选取的是上海地区的OSM文件，大小54M。链接如下：

<https://s3.amazonaws.com/metro-extracts.mapzen.com/shanghai_china.osm.bz2>

对原始的OSM数据进行分析后，发现如下一些问题：

* 相同属性的取值不统一

1. 当tag为“way“时，k=oneway对应的v值不统一，根据官方的解释，应该为yes或者no但在数据集中有的值为-1。需要统一为yes或者no。 官方解释链接如下：

<http://wiki.openstreetmap.org/wiki/Way>

1. highway和Hwy 不一致，需要转换Hwy到highway。

* 当tag为node或者way时，k=name对应的v是中文名字时，有的node节点没有k=name-en属性。
* 当tag为node时，user的值有些是中文。但绝大部分的user取值是英文字母

根据以上列出的问题，相应的解决办法如下：

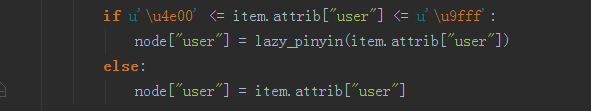
* 属性取值不统一的，在抽取数据时，把不规范或者不统一的值修改成规范或者统一值。
* 增加k=name-en属性，其v的取值采用百度开放的翻译API接口翻译成英文。

代码片段如下：



* 将user的取值为中文的，用汉语拼音方式改写。

代码片段如下：



数据抽取转换

* 完成以下内容

1. 只处理两种类型的顶级标记：“节点（node）”和“道路（way）”
2. CREATED 数组中的属性应该添加到键“created”下
3. 经纬度属性应该添加到“pos”数组中，以用于地理空间索引编制。确保“pos”数组中的值是浮点型，不是字符串。
4. 数据存为JSON格式并导入Mongo 数据库