



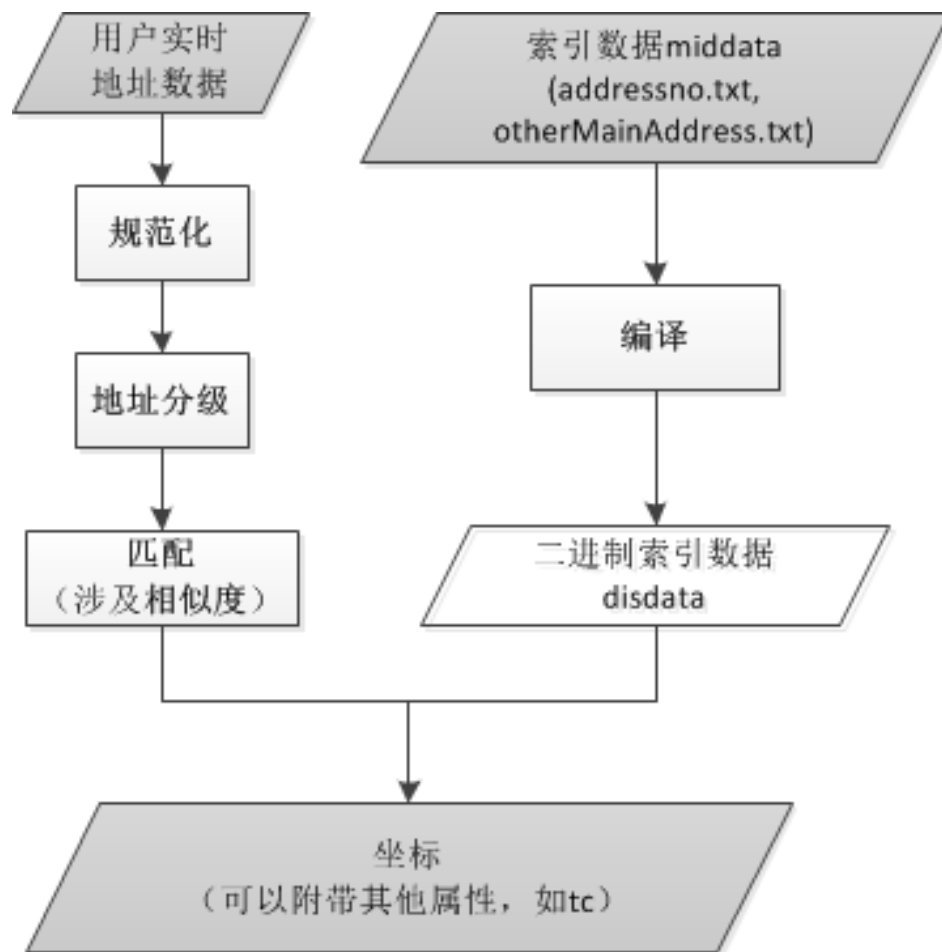
地理编码算法简介

地理信息研发中心
2018/4/9

地理编码基本概念

1. **地理编码**：也叫正向地理编码，是将地名或地址描述转换为地球表面上的相应位置。
2. **逆地理编码**：也叫反向地理编码，是将地球表面的地址坐标转换为标准地址。
3. **坐标**：这里指地理坐标经纬度。
4. **adcode**：行政区划代码，由中华人民共和国国家统计局发布，用六位阿拉伯数字分层次代表我国的省、市、区的名称。
5. **编译**：把人们熟悉的语言转换成计算机能识别的二进制语言。
6. **地址分级**：也叫切词、分词，是在通用中文分词原理的基础之上，对大量地址数据进行分析、挖掘，并依据分析模型，将用户地址数据切分为独立且有意义的小块，比如行政区划、道路、POI、门牌号等。
7. **匹配**：通俗地讲，就是先搜索，再在搜索结果中给出最优解。

地理编码算法流程



地理编码算法索引数据

1. middata：按省份分文件夹，和district.csv（全国省市区）。

每个省份文件夹下包含以下文件：

- town_village.txt：本省的街道、村、开发区
- road_line.txt：本省路网
- addressno.txt：本省门牌号和支门牌
- otherMainAddress.txt：本省POI和楼栋
- road_line_1.txt：路由分单项目独有

2. distdata：一个文件夹，包含多个文件，每个文件包含所有城市的数据。

注意：marked.csv文件是路由分单项目独有。

地理编码算法——规范化

- **繁简转换**
- **行政区划补全**
- **预处理**：全角转半角、HTML实体转换、字母转大写、去除开头结尾的空格和制表符
- **中文数字转换阿拉伯数字**
- **根据正则表达式替换字符**
- **字符替换**：地名标准化、去除人名等
- **后缀删除**：很多配送信息、订单信息等会写在结尾
- **根据配置表删除信息**：删除人名、邮编等
- **删除无意义字符**：删除空格、句号等标点符号
- **括号内信息处理**

地理编码算法——地址分级

地址分级模型——18级标准地址

编号	级别	示例
1	省、直辖市、特别行政区	广东省，北京市，香港特别行政区
2	地级市	深圳市
3	区、县、县级市	南山区、仁化县、乐昌市
4	开发区、工业园等	广州经济技术开发区
5	街道、镇、乡	粤海街道（办事处）、新塘镇、南阳乡
6	社区、村	粤桂社区（居委会）、风光村
7*	组、队	风光村 8组
8*	商圈	龙岗中心区
9	主路	友谊路
10	支路	龙苑新村 15巷 、友谊路 西6巷
11	门牌号	学府路 84号
12	支门牌号	学府路84- 1号
13	POI	桂园小区
14	楼栋号	桂园小区 8栋
15	单元号	桂园小区8栋 三单元
16	楼层号	桂园小区8栋三单元 3楼
17	房间号	桂园小区8栋三单元3楼 801室
18	地址描述信息	中油大厦 旁 、软件产业基地 的 1A栋

地理编码算法——地址分级

地址分级属性值

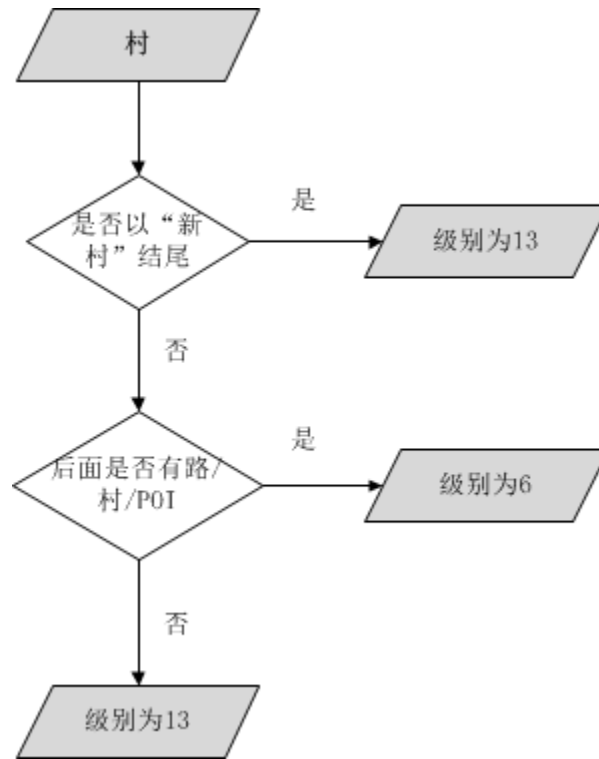
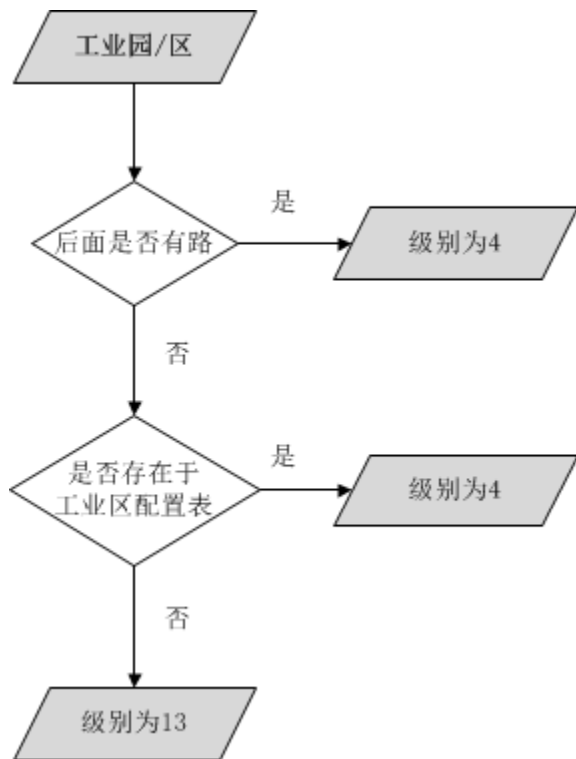
属性值	说明		
1	表示从词典中查询到的		
2	表示通过规则和逻辑推理出的		
3~7	用于 POI 的判定	3	某一类场所的统称，单独做POI没有意义；或者不会长期存在的POI
		4	主要指大桥、商圈等一类POI
		5	可共用坐标的POI，在地理位置上认为它们非常接近
		6	属于前面POI的一部分，且依赖于前面的POI而存在
		7	疑似POI

地址：广东省深圳市南山区粤海街道登良路8-4号蔚蓝海岸3期29栋2902

分词：广东省^11,深圳市^12,南山区^13,粤海街道^15,登良路^19,8^211,-4号^212,蔚蓝海岸^113,3期^613,29栋^214,2902^217
或者：广东省^11|深圳市^12|南山区^13|粤海街道^15|登良路^19|8^211|-4号^212|蔚蓝海岸^113|3期^613|29栋^214|2902^217

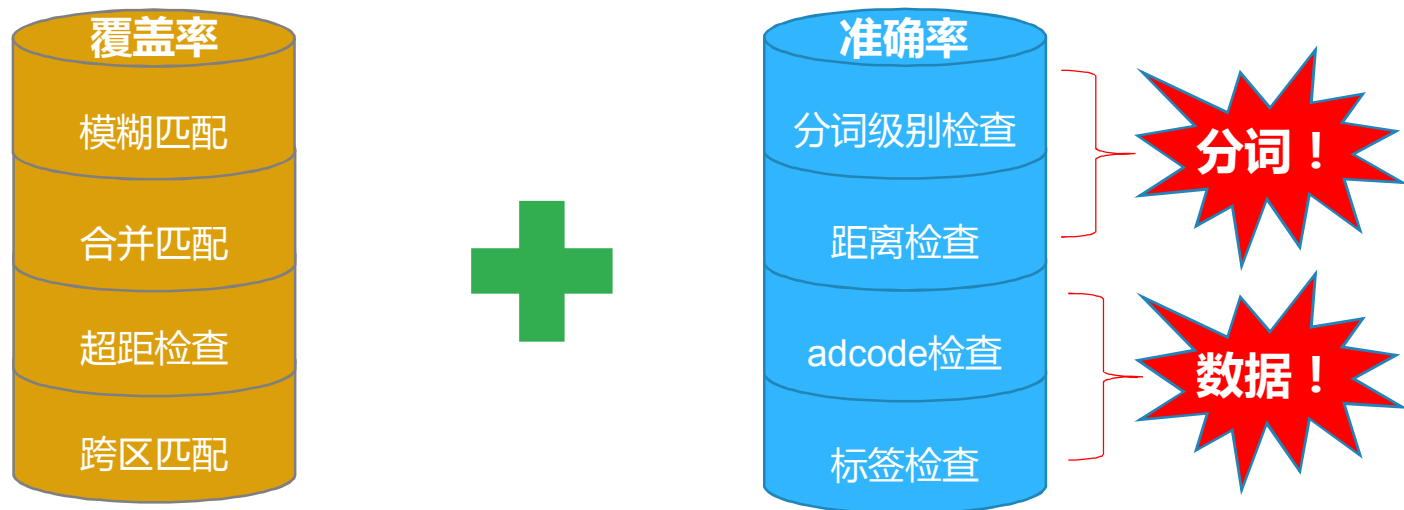
地理编码算法——地址分级

几类特殊级别的逻辑推断



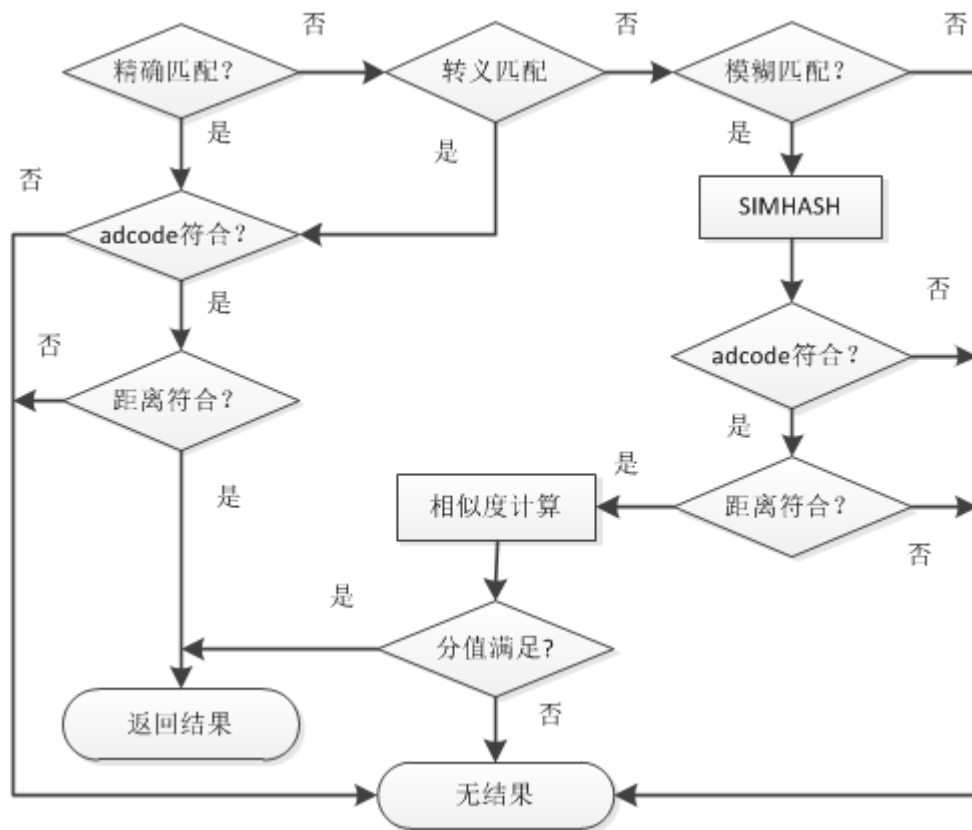
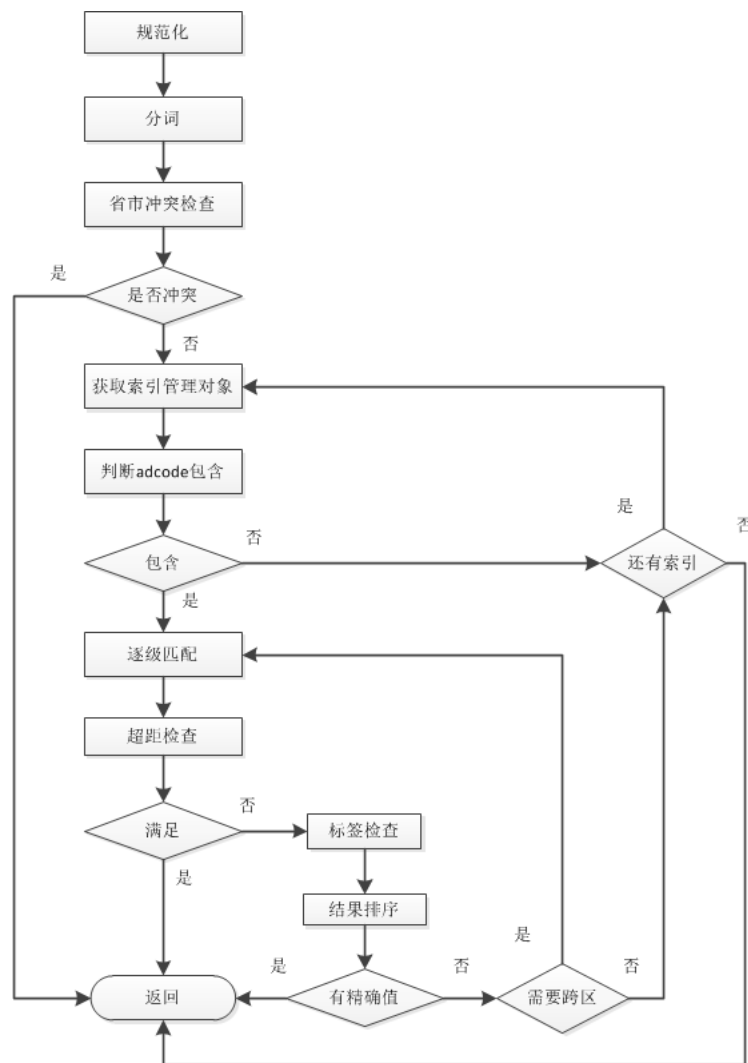
地理编码算法——匹配

1. **要求**：建议输入城市信息（鉴于全国同名地址太多）
2. **原则**：从前往后匹配，返回分词级别最低的部分。



注1：只对最终匹配的道路/门牌/支门牌/POI/楼栋 进行标签检查。

地理编码算法——匹配



注1：转义匹配仅省市区、街道、村执行；

注2：省市区不执行模糊匹配。

地理编码算法——匹配

几类特殊级别的匹配

1. **道路**：兼容各种用户书写习惯，添加/删除“东/南/西/北/中/公”以增大搜索范围，并且允许模糊匹配（道路相似度）。

match：深汕路

search：深汕东路、深汕南路、深汕西路、深汕北路、深汕中路、深汕公路、深汕路

- **注1**：当“**路”和“**东/南/西/北/中/公路”确认不是同一条路时，需添加进**道路冲突配置表**conflictRoad.cfg
- **注2**：**道路冲突配置表**不要和**别名停止词配置表**弄混，编译程序自动创建道路别名（例如给学府路建别名学府街），可关闭别名建立或添加进别名停止词配置表stopalias.cfg

2. **门牌号**：门牌号在索引中以挂接在道路下面的形式存在，所以匹配时只搜索已匹配的道路下的门牌号，不做距离检查。

3. **合并匹配**：支路（切词词性为10），必须与前面的主体合并匹配；POI附属词（切词词性为613），必须与前面的主体合并匹配；楼栋号（切词词性为14），当POI无该楼栋时，和前面POI主体合并在地标中搜索。

13+18/13+13/13+13+18/2+13/4+13/6+13/2+13+18/4+13+18/6+13+18/13+18+14(18为单个字)

地理编码算法——匹配

几类特殊级别的匹配

4. 连锁店类POI：标签检查时判断是否为连锁店及连锁店类型，对应不同的标签检查严格度。

chain	含义			
-1	连锁例外词，存在于exception_chain.cfg			
0	非连锁，不存在于任何配置			
1	连锁，存在于ChainStore.cfg	type=0	某个区县存在多个，其每个街道下存在多个（街道以下的标签必须匹配）	塞格科技园,440300,440304#:华强北麦当劳,0,0
		type=1	每个区县最多存在一个（POI前的信息能确定唯一区县）	卓越大厦,440300,440306
		type=2	某个区县存在多个，每个街道最多存在一个（POI前的信息能确定唯一街道）	赛格电子市场,440300,440306#
		type=3	0和2的混合型，有的街道有多个，有的街道最多一个	同富裕,440300, 440306#:沙井+福永+光明 440307#:坪地

注1：对于街道下有多个的连锁店，如果输入地址里包括门牌号，则在进行标签检查时将检查该POI的门牌号标签，如果不符合，则检查该POI距离门牌号的距离，如果超过距离认为该POI非高精，默认距离配置为300米，可以通过配置文件中的StreetNOLimit字段来配置该距离。

地理编码算法——匹配

几类特殊级别的匹配

POI进行模糊匹配的条件：

- POI字数大于2个字
- 没有精确值
- 有精确值，但是没有通过级别检查、adcode检查或距离检查 => 如果只是标签检查未通过，则不会回去模糊匹配
- POI在搜索时，出现同名的村（名称相同但级别为6），则不会进行模糊匹配
- 输入的adcode非0

地理编码算法——匹配

广东省^1|深圳市^2|宝安区^3|西乡街道^5|固戍^6|南昌村^13|东10巷^10|12^11|-4号
^12|蔚蓝海岸^13|3期^613|29栋^14|29楼^16|02室^17

注1：距离检查的各级距离阈值（经验阈值）如下：

区县：100000 m，开发区、街道：20000 m，社区/村：5000m，POI:1000m

道路交叉点：1500m，点状道路：1000 m，

线状道路：41000 高速公路:2000m；42000 国道:1500m；51000 省道:1000m；

52000 县道:500m；53000 乡公路:300m；54000 县乡村内部道路:200m；、43000 主

要大街、城市快速道:300m；44000 主要道路:800m；45000 次要道路:400m；

47000 普通道路:300m；49 非导航道路:200m

优先使用线状道路的距离限制；

注2：门牌号及子门牌与其后的数据无距离限制；

注3：对于613的距离限制，将参考前面的主体词性，使用其限制距离的一半；

注4：配置poiscope.cfg可以配置其搜索距离；

注5：索引中的精度字段可以扩大其搜索距离。

这些值将与配置文件中的AllowDistance字段作比较，取较大值（目前配置为1000m）

地理编码算法——配置文件

Def_Num=4	---索引源数量
Def_Name=yy	---索引1的名称
Def_Adcode=0	---索引1包含的adcode，多个以“ ”分隔
Def_Path=E:/问题/distdata_20180	---索引1目录
Def_Name1=sf	
Def_Adcode1=0	
Def_Path1=E:/问题/distdata_20175	
Def_Name2=sw	
Def_Adcode2=0	
Def_Path2=E:/问题/distdata_20177	
Def_Name3=gf	
Def_Adcode3=0	
Def_Path3=E:/问题/distdata_20179	
Patch_Path=E:/问题/disdata_patch	---索引在线更新时的补丁目录
Dict_Path=./data/dict	---匹配配置字典目录（存储连锁、连锁例外、道路冲突词等配置）
Segment_Path=./data/segment	---分词配置字典目录
Similarity_Path=./data/similarity	---相似度配置字典目录
Normalize_Path=./data/normalize	---规范化配置字典目录
Normalizer_IP=10.202.43.210	---规范化服务IP地址（目前规范化无windows版本）
Normalizer_Port=1323	---规范化服务端口号
AllowDistance=1000	---允许的最小搜索范围
NotCheckDistance=0	---是否不检查与上级的距离，0表示检查，1表示不检查
NotFuzzy=0	---是否不模糊匹配，0表示允许模糊匹配
StreetNOLimit=300	---连锁店距离门牌的距离限制
Precision=1,2,8	---认为是精确匹配的filter值
ScoreFilterPOI=	---认为POI匹配上的相似度分值，默认为0.9
ScoreFilterRoad=	---认为道路匹配上的相似度分值，默认为1
ScoreFilterMin=0.5	---认为有效的相似度分值，低于此值认为匹配无效，检查时用
GetAll=0	---是否获取所有可能匹配结果，数据生产用
NoNearby=0	---是否返回附近或插值结果，路由分单该值为1

地理编码算法——配置文件

索引目录

详细地址

address:

adcode: 索引:

☒ 检查标签 ☐ 跨区域匹配 ☐ 信任相似度 ☐ 门牌号优先 ☐ 不调规范化

查找

指定搜索的索引名称

adcode

单词	等级	属性	匹配	原始	指定搜索的索引名称	Value
广东省	1	1	0	1	广东省深圳市宝安区西乡街道固戍南昌村	
深圳市	2	1	1	2	source	sf
宝安区	3	1	1	3	address	固戍南昌村
西乡街道	5	1	1	5	id	823D3BD783E511E7AEA3000C29E2...
固戍	13	1	1	6	standardization	广东省深圳市宝安区西乡街道固戍...
南昌村	13	6	1	13	group	D33979BA83E811E7B1C0000C29E2...
东10巷	10	2	0	10	adcode	440306
12	11	2	0	11	coordinate	113.841120,22.596200
-4号	12	2	0	12	mainid	321814
蔚蓝海岸	13	1	0	13	subid	
3期	13	6	0	13	key	4 5
29栋	14	2	0	14	floor	0
29楼	16	2	0	16	type	0
02室	17	2	0	17	cell1	
					cell2	

分词结果

匹配结果

匹配过程

广东省深圳市宝安区西乡街道固戍南昌村东10巷12-4号蔚蓝海岸3期29栋29楼02室
NORMALIZE:广东省深圳市宝安区西乡街道固戍南昌村东10巷12-4号蔚蓝海岸3期29栋29楼02室.
SEGMENT TYPE:regular
SEGMENT:WORD=广东省,LEVEL=1,PROP=1
SEGMENT:WORD=深圳市,LEVEL=2,PROP=1
SEGMENT:WORD=宝安区,LEVEL=3,PROP=1
SEGMENT:WORD=西乡街道,LEVEL=5,PROP=1

地理编码算法——匹配示例

address=广东省深圳市南山区粤海街道登良路8-4号蔚蓝海岸3期29栋2902&adcode=440300↓

输入

NORMALIZE:广东省深圳市南山区粤海街道登良路8-4号蔚蓝海岸3期29栋2902. ↓

规范化

SEGMENT:WORD=广东省, LEVEL=1, PROP=1↓

SEGMENT:WORD=深圳市, LEVEL=2, PROP=1↓

SEGMENT:WORD=南山区, LEVEL=3, PROP=1↓

SEGMENT:WORD=粤海街道, LEVEL=5, PROP=1↓

SEGMENT:WORD=登良路, LEVEL=9, PROP=1↓

SEGMENT:WORD=8, LEVEL=11, PROP=2↓

SEGMENT:WORD=-4号, LEVEL=12, PROP=2↓

SEGMENT:WORD=蔚蓝海岸, LEVEL=13, PROP=1↓

SEGMENT:WORD=3期, LEVEL=13, PROP=6↓

SEGMENT:WORD=29栋, LEVEL=14, PROP=2↓

SEGMENT:WORD=2902, LEVEL=17, PROP=2↓

分词

CHECKADCODE:INPUT ADCODE = 440300, LEVEL = CITY↓

CHECKADCODE:PROVINCE 广东省, ADCODE = 440000. ↓

检查adcode

CHECKADCODE:CITY 深圳市, ADCODE = 440300. ↓

MATCH:WORD=广东省↓

SEARCH:ADMINNAME=广东省, ID=19, LEVEL=1, ADCODE=440000, ALIAS:0↓

MATCH:WORD=深圳市↓

SEARCH:ADMINNAME=深圳市, ID=230, LEVEL=2, ADCODE=440300, ALIAS:0↓

匹配

MATCH:WORD=南山区↓

SEARCH:ADMINNAME=南山区, ID=2118, LEVEL=3, ADCODE=440305, ALIAS:0↓

SEARCH:ADMINNAME=南山区, ID=1000, LEVEL=3, ADCODE=230404, ALIAS:0↓

MATCH:WORD=粤海街道↓

SEARCH:LANDMARK=粤海街道, ID=157909, LEVEL=5, ADCODE=440305, ALIAS:0, COORD=113.928101, 22.513814↓

FILTER:ID=157909, UPLEVEL=3, UPID=2118, LIMIT=100000 m, DIST=2148.84 m↓

地理编码算法——匹配示例

SEARCH:ADMINNAME=广东省, ID=19, LEVEL=1, ADCODE=440000, ALIAS:0↓
MATCH:WORD=深圳市↓
SEARCH:ADMINNAME=深圳市, ID=230, LEVEL=2, ADCODE=440300, ALIAS:0↓
MATCH:WORD=南山区↓
SEARCH:ADMINNAME=南山区, ID=2118, LEVEL=3, ADCODE=440305, ALIAS:0↓
SEARCH:ADMINNAME=南山区, ID=1000, LEVEL=3, ADCODE=230404, ALIAS:0↓
MATCH:WORD=粤海街道↓
SEARCH:LANDMARK=粤海街道, ID=157909, LEVEL=5, ADCODE=440305, ALIAS:0, COORD=113.928101, 22.513814↓
FILTER:ID=157909, UPLEVEL=3, UPID=2118, LIMIT=100000 m, DIST=2148.84 m↓
MATCH:WORD=登良路↓
SEARCH:ROADNAME=登良西路, ID=1206409, LEVEL=10, ADCODE=440305, ALIAS:0, COORD=113.921067, 22.513980↓
FILTER:ID=1206409, UPLEVEL=5, UPID=157909, LIMIT=20000 m, DIST=722.77 m↓
SEARCH:ROADNAME=登良路, ID=1259612, LEVEL=10, ADCODE=440300, ALIAS:0, COORD=113.923708, 22.513652↓
SEARCH:ROADNAME=登良路, ID=1247037, LEVEL=10, ADCODE=440300, ALIAS:0, COORD=113.937429, 22.512458↓
SEARCH:ROADNAME=登良路, ID=233030, LEVEL=9, ADCODE=440305, ALIAS:0, COORD=113.948153, 22.515908↓
SEARCH:ROADNAME=登良路, ID=231655, LEVEL=9, ADCODE=440305, ALIAS:0, COORD=113.930757, 22.509918↓
SEARCH:ROADNAME=登良路, ID=224002, LEVEL=9, ADCODE=440305, ALIAS:0, COORD=113.921471, 22.514043↓
FILTER:ID=224002, UPLEVEL=5, UPID=157909, LIMIT=20000 m, DIST=421.10 m↓
FILTER:ID=231655, UPLEVEL=5, UPID=157909, LIMIT=20000 m, DIST=414.62 m↓
FILTER:ID=233030, UPLEVEL=5, UPID=157909, LIMIT=20000 m, DIST=2029.28 m↓
MATCH:WORD=8↓
STREETNO:2栋, ID=83650, MAINID=231655, ADCODE=440305, LEVEL=16, COORD=113.934311, 22.509729↓
STREETNO:5, ID=47042, MAINID=231655, ADCODE=440305, LEVEL=14, COORD=113.936356, 22.510153↓
STREETNO:6栋, ID=149270, MAINID=231655, ADCODE=440305, LEVEL=16, COORD=113.930573, 22.510289↓
STREETNO:7座, ID=23759, MAINID=231655, ADCODE=440305, LEVEL=16, COORD=113.934977, 22.509632↓
STREETNO:8座, ID=76701, MAINID=231655, ADCODE=440305, LEVEL=16, COORD=113.933429, 22.510137↓
STREETNO:19号, ID=120424, MAINID=224002, ADCODE=440305, LEVEL=14, COORD=113.925010, 22.511170↓
STREETNO:62号, ID=126634, MAINID=1206409, ADCODE=440305, LEVEL=14, COORD=113.921067, 22.513980↓
MATCH:WORD=-4号↓

三级行政区划的匹配: adcode限制

街道级匹配: adcode限制、距离限制, 允许粤海、粤海街道、粤海镇相互匹配

道路级匹配: adcode限制、距离限制, 允许某路、某东/南/西/北/中路相互匹配 (除道路冲突配置外)

ALIAS:1, 表示建索引时生成的别名, 可以通过别名停止词限制该别名的建立

门牌号匹配: adcode限制, 只匹配数字、不匹配后缀 (号、栋、座、号楼等), 只搜索已匹配道路下挂的门牌号

支门牌匹配: 同门牌匹配, 只搜索已匹配门牌号下挂的支门牌号

地理编码算法——匹配示例

MATCH:WORD=-4号↓

MATCH:WORD=蔚蓝海岸↓

SEARCH:LANDMARK=蔚蓝海岸, ID=1114079, LEVEL=13, ADCODE=440303, ALIAS:0, COORD=113.937429, 22.512458↓

POI匹配: adcode限制、距离限制, 允许模糊匹配(经验阈值)

SEARCH:LANDMARK=蔚蓝海岸, ID=660554, LEVEL=13, ADCODE=440307, ALIAS:0, COORD=114.498144, 22.607432↓

SEARCH:LANDMARK=蔚蓝海岸, ID=599748, LEVEL=13, ADCODE=440305, ALIAS:0, COORD=113.935520, 22.508067↓

SEARCH:LANDMARK=蔚蓝海岸, ID=385449, LEVEL=13, ADCODE=440305, ALIAS:0, COORD=113.932942, 22.507521↓

FILTER:ID=385449, UPLEVEL=9, UPID=231655, LIMIT=1000 m, DIST=267.41 m↓

FILTER:ID=599748, UPLEVEL=9, UPID=231655, LIMIT=1000 m, DIST=211.09 m↓

MATCH:WORD=3期↓

MATCH:WORD=蔚蓝海岸3期↓

613匹配: 必须和13合并匹配, adcode限制、距离限制(上级不是合并的13)

SEARCH:LANDMARK=蔚蓝海岸3期, ID=648998, LEVEL=13, ADCODE=440305, ALIAS:0, COORD=113.937429, 22.512458↓

SEARCH:LANDMARK=蔚蓝海岸3期, ID=505195, LEVEL=13, ADCODE=440305, ALIAS:0, COORD=113.936587, 22.511729↓

FILTER:ID=505195, UPLEVEL=9, UPID=231655, LIMIT=1000 m, DIST=192.27 m↓

FILTER:ID=648998, UPLEVEL=9, UPID=231655, LIMIT=1000 m, DIST=270.42 m↓

MATCH:WORD=29栋↓

楼栋匹配: 和POI合并匹配, adcode限制、距离限制, 使用POI相似度算法计算POI相似度

MATCH:WORD=蔚蓝海岸3期29栋↓

BEGIN FUZZY SEARCH. ↓

只允许在阈值范围内进行模糊匹配

FILTER:ID=480884, UPLEVEL=9, UPID=231655, LIMIT=1000 m, DIST=267.41 m↓

FUZZY:LANDMARK=蔚兰海岸3期29号楼, ID=480884, LEVEL=13, ADCODE=440305, COORD=113.932942, 22.507521, SCORE=1.00, FLAG=0↓

INFO:ID=480884, INDEXFLOOR=220688. ↓

楼层匹配: 只匹配数字、不匹配后缀(楼/层), 可以从房间号推断楼层

MATCHFLOOR:LANDMARK=2902, FLOOR IS EMPTY. ↓

CHECKTAG:LANDMARK=蔚兰海岸3期29号楼, ID=480884, CHAIN=0, Type=0, ↓

判断是否是连锁店及连锁店类型, 标签检查严格度: 连锁店 (type:0>3>2>1) >非连锁店

TAG=3F2193E8B2FB11E7A9FC000C29E27523:903F47DEB2FF11E7AAE2000C29E27523:广东省深圳市南山区登良路蔚兰海岸3期29号楼:755DL,,755DL022:~登良路~9~1↓

TAGMATCH:LEVEL=9, NAME=登良路, TAG=登良路, SCORE=1.00, FLAG=0↓

标签检查: 只有该主体之前的已匹配上的部分才能作为标签参与检查

TAG MATCHED!↓

相似度计算: 使用道路相似度再次计算相似度, 避免POI相似度bug

COMPARE:蔚兰海岸3期29栋 AND 蔚兰海岸3期29号楼 assimilate the end word. ← 特殊后缀词的相似度处理: 号、栋、幢、座、号楼可以相互匹配

地理编码算法——匹配反例

切分错误：新区东二环路二号富士康科技集团营建采购处

级别错误：西乡共乐铁仔路55号四楼

adcode错误：广东深圳市坪山区沙博新村

坐标错误：高新技术产业园清华信息港研发楼A301

标签错误：南山区软件产业基地肯德基

相似度不够：南山区软件产业基地园区

距离超限：广东深圳市南头街道桃园路软件产业基地1栋

地理编码引擎

1. 通信方式：TCP/HTTP

2. 编码方式：GBK

3. 请求串：

http://10.202.43.221:9098/?**user_id**=ronghe&**query_type**=GEOCODE&**address**=南山区学府路83号软件产业基地1栋A座15楼&**ret_splitinfo**=2&**filter_uprecision**=2&**adcode**=110000&**opt**=sf&**output**=xml

字段	取值及说明
user_id	STREETNO:优先返回门牌号；其他：返回最精细级别
query_type	GEOCODE:地理编码；GEOGETALL:返回所有匹配结果；UPDINDEX:更新索引；QUERYSTATE:查询索引更新状态
address	待匹配地址
ret_splitinfo	0:不包含分词信息和标准地址;1:包含分词信息，不包含标准地址; 2:有标准地址；3:无分词信息，有标准地址
filter_uprecision	0：不进行过滤检查，不检查标签，当结果非精确时尝试跨区匹配；1：当结果非精确时尝试跨区匹配；2：做标签检查；3：检查标签，当结果非精确时跨区匹配
adcode	三级行政区划代码
output	数据输出格式，json:json格式,xml:xml格式，默认为xml格式
opt	指定源名称，默认为所有源

地理编码引擎

4. 返回结果说明：

```
<status>0</status> //0 OK, 5 省市冲突
<count>1</count> //匹配结果数量
<source>yy</source> //匹配结果来源(索引名称)
<list>
<poi>
<name>广东省深圳市南山区软件产业基地 1 栋 a 座</name> //根据 adcode 拼出的名称, 非小组名和标准名
<floor>15</floor> //楼层号
<id>2EDE5BAAEABE11E7BF550894EF0EEB3C</id> //小组 id
<province>广东省</province> //省
<city>深圳市</city> //市
<district>南山区</district> //区
<address>软件产业基地 1 栋 a 座</address> //匹配地址, 小组名
<adcode>440305</adcode> //行政区划代码
<level>GL_POI</level> //匹配级别
<mainid>4769101</mainid> //编译后的主地址 id
<subid></subid> //编译后的子地址 id
<key>3|4|5</key> //匹配的序号 (对应分词结果)
<x>113.939168</x> //经度
<y>22.52451</y> //纬度
<score>1</score> //相似度分值
<sflag>0</sflag> //相似度分值标记
<filter>1</filter> //检查过滤值
<poi_typecode>0</poi_typecode> //POI 类型
<dataSrc>20</dataSrc> //数据来源
<standardization>广东省深圳市南山区粤海街道学府路深圳市软件产业基地 1 栋 A 座</standardization> //标准地址
<group>D73DE3AEEAFC11E7B5520894EF0EEB3C</group> //标准地址 id
<cell1>,,</cell1> //zc,sss_tc,sch_tc
<cell2></cell2> //预留字段
</poi>
</list>
<splitResult>南山区^13,学府路^19,83 号^211,软件产业基地^213,1 栋^214,a 座^214,15 楼^216;16</splitResult> //分词结果
<splitType>0</splitType> //分词类型, 0 规则分词, 100 机器分词
<addrSplitInfo> //分级匹配结果, match: 0 不匹配, 1 匹配, prop: 属性值; level: 分词级别
<as_info match="1" prop="1" level="3">南山区</as_info>
<as_info match="1" prop="1" level="9">学府路</as_info>
<as_info match="1" prop="2" level="11">83 号</as_info>
<as_info match="1" prop="2" level="13">软件产业基地</as_info>
<as_info match="1" prop="6" level="13">1 栋</as_info>
<as_info match="1" prop="6" level="13">a 座</as_info>
<as_info match="0" prop="2" level="16">15 楼</as_info>
</addrSplitInfo>
```

地理编码引擎

5. 返回的两个重要字段：

filter	含义
0	级别不够没做检查
1	精确值
2	标签不匹配
4*	标记的项
8	模糊匹配
16	附近
32	街道不符合
64	连锁店
128	相似度极低
6*	2+4
10	2+8
12*	8+4
40	32+8
66	64+2
70	64+2+4
74	64+2+8
96	64+32
...	...

GL_LEVEL	含义
GL_COUNTRY	国家
GL_PROVINCE	省
GL_CITY	地市级
GL_COUNTY	区县
GL_DEV_ZONE	开发区
GL_TOWN	乡镇、街道
GL_VILLAGE	村、社区
GL_GROUP	组、队
GL_BZONE	商圈
GL_LINE	有线坐标道路
GL_ROAD	点坐标道路
GL_ROAD_BRANCH	支路
GL_ROADINTER	道路交叉口
GL_POI	兴趣点
GL_STREETNO	门牌
GL_STREETNO_SUB	支门牌
GL_BUILDINGNO	楼栋号
GL_INTER	插值
GL_NEARBY	附近

匹配上？

GL_LEVEL=GL_POI/GL_STREETNO/GL_STREETNO_SUB/GL_BUILDINGNO
&& filter=1/2/8/4/6/12

地理编码服务

编号	服务
1	获取经纬度坐标
2	地址规范化服务
3	地址分级服务
4	相似度服务
5	地址标准化
6	地址纠错服务
7	楼层识别服务
8	POI类型识别服务

类别	项目
内部	标准库（四源数据生产）
	路由分单项目
	地址筛单项目
	小哥行为规范项目
	输入提示项目（缺失库）
外部	舒华、CX、同城、仓配、理赔、冷链、车联网、医药、大数据.....



Thanks