两部分

匹配

合并

怎么匹配

1范围区间比较

2直接小数点多少位比较

3球大圆距离比较，多少米

怎么合并

合并地址文字部分

选取最长字符串

合并地址经纬度部分

1每次匹配成功都取旧的和新的经纬度，取算术平均值作为该地址的新的标准经纬度

2所有数据都保留，每个地址完成匹配给予它一个编号匹配程度编号与匹配到的地址编号相同，匹配失败给予新编号。循环结束后对编号进行升序排列，得到地址匹配表2号表。选取最优文字部分的进行保留，生成出真实地址表3号表

导入excel1号表原始表

第一个地址直接保留到新的表格

第二个地址开始，与前面所有保留地址进行对比，首先将该地址经度变成一个区间如

115.859247变成115.859247\*（1-0.0001）~115.859247\*（1+0.0001）精度有待确定，然后将已保存的经纬度与区间最大值最小值进行比较，

如果在区间内

同理比较纬度

纬度也在区间内

将其视为同一个地址，进行合并

合并规则，比较详细地址字符串长度，取较长的一个作为新地址收录保存，更新详细地址汉字部分

问题思考

两个完成匹配的经纬度怎么保留，经纬度怎么保存取舍？

取两个原始值均值？多次重复可能会造成误差

或者将匹配好的经纬度也保留下来？

这样的话，我的想法是完成匹配之后无论匹配成功与否都给他一个编号，相同地址相同标号，不同地址编号顺延，这样输出出来的还是4600多个数据，然后会出现123333456777这种情况，然后提取出每个编号相同的地址们中详细地址最长的一个作为最优的，保留到地址库里，以后比较的时候还需要对4600个数据做遍历处理

但也可能会产生问题

如果精度过高，有可能匹配不上

精度过低造成匹配成功率过高，造成bug，将同一地址可以匹配范围放大，造成不该匹配到的匹配到

还需要考虑如果匹配成功两个地址该怎么办？

取最先出现的那个？

如果经纬度有一项不匹配则视为两个不同地址，将该地址视为新地址收录，保留到新表格

完成循环

输出成新的excel

比较两个经纬度之间的距离，需要球大圆公式，然后距离在多少米之内认为同一地址

（或直接比较经纬度数据的前多少位字符，只要重合，则认为是同一地址，合并）

只考虑前多少位的优点

可能会解决上述问题思考的几个问题

但精确度需要考证