

OpenStreetMap 项目报告

一. 地图区域:

温州市, 中华人民共和国

<http://www.openstreetmap.org/relation/3289341>

我的家乡在温州, 我对这里相对比较熟悉, 所以我更想通过这个项目深入了解自己的家乡。

二. 地图遇到的问题:

通过对采样文件进行解析, 发现了部分问题。

1. 信息不完整: 比如 X975、X971 等路, 后面没有跟中文说明。
2. 信息过时: S41 省道是过去的路名, 现在已经更改为 S223。
3. 信息格式不一致: 比如 node 标签下 way 子标签中 name 属性值里对 104 国道的表示方式有好几种 (“National Road 104”, “104 National Road”, “104 国道线”, “G104 国道”); node 及 way 标签的 tag 子标签中, 少部分属性值含有空格、以及单纯用拼音来表示。
4. 信息错误: 极少部分 tag 子标签中出现 “uu” 或者字母加数字等无法理解的表述。
5. Way 标签下 tag 子标签中 Name.zh 的属性值少部分值为拼音, 而绝大多数的值是汉字。
6. Tag 子标签中的 source 字段, 出现单词相同但是大小写不一的情况: 比如 “Bing” 和 “bing”。

三. 清理方案:

1. 统一国道、高速公路表示格式: 如 “G104 国道”、“S325 省道” 等。
2. 去除属性值中的空格。
3. 对 tag 子标签中的 name 及 name.zh 属性值, 只保留含有中文的表述。

四. 数据概览: 这部分包括数据集的基本统计信息

1. 文件大小:

```
wenzhou.osm ----- 80695KB
OSM.db ----- 41796KB
Nodes.csv -----34082KB
nodes_tags.csv ----- 294KB
ways.csv -----1586KB
ways_nodes.csv -----11555KB
ways_tags.csv -----2349KB
```

2. 节点数:

```
sqlite> SELECT COUNT(*) FROM nodes;
422239
```

3. 道路数:

```
sqlite> SELECT COUNT(*) FROM ways;  
27483
```

4. 唯一用户数:

```
sqlite> SELECT COUNT(DISTINCT(e.uid)) from (SELECT uid FROM nodes UNION  
ALL SELECT uid FROM ways) as e;  
199
```

5. 贡献度最高的十个用户:

```
sqlite> SELECT e.user, COUNT(*) as num FROM (SELECT user FROM nodes UNION  
ALL SELECT user FROM ways) as e GROUP BY e.user ORDER BY num DESC LIMIT 10;  
katpatuka|92100  
aighes|61334  
jamesks|45673  
ff5722|31419  
ABRob|30008  
u_kubota|22417  
Qiboua|20387  
AUV|17139  
来自青山下|14616  
ediyes|11372
```

五. 其他想法:

从用户数来看, 总共用户才 199 个, 贡献度排名前 10 的用户总计贡献占比为 77%。整体来说, 在这个地区没有多少人知道或使用 openstreetmap。

六. 附加数据探索:

1. 前十大设施:

```
SELECT value, COUNT(*) as num FROM nodes_tags WHERE key='amenity' GROUP BY  
value ORDER BY num DESC LIMIT 10;  
bicycle_rental|70  
place_of_worship|41  
fuel|24  
restaurant|23  
bus_station|13  
hospital|13  
fast_food|11  
cafe|10  
school|10  
bank|9
```

2. 用户贡献地图数据时使用最多的前 5 种采集方式:

```
sqlite> select e.value,count(*) as num from (select key,value from nodes_tags
UNION ALL SELECT key,value FROM ways_tags) as e where key = 'source' group by
e.value order by num desc limit 5;
```

BING|8430

GPS|1493

PGS|395

MAPBOX|317

GNS|144

3. 宗教信仰统计：（最多的是佛教和道教）

```
SELECT nodes_tags.value, COUNT(*) as num FROM nodes_tags JOIN (SELECT
DISTINCT(id) FROM nodes_tags WHERE value='place_of_worship') as i ON
nodes_tags.id=i.id WHERE nodes_tags.key='religion' GROUP BY nodes_tags.value ORDER
BY num DESC;
```

buddhist|15

taoist|7

七. 结论：

经过对地图数据的审查，感觉整体数据不够详细，缺少很多地方的信息。比如市区的很多道路在里面都没有列出，究其原因还是用户数量不够，提交的数据量不够全面，建议如下：

建议 1：在用户界面呈现一个当前地区的用户活跃度排行榜，并显示自己的排名，以此刺激用户进行更多的提交

好处：提升用户活跃度，提高地图完整度

预期的问题：对 **openstreetmap** 组织来说，可能会增加开发成本

建议 2：导入百度或者高德地图的基础数据包

好处：提高地图数据完整性

预期的问题：不同地图软件的数据格式，命名方式可能不尽相同，合并数据并删除重复数据会产生巨大的工作量