
中文地址地理编码方案 Documentation

Release 1

liuyou

November 22, 2013

CONTENTS

1	安装 CRF	1
2	安装 postgresql	3
3	安装 nlpbamboo	5
3.1	下载 bamboo 索引文件	5
3.2	安装 postgresql 中文分词模块	5
4	导入数据	7
4.1	安装 python 的 mysql postgresql 模块	7
5	测试步骤	9

安装 CRF

```
wget http://crfpp.googlecode.com/files/CRF%2B%2B-0.58.tar.gz
tar xzf CRF++-0.58.tar.gz
cd CRF++-0.58
./configure
make
make install
echo "/usr/local/lib" >> /etc/ld.so.conf
ldconfig
```


安装 POSTGRESQL

debian:

```
apt-get install postgresql-9.1 postgresql-contrib-9.1 postgresql-server-dev-9.1
```


安装 NLPBAMBOO

```
wget http://nlpbamboo.googlecode.com/files/nlpbamboo-1.1.2.tar.bz2
tar jxf nlpbamboo-1.1.2.tar.bz2
cd nlpbamboo-1.1.2
mkdir build
cd build
cmake .. -DCMAKE_BUILD_TYPE=release
make all
```

3.1 下载 bamboo 索引文件

```
wget http://nlpbamboo.googlecode.com/files/index.tar.bz2
tar jxf index.tar.bz2 -C /opt/bamboo
```

3.2 安装 postgresql 中文分词模块

```
cd /opt/bamboo/exts/postgres/pg_tokenize
make && make install
cd ../chinese_parse
make && make install

service postgresql restart
su - postgres
psql
\i /usr/share/postgresql/9.1/contrib/pg_tokenize.sql
\i /usr/share/postgresql/9.1/contrib/chinese_parser.sql
```


导入数据

4.1 安装 python 的 mysql postgresql 模块

debian:

```
apt-get install python-mysql python-psycopg2
```

importdata.py:

```
#!/usr/bin/env python
#coding=utf-8
import MySQLdb
, psycopg2
import sys

#psycopg2.paramstyle='qmark' #psycopg2.paramstyle 失效 ,全用%s
pconn = psycopg2.connect(host='172.16.147.133', user='postgres', password='1', database='address')
pc = pconn.cursor()

def insert(row, table):
    lens = len(row)
    str = "insert into %s values (%s);" % (table, (',%s' * lens)[1:])
    try:
        pc.execute(str, row)
    except:
        pass

if __name__ == "__main__":
    table = sys.argv[1]
    mconn = MySQLdb.connect(charset = 'gbk', host="192.168.1.16", user='caiwu', passwd='cai_xxxx', db='exp_address')
    mc = mconn.cursor()
    mc.arraysize = 10000
    mc.execute("select * from %s" % table)
    while 1:
        rows = mc.fetchmany()
        if not len(row):
            break
        for row in rows:
            insert(row, table)
    pconn.commit()
```

使用方法：

python importdata.py 表名

数据整理：

```
CREATE TABLE ldb1
(
  ctry integer,
  addr character varying(600),
  comp integer
)
```

```
insert into lab1 values
select
  recv_ctry integer,
  recv_addr ,
  entry_comp
from ldb
group by recv_ctry integer,
  recv_addr ,
  entry_comp
```

删除空数据：

```
delete from ldb1
where recv_ctry =0
  or recv_addr is null
  or entry_comp = 0
```

测试步骤

```
CREATE TABLE wd
(
    bm integer NOT NULL,
    sheng integer,
    shi integer,
    szd integer,
    psfw text,
    bpsfw text,
    mc character varying(100),
    CONSTRAINT wd_pkey PRIMARY KEY (bm)
);

CREATE TABLE psfw
(
    id integer NOT NULL DEFAULT nextval('areas_id_seq'::regclass),
    bm integer,
    dz character varying(80),
    szd integer,
    pslb integer, -- 1 全送排除 2 全不送排除 3 送 4 不送
    CONSTRAINT areas_pkey PRIMARY KEY (id)
);

CREATE INDEX psfw_dz_idx
    ON psfw
    USING btree
    (dz COLLATE pg_catalog."default");

CREATE TABLE psfw_lu
(
    id integer NOT NULL DEFAULT nextval('roadnums_id_seq'::regclass),
    psfw_id integer,
    xiao integer,
    da integer,
    dsh integer, -- 0 不区分单双号 1 单号 2 双号
    CONSTRAINT roadnums_pkey PRIMARY KEY (id)
);

CREATE INDEX psfw_lu_xiao_da_idx
    ON psfw_lu
    USING btree
    (xiao, da);
```

psfw 表存储 szd-区县 dz-地址 pslb 类别 dzidx 测试索引列

psfw_lu 表为 psfw 子表，当 psfw 表中 dz 为路且派送或者不派送范围不为全路段时，在

psfw_lu 表中记录门牌号码, xiao 起始号码 da 终止号码, dsh, 标记奇偶数号码

用 imdata1.py 把 mysql gsjj 表导入到 pgsqlldb 中

用 addip.py 处理 pgsqlldb 表, 数据进 psfw 表

然后, 增加一个 tsquery 列, 做全文索引查询用:

```
ALTER TABLE psfw ADD COLUMN dzidx tsquery;
CREATE INDEX psfw_dzidx_idx
    ON psfw
    USING gist
    (dzidx);
update psfw set dzidx=to_tsquery('chineseconfig',dz);
```

由于测试数据不规范, 只进行 bm=213300 szd=320481 的公司进行测试, 从 ldb 表中处理路、小区数据进入 psfw psfw_lu 表, 然后用 ldb 表中数据进行测试

用 procdatal.py 处理 ldb 表中 bm=213300 数据进入 psfw .

更新 213300 派件网点所在地为 999169, 测试用 (由于导入数据不准)

```
update psfw set szd =999169 where bm=213300
```

建立联合索引

```
CREATE INDEX psfw_szd_dzidx_idx
ON psfw
USING btree
(szd, dzidx);
```

查询语句, 耗时 10ms 级

```
select * from psfw where szd=999169 and to_tsvector('chineseconfig',' 骆驼街道贵驷里洞桥村洞桥路 60 号') @@ dzidx
```

查询门牌号 60

```
select * from psfw p inner join psfw_lu l
on p.id=l.psfw_id
where p.szd=999169 and to_tsvector('chineseconfig',' 骆驼街道贵驷里洞桥村洞桥路') @@ p.dzidx
and (60 between l.xiao and l.da)
```

取出数据结果后, 依据 psfw 表 psfb 来判断是否派送。门牌号取法采用正则, 见脚本 test.py 方法 procsb

测试方法 ldb 表中 bm=213300 任一记录 recv_addr 列去 psfw 中查询。

测试门牌号请插入以下数据至 psfw_lu, 测试效率可加入更多数据, 建立索引。:

```
psfw_id;xiao;da;dsh
820246;1;100;0
820246;101;200;1
820246;150;300;2
```

部分插入的测试地址:

```
213300;"台州路";999169;3
213300;"光明路地税局";999169;3
213300;"红星路";999169;3
213300;"粮食局";999169;3
213300;"迎江路";999169;3
213300;"东渡路";999169;3
213300;"厦门高崎五组 5043 号";999169;3
213300;"厦门高崎五组";999169;3
```

213300;"仙岳路";999169;3
213300;"历历七里河小区";999169;3
213300;"董村";999169;3
213300;"香港中路 132 号院";999169;3
213300;"南阁街";999169;3
213300;"河南师范大学";999169;3
213300;"神路";999169;3
213300;"承天路";999169;3
213300;"荆东路";999169;3
213300;"柏林馨苑";999169;3
213300;"解放东路";999169;3
213300;"禅禅湖景路";999169;3
213300;"禅禅季华五路";999169;3
213300;"潮汕路";999169;3
213300;"福田保税区南光紫荆苑";999169;3
213300;"人民医院西药房";999169;3
213300;"文昌路";999169;3
213300;"雀路";999169;3
213300;"柳邕路";999169;3
213300;"潭中西路";999169;3
213300;"鱼";999169;3
213300;"一环路南二段";999169;3
213300;"3";999169;3
213300;"北京南锣鼓巷";999169;3
213300;"十里堡镇";999169;3
213300;"南大街";999169;3
213300;"东城地纬路";999169;3
213300;"中山东路";999169;3
213300;"附城村";999169;3
213300;"附城村西小区";999169;3
213300;"平阳路";999169;3
213300;"东风新村";999169;3
213300;"上海巨鹿路";999169;3
213300;"上海宝胜路";999169;3
213300;"麒麟街";999169;3
213300;"建宁路";999169;3
213300;"周庄";999169;3
213300;"中联皇冠";999169;3
213300;"滨湖经济开发区华谊路";999169;3
213300;"隐秀苑";999169;3
213300;"维扬路";999169;3
213300;"花园岗街";999169;3
213300;"文三路";999169;3
213300;"柳汀街";999169;3
213300;"柯柯忠烈庙前";999169;3
213300;"邮电路";999169;3
213300;"塘下镇陈宅村友谊西路";999169;3
213300;"瑞安嘉宝锦园 H 栋";999169;3
213300;"下应街";999169;3
213300;"兴宁路";999169;3
213300;"环城北路东段 A 座 2 楼 (梦神商务中心";999169;3
213300;"环城北路东段";999169;3
213300;"普陀山路宁波天宇光电科技有限公司";999169;3
213300;"普陀山路";999169;3
213300;"骆驼街道贵驷里洞桥村";999169;3
213300;"骆驼街道贵驷里洞桥村洞桥路";999169;3
213300;"茗园 1 区包兴路";999169;3
213300;"茗园";999169;3

213300;"春江花城 78 号 104 室";999169;3
213300;"春江花城";999169;3
213300;"洞桥镇潘家套村二桥曦世工艺品有限公司";999169;3
213300;"梅墟街道社区卫生服务中心";999169;3
213300;"高新区梅墟梅福园 168101 室";999169;3
213300;"高新区梅墟梅福园";999169;3
213300;"科技园区竹泉路宁波乐星感应电子有限公司";999169;3
213300;"科技园区竹泉路";999169;3
213300;"前童镇老市场";999169;3
213300;"桃源中路财政局会计核算中心";999169;3
213300;"桃源中路";999169;3
213300;"太平洋商业街";999169;3
213300;"北三环长池支路开达电器有限公司";999169;3
213300;"北三环长池支路";999169;3
213300;"孙塘北路";999169;3
213300;"北定工业园";999169;3
213300;"龙港镇塑编 = 编工业区";999169;3
213300;"新区街";999169;3
213300;"牛山北路";999169;3
213300;"石桥头镇石黄路";999169;3
213300;"熟镇街";999169;3
213300;"黎明工业区金堂鞋业有限公司";999169;3
213300;"黎明工业区";999169;3
213300;"浙南农贸市场";999169;3
213300;"浙南农贸市场 2 区";999169;3
213300;"县府大院";999169;3
213300;"罗阳镇西大街";999169;3
213300;"虹桥镇幸福西路";999169;3
213300;"江诚大厦";999169;3
213300;"江诚大厦 A";999169;3
213300;"石帆第二工业区信本服饰包装有限公司";999169;3
213300;"乐成镇南马巷 1 号改水办";999169;3
213300;"乐成镇南马巷";999169;3
213300;"嘉善经济开发区";999169;3
213300;"中山西路";999169;3
213300;"嘉禾北京城 23 幢 306";999169;3
213300;"嘉禾北京城";999169;3
213300;"金穗月亮湾 52 幢 302 室";999169;3
213300;"金穗月亮湾";999169;3
213300;"中环西路 (文昌路口) 颐高广场";999169;3
213300;"中环西路";999169;3
213300;"新城街道档案局";999169;3
213300;"惠民镇惠民派出所";999169;3
213300;"姚庄";999169;3
213300;"海盐河南西路 (海盐城投集团);999169;3
213300;"海盐河南西路";999169;3
213300;"农业综合开发区启二路";999169;3
213300;"海洲街道世纪花园";999169;3
213300;"海洲街道紫薇花园";999169;3
213300;"当湖镇当湖西路";999169;3
213300;"新华爱心高级中学";999169;3
213300;"永丰路永丰西村";999169;3
213300;"职业中专高一 15 班 (fantasticbaby 创业社团) ";999169;3
213300;"职业中专高一";999169;3
213300;"文华路桂花城银桂苑";999169;3
213300;"百官工业园";999169;3
213300;"河西社区 105 幢 1 号";999169;3
213300;"河西社区 105 幢";999169;3

213300;"劳动路";999169;3
213300;"闻波兜 31 幢 303 室";999169;3
213300;"闻波兜";999169;3
213300;"龙溪北路";999169;3
213300;"莫干山镇集镇南路";999169;3
213300;"皇家湾名邸 4 号楼 1 单元 402 室";999169;3
213300;"皇家湾名邸";999169;3
213300;"雉城镇新塘村";999169;3
213300;"安电信局";999169;3
213300;"袍江越东南路";999169;3
213300;"花园";999169;3
213300;"花园岬南区";999169;3
213300;"灵芝镇小观村";999169;3
213300;"绍兴文理学院";999169;3
213300;"新建北路住宅 4 幢 601 室 (近县前街";999169;3
213300;"新建北路住宅";999169;3
213300;"绍柯桥万国中心";999169;3
213300;"绍柯桥万国中心 A 座";999169;3
213300;"浙江友谊路";999169;3
213300;"柳湖花园";999169;3
213300;"滨江小区";999169;3
213300;"滨江小区 B 幢";999169;3
213300;"岩头镇岩四村九区 (岩头小学";999169;3
213300;"岩头镇岩四村九区";999169;3
213300;"兰江街道上叶村";999169;3
213300;"工人北路";999169;3
213300;"国际商贸城 f 区四楼 19302 店面";999169;3
213300;"国际商贸城 f 区四楼";999169;3
213300;"苏溪镇人民路";999169;3
213300;"江东金村";999169;3
213300;"江东金村 F 区 6 栋";999169;3
213300;"上洪村东区群英路 (麦杰进出口有限公司";999169;3
213300;"上洪村东区群英路";999169;3
213300;"白云街";999169;3
213300;"白云街道鑫园路海德国际社区檀香 5 路 7 街";999169;3
213300;"横店镇江南路 (浙江普洛得邦制药有限公司";999169;3
213300;"横店镇江南路";999169;3
213300;"城南路";999169;3
213300;"巨化孔家一条街";999169;3
213300;"南市街";999169;3
213300;"国际商贸 6565 店";999169;3
213300;"国际商贸";999169;3
213300;"普陀公安局";999169;3
213300;"沈家门东海西路 17 弄";999169;3
213300;"沈家门街";999169;3
213300;"沈家门街道西大街";999169;3
213300;"高亭沿港中路";999169;3
213300;"黄岩东城开发区";999169;3
213300;"开发大道台州华胜汽车服务有限公司";999169;3
213300;"开发大道";999169;3
213300;"双江小区";999169;3
213300;"大麦屿兴港西路港行大楼";999169;3
213300;"人民西路";999169;3
213300;"城西街";999169;3
213300;"太平街";999169;3
213300;"太平街道方城路";999169;3
213300;"白云花苑";999169;3
213300;"城北 11 幢 1 单元 201 室";999169;3

213300;"城北";999169;3
213300;"人民医院泌尿科";999169;3
213300;"洛河经济开发区长壁机械有限责任公司";999169;3
213300;"蚌埠百货大楼";999169;3
213300;"华光大道";999169;3
213300;"曹市镇袁店二矿保卫科";999169;3
213300;"谯谯富荣花园";999169;3
213300;"谯谯富荣花园 A";999169;3
213300;"谯谯光明东路中国人民财产保险公司分公司";999169;3
213300;"谯谯光明东路";999169;3
213300;"谯谯亳州市国税局";999169;3
213300;"门台工业园";999169;3
213300;"皖紫薇花园";999169;3
213300;"清流路";999169;3
213300;"天长东路置业花苑";999169;3
213300;"交通局";999169;3
213300;"解放四大街";999169;3
213300;"中原路";999169;3
213300;"剑光城 7 号楼 202 室";999169;3
213300;"剑光城";999169;3
213300;"镜湖新村";999169;3
213300;"人民中路";999169;3
213300;"阜阳国贸商城四楼步步高学习机专柜";999169;3
213300;"国贸商城四楼家电部步步高点读机柜组";999169;3
213300;"淝河镇人民政府";999169;3
213300;"望江东路合肥水泥研究设计院检测中心";999169;3
213300;"望江东路";999169;3
213300;"金寨南路姚公庙东风新村";999169;3
213300;"马鞍山路中段中国联通合肥分公司";999169;3
213300;"石塘路肥东锦弘中学";999169;3
213300;"石塘路";999169;3
213300;"石塘路锦弘中学";999169;3
213300;"高新区高新区高新区天波路天怡国际商务中心";999169;3
213300;"巢湖中路中国银行";999169;3
213300;"巢湖中路";999169;3
213300;"长江中路";999169;3
213300;"阜南路测绘大厦 3 楼申银万国证券公司";999169;3
213300;"阜南路";999169;3
213300;"藕塘路与茨河路交叉口金荷苑";999169;3
213300;"藕塘路与茨河路交叉口金荷苑小区";999169;3
213300;"濉溪路财富广场";999169;3
213300;"濉溪路财富广场 B 座西楼";999169;3
213300;"其它区其它区滨湖新区康园小区";999169;3
213300;"其它区其它区经济技术开发区芙蓉路安徽医科大学";999169;3
213300;"其它区其它区新站区工业园";999169;3
213300;"其它区其它区新站区工业园宝业集团安徽有限公司";999169;3
213300;"创业大道丰乐加工包装中心";999169;3
213300;"创业大道";999169;3
213300;"合作化南路（原法院）稻香村";999169;3
213300;"合作化南路";999169;3
213300;"梅山路银保大厦人民银行合肥中心";999169;3
213300;"梅山路";999169;3
213300;"祁门路安徽国贸大厦";999169;3
213300;"祁门路";999169;3
213300;"长江东路银洁烟酒店";999169;3
213300;"长江东路";999169;3
213300;"新站区站前路";999169;3
213300;"新站区站前路浙江商贸城 C 座";999169;3

213300;"胜利路中环国际大厦";999169;3
213300;"胜利路中环国际大厦 A 座";999169;3
213300;"一路";999169;3
213300;"一路公交总站向北";999169;3
213300;"花园路";999169;3
213300;"梅苑路国税局";999169;3
213300;"梅苑路";999169;3
213300;"九龙岗肿瘤医院";999169;3
213300;"柏园南村";999169;3
213300;"东方医院总院原矿三院急诊科";999169;3
213300;"公安局";999169;3
213300;"泉林村";999169;3
213300;"蔡家岗街道社区卫生服务中心";999169;3
213300;"第四人民医院山下门诊蔡家岗长途车站对面";999169;3
213300;"徽山路中华职业学校";999169;3
213300;"徽山路";999169;3
213300;"马马迎江路";999169;3
213300;"马马马杨沫小区";999169;3