-unfold unfold" data-click-name="更多信息"></a>
.v class="other J-other Hide"></div>
.v class="action"></div>
iv id="tuiguangAd-wrapper"></div>
</div>
}
d[class^="zl-"] {
width: 14px;
height: 30px;
margin-top: -9px;
background-image:
url(//s3plus.meituan.net/v1/mss_0a06a47.../svgtextcss/ab22723....svg);
background-repeat: no-repeat;
display: inline-block;
vertical-align: middle;
margin-left: -6px;
}
.zl-gc5M {
background: -36.0px -7.0px;
}

```

点击可放大图片

现在来分析右上的 `d[class^="zl-"]` 部分。这里的意思是说，以 `zl-` 开头的映射对应的 `background-image` 为：
`url(//s3plus.meituan.net/v1/mss_0a06a47.../svgtextcss/ab22723....svg)`——这个URL就是数字图片的地址，将这个链接打开会看到这样一组数字图片：



再来看右下 `.zl-gc5M` 和 `.zl-tvPf` 部分。`background: -36.0px -7.0px;`
`background: -106.0px -7.0px;` 懂一点CSS的都知道，这是图片平移，分别往左平移36个像素点和106个像素点。回到上副图，将数字2往左平移36个像素点，把数字7往左平移106个像素点，数字2和7就可以在网页上显示出来。这就是文字图片加密的原理。

找到对应的CSS文件，建议不要格式化，全部复制存为字符串。这里要说明一下，大众点评把所有文字、数字的平移放在同一个CSS文件里，所以这个CSS格式化以后有19301行。

数字的分类字母为‘zl-’：



```

c91d98607cad828...ea5c4cf1c7.css* X
11841     background: -420.0px -246.0px;
11842 }
11843
11844 .ku-zRay {
11845     background: -0.0px -756.0px;
11846 }
11847
11848 .fu-0d1I {
11849     background: -168.0px -1687.0px;
11850 }
11851
11852 .fu-JQ15 {
11853     background: -252.0px -127.0px;
11854 }
11855
11856 .fu-TL1X {
11857     background: -336.0px -517.0px;
11858 }
11859
11860 .fu-FEIE {
11861     background: -434.0px -2377.0px;
11862 }
11863
11864 .fu-ONHv {
11865     background: -14.0px -1777.0px;
11866 }
11867
11868

```

数字的分类字母是 'zl-', 从上面的CSS文件里将 'zl-'调出:

```

zl-tvPf{background:-1060px -70px;}
zl-Jvp2{background:-220px -70px;}
zl-giSW{background:-780px -70px;}
zl-Cg3x{background:-1340px -70px;}
zl-gc5M{background:-360px -70px;}
zl-htaN{background:-640px -70px;}
zl-TohQ{background:-500px -70px;}
zl-JTyc{background:-1200px -70px;}
zl-FhcV{background:-80px -70px;}
zl-Bthl{background:-910px -70px;}

```

可以看到不同的class只有左右偏移量不同, 最小偏移量为80Px, 最大的偏移量为1340Px。这是一个公差为140的等差数列, 所以整体思路是: 根据平移的多少来和数字串'9426831705'匹配, 即偏移量为80对应数字9, 偏移量为220对应数字4.....这里要说明一下的是偏移量为910的zl-Bthl, 它对应的是数字1, 但是大众点评的1是没有加密的, 就把数字1对应的真正的偏移量920改为了910。这里对整体是不影响的。

```

def number():
    #正文不能超过50000字, 这个q先删了, 小伙伴自行去css文件复制即可
    q = q.split('.')
    number = '9426831705'
    data,changedic,s ,w,j,p= {},{},{},'',0,0
    for i in q:

```

```

w = w + i
if j == 3:
    if 'zl-' in w:
        #print(w)
        data[int(re.findall(r'background:-(.*)px', w)[0])] = w.split()
        s.append(int(re.findall(r'background:-(.*)px', w)[0]))
        w, j = '', 0
    j = j + 1
for i in sorted(s, reverse=False)[: -1]:
    changedic['<d class="{0}"></d>'.format(data[i])] ,p= number[p],p+1
print(changedic)
return changedic

```

这是匹配的结果：

```
{'zl-FhcV': '9', 'zl-Jvp2': '4', 'zl-gc5M': '2', 'zl-TohQ': '6', 'zl-htaN': '8', 'zl-giSW':
'3', 'zl-Bthl': '1', 'zl-tvPf': '7', 'zl-JTyc': '0', 'zl-Cg3x': '5'}
```

找了美食、健身馆、游泳馆、按摩、景点五种网页进行实验，所有的数字都成功地识别出来了。

```
{'tel': '021-59991297', 'review': '117', '人均': '27', '口味': '7.8', '环境':
{'tel': '021-64161059', 'review': '1326', '人均': '340', '技师': '9.3', '环:
{'tel': '16621127940 \xa0 17321127275', 'review': '133', '人均': '755', '水
{'tel': '021-67734074', 'review': '8183', '费用': '77'}
{'tel': '13611752158', 'review': '66', '人均': '', '设施': '9.1', '环境': '9

```

这里把最核心的代码贴出来，完整版会上传到GitHub。

```

def change(dic,text):
    for key,value in dic.items():
        text=text.replace(key,value)
    return text

def score(changedic,urls):
    for url in urls:
        inf={}
        inf['tel']=re.findall(r'</span> (.*) </p>',change(changedic,str(sou
        inf['review']=re.findall(r'> (.*) 条评论',change(changedic,str(sou
#不同的网页结构不同，所以加了两个try/except使代码适用于全部网页
        try:
            inf[re.findall(r'>(.*) 元',change(changedic,str(soup.find('spa
        except:
            inf[soup.find('span', id='avgPriceTitle').text.split(':')[0]]=
        try:

```



```
soup = soup.find('span', id='comment_score').find_all('span', <
for i in soup:
    k = re.findall(r'<span class="item">(.*?) </span>', change
    inf[k.split(':')[0]] = k.split(':')[1].strip()
except Exception:
    continue
finally:
    print(inf)
```

至此，大众点评的数字解密就讲完啦。



看到这里的小伙伴都很优秀噢，接下来讲的会比数字难很多。Are you ready?

现在开始看评论，以（<http://www.dianping.com/shop/77335766>）的第一条评论为例：



果果子Miku Lv4

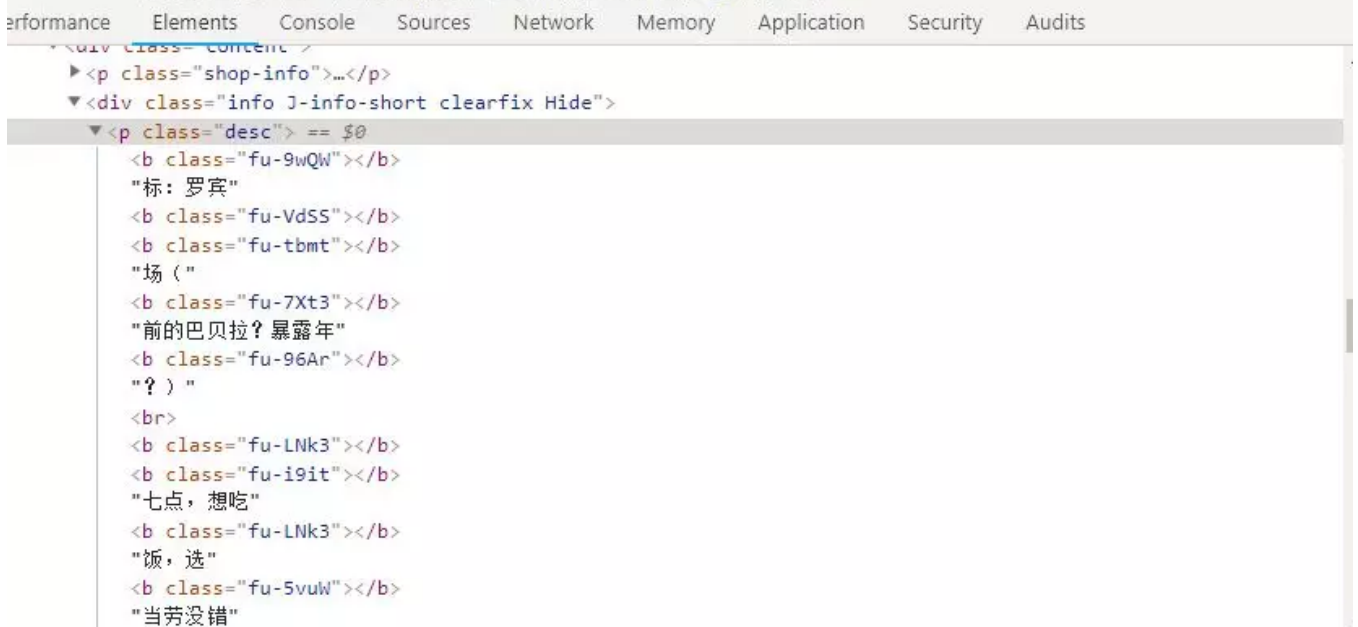
★★★★★ 人均: 16元

坐标: 罗宾森广场 (以前的巴贝拉? 暴露年纪?)

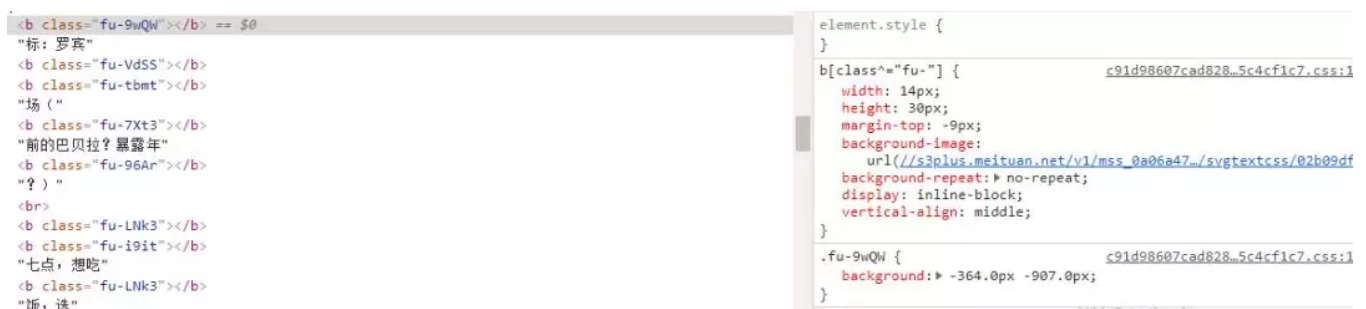
早上七点, 想吃早饭, 选麦当劳没错[馋]上下两层, 楼下只卖冰淇淋🍦 (早上没有的), 二楼点餐, 座位蛮多的, 随便找了个座位~

选了最爱的猪柳蛋汉堡配豆浆¥16! 口味就是连锁店的味道 (都一样, 没有任何差别) ~

最近推送的新品看着就好想吃[笑哭]但是早上木有[笑哭]那个鸡排以前吃过, 超级香的! 新出的冰淇淋颜色超好看, 名字好绕口好奇怪! 下次有机会再来打个甜品的卡吧[机智][机智][机智]



可以看到评论里的部分文字都被替换成了标签, 显示出的文字完全是读不通的, 和数字一样, 它有一张文字图 (background-image), 通过移动各个文字的位置 (background) 将文字显示出来。



打开 (background-image) 的URL, 可以看到有很多文字, 大众点评所有加密的文字都在这里啦。和数字一样, 是通过将单个字往左, 往上平移显示出来; 找到相应的CSS文件, 再将文字与标签匹配。



患大骤怠芒孔苹服戒举稼狄领疾蕉彝橘吸醉姓杀姐呈未暖脏涂嫁众巴厦库
期榜责遍炉饶毡弓太冀残饕晒零造训掙浑朝竹惨戏制剩枪
猛响章隔甲咏帆窗御江诊屈搬锅秋劈窠卧桐茅展街稍即辱纹宽愧祸奖再
壮龟绩卧悠首其爹员牙治净伴痰陈控匙符键虎舞怨敲爷紧争犹严亚灯尽扰回惭
庆识懂效锈泻此趣锁帅彩源捞副俯囊针久滑州俱竟筛慧提体豆配幕桂济卓茂辈杆甜占萝翠儿骨
踢餐靠刊絮包殖吴肺假们傻港坟鸡嫩僻九眯沸旦虽
寻痕陷慰止沟党巾检丑狂益素艇川拉特荒挖巨顾愚阁暴条吓蛛兵趁
周攻附授认昏鼓摩秀邪恋赏戴牵禾宇变膝架叹价份搞护忧闯杆沈股距胜妇尘
宋旷刃会瓶备买车绕捧耍菠干腰艰市掏床孙弄束抄较意朵宙钥
出滔因秩木眼改茫勉峡撞萝骄由帐汇昼汁承校空踪叠档鸽现侵九储镰编极缩血两播闭扛夏恢
桃规域咱节拍粗臆皇穷追象礼膀莖善仇锯捷沃店童厚幻入任愈绢得早肢勾驾恰滚蚂颞
盐称税恨揽亲歌楼柱织访灿怀温脑倘杂牺伍粒光组貉挑智类队候剖登穴瓣蝶升虹贪萍下欣
俩已鸟咬鹄恼裁返逃肢亮肿嘉疫纱水驴央危威一贺耗贤侵道恒辽绒真闷雁膝虐辜
促第声榨松资钻斩岁共银告锄缘欧庙秃截慨淡咽衫二
颈锤饺更对战芳粟材强疆勇斑过人果慨英站钳乳鹈历卷丁式迈住适毕核蒸区嫖够爆顷李民笑徽磨
笔友虫蓬厂卫龄贞午淹低利榴谨于屠葛值斯歎手事火名暑
椅殊艳祖枕八鑫悉丙勾未骂花闹昌负贩忍拜乞次绍顶碎帆款遗
恶丢羹丰勒描后倚隶棋力背军院零丽渐断知扔担忠歇四俊勾桶寄圣雕番悲维普焦莫谢退叫
切号赌级腔陕肤梳晴螺刻兼吞含械蜡情栋援射雪便盼狸触葵诱炸纹归密
珍脊艺旗界了燥闹喘偷秤哥副衡小凶使脖喉虚女饥鲜夕孟忘记烘可兰
航腑噪裂喷舰罗位扎逆纵直轧侮余略葡犁膛丹爪叛妄林健法夹夺混赠料货连宏摊结旋
起寒剑杜治群续堆耀替胞肌龙势渴重浇绘传里别君挡伞脸竟

仔细看下图，第一个数字对应的左移，第二个数字对应的上移，下面把左移称为第一组数据，上移称为第二组数据。

第一组是以0为首项，140为公差的等差数列，最大为5740，对应第42列；第二组是以70为首项，公差为300的等差数列，最大值为23770，对应第80行。整个思路就是做一个循环，一行一行的匹配，第一行对应70px，第二行370px，依此类推。在每一行中，根据它的第一组数据（左移量）从小到大排列，和每行的文字一一对应。


```

fu-jf4L {background:-4480px -11470px;}
fu-yBH4 {background:-700px -20470px;}
fu-rzDW {background:-00px -7570px;}
fu-GLS2 {background:-2940px -9370px;}
fu-B38y {background:-420px -13570px;}
fu-Deav {background:-4760px -6670px;}
fu-GU69 {background:-2660px -7570px;}
fu-VjCn {background:-140px -15370px;}
fu-mdzS {background:-980px -5170px;}
fu-qsfF {background:-3080px -13870px;}
fu-lf5Z {background:-2520px -21970px;}
fu-SW1F {background:-2240px -15370px;}
fu-4yPu {background:-4480px -10870px;}
fu-rFEq {background:-840px -11170px;}
fu-iX2T {background:-840px -7270px;}
fu-8pc3 {background:-1820px -9370px;}
fu-vWzu {background:-2380px -12970px;}
fu-DvOs {background:-280px -6370px;}
fu-eOYy {background:-1400px -11770px;}
fu-kvyq {background:-3080px -12070px;}

```

注意一下，它的换行表达为
以及空格对应\xa0。所以在匹配的时候，用。代替
，用；代替\xa0。即result['
']='。'；result['\xa0']='；'。匹配的结果：

```

class="fu-BGr2">。 杆 class="fu-Xner">。 甜 class="fu-9qu3">。 占 class="fu-FzHk">。 萝 class="fu-a8ge">。 草 class="fu-2hUo">。 儿 class="
class="fu-eAPp">。 骨 class="fu-s7bD">。 踢 class="fu-0TSh">。 餐 class="fu-hcHu">。 靠 class="fu-tXj4">。 刊 class="fu-o9EX">。 絮 class="
class="fu-K7ks">。 包 class="fu-e8IK">。 痘 class="fu-96Fy">。 吴 class="fu-qXi">。 肺 class="fu-jsu0">。 假 class="fu-afhb">。 们 class="
class="fu-5Vej">。 俊 class="fu-Qq80">。 港 class="fu-h1lj">。 坟 class="fu-wZTi">。 鸡 class="fu-6ge9">。 梢 class="fu-Wqet">。 僻 class="
class="fu-Cbtu">。 丸 class="fu-CSql">。 胀 class="fu-Faj9">。 薄 class="fu-ZdgK">。 旦 class="fu-IWIR">。 虽 class="fu-PcPe">。 寻 class="
class="fu-5z5U">。 痕 class="fu-B8xy">。 恼 class="fu-Y5tX">。 慰 class="fu-tjgH">。 止 class="fu-Zzfb">。 沟 class="fu-CaGL">。 觉 class="
class="fu-ah5q">。 巾 class="fu-05Ie">。 榆 class="fu-EDM7">。 丑 class="fu-0mcp">。 狂 class="fu-s8td">。 益 class="fu-tyCd">。 素 class="
class="fu-8eVr">。 蟹 class="fu-ZTV1">。 川 class="fu-0CS5">。 抬 class="fu-QUgW">。 特 class="fu-6Zr0">。 荒 class="fu-W0nq">。 挖 class="
class="fu-2mYu">。 巨 class="fu-gmI1">。 顾 class="fu-4v7F">。 愚 class="fu-BDF2">。 鄙 class="fu-8YqS">。 累 class="fu-9CQ1">。 奈 class="
class="fu-GpG">。 吓 class="fu-GMSz">。 蛛 class="fu-WyqW">。 兵 class="fu-FJds">。 赴 class="fu-BzEj">。 周 class="fu-7GTh">。 攻 class="
class="fu-uPh">。 附 class="fu-tae6">。 搜 class="fu-m0Hf">。 认 class="fu-aYkv">。 管 class="fu-vv76">。 鼓 class="fu-YCra">。 摩 class="
class="fu-Wg01">。 杆 class="fu-QQ00">。 邪 class="fu-uluc">。 恋 class="fu-Ytha">。 冀 class="fu-zdpl">。 戴 class="fu-stbn">。 牵 class="
class="fu-Pokn">。 禾 class="fu-u8W8">。 宇 class="fu-j0tY">。 宴 class="fu-BR02">。 膝 class="fu-Tv7B">。 架 class="fu-5ji0">。 叹 class="
class="fu-9uDe">。 价 class="fu-ah8c">。 时 class="fu-0hm7">。 播 class="fu-B21e">。 护 class="fu-m05q">。 忧 class="fu-Led2">。 闷 class="
class="fu-1CaH">。 杆 class="fu-bdM0">。 沈 class="fu-s98t">。 般 class="fu-PAM4">。 距 class="fu-v0LH">。 脏 class="fu-ch59">。 妇 class="
class="fu-193p">。 空 class="fu-lyZv">。 宋 class="fu-6Zpt">。 矿 class="fu-bd3B">。 刃 class="fu-v0tY">。 舍 class="fu-iEjS">。 藏 class="
class="fu-C6am">。 备 class="fu-0Dmb">。 买 class="fu-um4">。 主 class="fu-mJ8T">。 痰 class="fu-VD1b">。 捷 class="fu-spgw">。 耍 class="
class="fu-RNMu">。 菠 class="fu-BurF">。 干 class="fu-1hYH">。 腰 class="fu-81LL">。 眼 class="fu-ZvYy">。 市 class="fu-Diq3">。 拘 class="
class="fu-m7f4">。 床 class="fu-X4mM">。 孙 class="fu-Zpel">。 弄 class="fu-Jw0B">。 束 class="fu-1BqS">。 抄 class="fu-rza5">。 转 class="
class="fu-04QP">。 息 class="fu-aLc">。 采 class="fu-FFhx">。 宙 class="fu-1gGL">。 钳 class="fu-DM6Q">。 出 class="fu-h9a2">。 滔 class="
class="fu-1FTu">。 因 class="fu-0uE3">。 扶 class="fu-RdHu">。 木 class="fu-jJFe">。 眼 class="fu-g0X5">。 改 class="fu-FQcs">。 茫 class="
class="fu-13CX">。 触 class="fu-1LJ9">。 峙 class="fu-XtH">。 捷 class="fu-QHRu">。 萝 class="fu-7SFC">。 鄂 class="fu-vyX8">。 由 class="
class="fu-RPZ7">。 帐 class="fu-XQa8">。 汇 class="fu-oaFe">。 暨 class="fu-Fae2">。 汁 class="fu-ekvh">。 承 class="fu-tFeb">。 枝 class="
class="fu-a0af">。 空 class="fu-1WYH">。 踪 class="fu-V6B5">。 蕊 class="fu-5Ltd">。 档 class="fu-0H10">。 绝 class="fu-00pa">。 现 class="

```

代码如下：

```

def chinese(text):
    q = q.split('.')
    w,j,s,data,col,row='','0,[],{},{},[],[]
    for i in q:
        w = w + i
        if j == 3:
            if 'fu-' in w :
                data[re.findall(r'px -(.*?)px', w)[0]+'/' +re.findall(r'bac1
                col.append(re.findall(r'background:-(.*?)px', w)[0])
                row.append(re.findall(r'px -(.*?)px', w)[0])

```

```

        w ,j,s= ' ',0,[]
        j = j + 1
    result={}
    for i in range(70,23770+300,300):
        t=text[int((i-70)/300)]
        l=len(t)
        for j in range(0,l):
            if j==0:
                key=str(i)+'/'
            else:
                key=str(i)+'/'+str(j*140)
            try:
                result['<b class="{0}"></b>'.format(data[key])]=t[j]
            except KeyError:
                continue
    result['<br/>']='。'
    result['\xa0']=';'
    return result

```

把标签和字对应后就试着爬一下评论吧。在大众点评里，长的评论会只显示部分，它的全部内容class='desc J-desc'的<p>标签，短的评论在class='desc'的<p>标签内。



心雨 lv6
★★★★☆

走进麦当劳， 只要一推门进去，就会闻到一种特别的味道，这味道是汉堡包和薯条混合起来的味道。而且这种味道只有麦当劳才有。

麦当劳的天花板，挂着一个牌子来.....

展开评论 ∨

```

Performance  Elements  Console  Sources  Network  Memory  Application  Security  Audits
▼<div class="content">
  ▶<p class="shop-info">...</p>
  ▼<div class="info J-info-short clearfix">
    ::before
    ▶<p class="desc">...</p>
    ▶<a href="javascript:void(0);" class="more fold J-more" data-click-name="展开评论5" data-click-title="文字">...
  </a>
    ::after
  </div>
▼<div class="info J-info-all clearfix Hide"> == $0
  ▼<p class="desc J-desc">
    <b class="fu-DtAS"></b>
    "进"
    <b class="fu-5vuw"></b>
    <b class="fu-lsjP"></b>
    <b class="fu-8iBy"></b>
    ", 只要一推"
    <b class="fu-RicI"></b>
    "进去，就会闻到一种特别"
    <b class="fu-taIC"></b>

```

和前面一样，选择了美食、按摩、健身、旅游、游泳五个网页进行测试，每组只显示前三条评论，得到的结果如下：

<http://www.dianping.com/shop/77335766>

['坐标：罗宾森广场（以前的巴贝拉？暴露年纪？）。早上七点，想吃早饭，选麦当劳没错了[馋]上下
['一大早来一份差评其实我也不想的。终端自助点餐，选择的是堂吃，结果在店里根本不忙的情况下还
['在支付宝的麦当劳生活号领了一个大脆鸡扒麦满分的早餐免费券，今天来换汉堡，虽然一个不大，但

<http://www.dianping.com/shop/9799086>

['最近身体各处都很紧绷；硬帮帮的；肯定大众点评上面推荐的这家；感觉还不错就预定了；中午订好
['去过很多Spa馆，这一家真的可以算得上上海非常专业的了～；靠谱～全程都是柬埔寨小姐姐英文跟
['环境一般，但是手法不错。洗个澡做个精油很舒服，全是柬埔寨人，男女技师都有。门面倒是真的不

<http://www.dianping.com/shop/92871773>

['不喜欢大班课，所以买的一对二课程，团购560，然后游泳门票30-40/次，教练门票一共240（120
['因为零基础，所以选择了一对一的课程，。看评论说常教练很好，所以请客服安排了她，很漂亮的一
['客服安排和一个可爱的妹子拼了一对二的6节课程，一开始超怕的，以前从来没有游起来过，不过我1

<http://www.dianping.com/shop/23635007>

['[爱心]行走的英伦风格，小时代的拍摄点，整个小镇弥漫着文艺气息，兜兜转转可以逛一天，因为茸
['小雪已过天仍暖，银杏满地金灿灿。。小镇畅游把景看，回味无穷是此山。。上海佘山世茂洲际酒店
['这是我记忆中第二次来泰晤士小镇了！记得好几年前来的时候，这里还有很多店没有开出来，吃的也

<http://www.dianping.com/shop/78965850>

['喜欢这的教练（我的美女教练），喜欢这的关键（安静），也喜欢这的氛围（和谐）。练了快两个月
['里面小小的；不是特地来还真没想到这是个这么小的，就跟家里房间那样的大小。放了一些器材。隔
['说实话；我本身是不爱运动的人；真的是下了很大的决心才决定来这里尝试一下的；等上了几节课后

第一条评论就是前面举例的那条，可以对比一下噢，内容是完全一样的。

大众点评的解密到这里就全部讲完了，小伙伴可以自己动手做一次。所有代码已全部上传到Github（<https://github.com/zuobangbang/text-decoding/tree/master/dianping>）。之前上传的比较乱，以后上传的代码会进行精简和说明，帮助大家更好理解，欢迎在Github上关注我。

作者：左伊雅，目前在南京某 211 大学读研二，喜欢数据挖掘和爬虫。

声明：本文为作者投稿，文章仅用于学习交流，不得用于其他途径。

热 文 推 荐

👉 突发！5G 标准推迟三个月

👉 谁还能救小黄车？

👉 Python 分析 35 年的考研英语真题词汇，解读孤独的考研大军！