



(12) 发明专利申请

(10) 申请公布号 CN 106933883 A

(43) 申请公布日 2017. 07. 07

(21) 申请号 201511028783. 5

(22) 申请日 2015. 12. 31

(71) 申请人 中移(苏州)软件技术有限公司
地址 215163 江苏省苏州市高新区科灵路
78号(苏高新软件园)10号楼
申请人 中国移动通信集团公司

(72) 发明人 陈敏 金晶 梁俊 刘道飞

(74) 专利代理机构 北京派特恩知识产权代理有
限公司 11270
代理人 张振伟 姚开丽

(51) Int. Cl.
G06F 17/30(2006. 01)

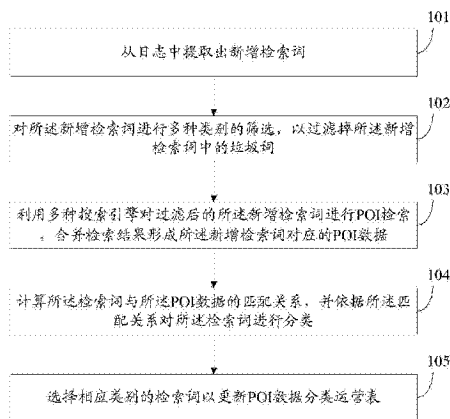
权利要求书3页 说明书14页 附图5页

(54) 发明名称

基于检索日志的兴趣点常用检索词分类方法、装置

(57) 摘要

本发明公开了一种基于检索日志的POI常用检索词分类方法、装置,所述方法包括:从日志中提取出新增检索词;对所述新增检索词进行多种类别的筛选,以过滤掉所述新增检索词中的垃圾词;利用多种搜索引擎对过滤后的所述新增检索词进行POI检索,合并检索结果形成所述新增检索词对应的POI数据;计算所述检索词与所述POI数据的匹配关系,并依据所述匹配关系对所述检索词进行分类;选择相应类别的检索词以更新POI数据分类运营表。



1. 一种基于检索日志的兴趣点POI常用检索词分类方法,其特征在于,所述方法包括:
从日志中提取出新增检索词;
对所述新增检索词进行多种类别的筛选,以过滤掉所述新增检索词中的垃圾词;
利用多种搜索引擎对过滤后的所述新增检索词进行POI检索,合并检索结果形成所述新增检索词对应的POI数据;

计算所述检索词与所述POI数据的匹配关系,并依据所述匹配关系对所述检索词进行分类;

选择相应类别的检索词以更新POI数据分类运营表。

2. 根据权利要求1所述的基于检索日志的兴趣点POI常用检索词分类方法,其特征在于,所述对所述新增检索词进行多种类别的筛选,以过滤掉所述新增检索词中的垃圾词,包括:

将所述新增检索词逐条与POI信息库分类表中的各个分类名称进行字符对比;若检索词与分类名称完全匹配,则将所述检索词从新增检索词中删除;和/或,

逐条验证所述新增检索词是否为预设的公交车线路/地铁线路或预设的公交车站点/地铁站点;是时,将所述检索词从所述新增检索词中删除;和/或,

逐条验证所述新增检索词是否指示路线查询;是时,将所述检索词从所述新增检索词中删除;和/或,

逐条验证所述新增检索词是否由区域和检索词组成;是时,将所述检索词从所述新增检索词中删除。

3. 根据权利要求1所述的基于检索日志的兴趣点POI常用检索词分类方法,其特征在于,所述利用多种搜索引擎对过滤后的所述新增检索词进行POI检索,合并检索结果形成所述新增检索词对应的POI数据,包括:

将过滤后的所述新增检索词逐条进入POI信息库或第一搜索引擎中搜索,获取结果列表;将结果列表为空的检索词汇总作为第二搜索引擎搜索的输入;将结果列表不为空的检索词汇总以待检索词合并;

将所述输入的检索词逐条进入第二搜索引擎中搜索,获取结果列表;对于结果列表为空的检索词,提示其他搜索城市选择的检索词汇总生成非归属地数据;将未提示其他搜索城市选择的检索词汇总生成高额赏金任务;对于结果列表不为空的检索词,将首条结果的POI名称作为替换检索词,并与原检索词建立关联关系;

将替换后检索词逐条进入所述POI信息库的第三搜索引擎中搜索,获取结果列表;将结果列表为空的检索词汇总生成查无数据;将结果列表不为空的检索词汇总以待检索词合并;

将以待合并的检索词进行合并,生成待匹配检索词。

4. 根据权利要求3所述的基于检索日志的兴趣点POI常用检索词分类方法,其特征在于,所述计算所述检索词与所述POI数据的匹配关系,并依据所述匹配关系对所述检索词进行分类,包括:

逐条判断所述待匹配检索词搜索结果中,是否存在任意一条poi名称包含完整的检索词;若不存在,保留全部搜索结果;

若存在,提取出poi名称包含完整的检索词类数据;若包含完整的检索词的poi名称只

有一条,更新POI常用检索词库,并将其匹配状态指示为一对一类成功匹配,同时更新一对一匹配文件;若包含完整的检索词的poi名称有多条,且多条数据类别属于同一个分类时,更新POI常用检索词库,并将其匹配状态指示为一对多类成功匹配,同时更新一对多匹配文件;若多条数据类别属于多个分类,则判断各个分类的占比,若存在分类占比大于等于第一阈值时,则将该分类数据与检索词匹配,更新POI常用检索词库,并将其匹配状态指示为一对多类成功匹配,同时更新一对多匹配文件;否则将数据放入无法识别类文件中;

计算剩余的待匹配检索词与其对应的搜索结果的匹配度,若检索词与所有结果的匹配度均低于第二阈值,则此检索词属于查无类,将所述检索词添加至查无数据文件中;若检索词与结果的匹配度高于第二阈值,将所述检索词与所述结果匹配,添加至人工待处理文件中。

5. 根据权利要求3所述的基于检索日志的兴趣点POI常用检索词分类方法,其特征在于,所述选择相应类别的检索词以更新POI数据分类运营表,包括:

对于一对多匹配文件中的检索词,逐条提取匹配结果在POI信息库中的数据类别,生成待分类运营文件;

根据待分类运营文件中的数据类别,在POI信息库分类运营表中确定对应的类别运营种类,并汇总形成自然增加类别的检索词数据文件、根据条件增加类别的检索词数据文件和独立词数据文件;

根据预制条件把条件增加类文件中的检索词分别添加到自然增加类文件和独立词数据文件中;

根据自然增加类别文件中的检索词,自然增加类别,更新POI信息库分类运营表;并根据新增分类情况,更新POI信息库中的数据类别;以及根据独立词数据文件中的检索词,更新独立词词库。

6. 一种基于检索日志的POI常用检索词分类装置,其特征在于,所述装置包括:

检索词提取模块,用于从日志中提取出新增检索词;

检索词过滤模块,用于对所述新增检索词进行多种类别的筛选,以过滤掉所述新增检索词中的垃圾词;

多引擎搜索模块,用于利用多种搜索引擎对过滤后的所述新增检索词进行POI检索,合并检索结果形成所述新增检索词对应的POI数据;

类型匹配模块,用于计算所述检索词与所述POI数据的匹配关系,并依据所述匹配关系对所述检索词进行分类;

分类运营模块,用于选择相应类别的检索词以更新POI数据分类运营表。

7. 根据权利要求6所述的基于检索日志的POI常用检索词分类装置,其特征在于,所述检索词过滤模块,还用于将所述新增检索词逐条与POI信息库分类表中的各个分类名称进行字符对比;若检索词与分类名称完全匹配,则将所述检索词从新增检索词中删除;和/或,逐条验证所述新增检索词是否为预设的公交车线路/地铁线路或预设的公交车站点/地铁站点;是时,将所述检索词从所述新增检索词中删除;和/或,逐条验证所述新增检索词是否指示路线查询;是时,将所述检索词从所述新增检索词中删除;和/或,逐条验证所述新增检索词是否由区域和检索词组成;是时,将所述检索词从所述新增检索词中删除。

8. 根据权利要求6所述的基于检索日志的POI常用检索词分类装置,其特征在于,所述

多引擎搜索模块,还用于将过滤后的所述新增检索词逐条进入POI信息库或第一搜索引擎中搜索,获取结果列表;将结果列表为空的检索词汇总作为第二引擎搜索的输入;将结果列表不为空的检索词汇总以待检索词合并;将所述输入的检索词逐条进入第二搜索引擎中搜索,获取结果列表;对于结果列表为空的检索词,提示其他搜索城市选择的检索词汇总生成非归属地数据;将未提示其他搜索城市选择的检索词汇总生成高额赏金任务;对于结果列表不为空的检索词,将首条结果的POI名称作为替换检索词,并与原检索词建立关联关系;将替换后检索词逐条进入所述POI信息库的第三搜索引擎中搜索,获取结果列表;将结果列表为空的检索词汇总生成查无数据;将结果列表不为空的检索词汇总以待检索词合并;将以待合并的检索词进行合并,生成待匹配检索词。

9. 根据权利要求6所述的基于检索日志的POI常用检索词分类装置,其特征在于,所述类型匹配模块,还用于逐条判断所述待匹配检索词搜索结果中,是否存在任意一条poi名称包含完整的检索词;若不存在,保留全部搜索结果;若存在,提取出poi名称包含完整的检索词类数据:若包含完整的检索词的poi名称只有一条,更新POI常用检索词库,并将其匹配状态指示为一对一类成功匹配,同时更新一对一匹配文件;若包含完整的检索词的poi名称有多条,且多条数据类别属于同一个分类时,更新POI常用检索词库,并将其匹配状态指示为一对多类成功匹配,同时更新一对多匹配文件;若多条数据类别属于多个分类,则判断各个分类的占比,若存在分类占比大于等于第一阈值时,则将该分类数据与检索词匹配,更新POI常用检索词库,并将其匹配状态指示为一对多类成功匹配,同时更新一对多匹配文件;否则将数据放入无法识别类文件中;计算剩余的待匹配检索词与其对应的搜索结果的匹配度,若检索词与所有结果的匹配度均低于第二阈值,则此检索词属于查无类,将所述检索词添加至查无数据文件中;若检索词与结果的匹配度高于第二阈值,将所述检索词与所述结果匹配,添加至人工待处理文件中。

10. 根据权利要求6所述的基于检索日志的POI常用检索词分类装置,其特征在于,所述分类运营模块,还用于对于一对多匹配文件中的检索词,逐条提取匹配结果在POI信息库中的数据类别,生成待分类运营文件;根据待分类运营文件中的数据类别,在POI信息库分类运营表中确定对应的类别运营种类,并汇总形成自然增加类别的检索词数据文件、根据条件增加类别的检索词数据文件和独立词数据文件;根据预制条件把条件增加类文件中的检索词分别添加到自然增加类文件和独立词数据文件中;根据自然增加类别文件中的检索词,自然增加类别,更新POI信息库分类运营表;并根据新增分类情况,更新POI信息库中的数据类别;以及根据独立词数据文件中的检索词,更新独立词词库。

基于检索日志的兴趣点常用检索词分类方法、装置

技术领域

[0001] 本发明涉及地理信息检索领域,尤其涉及一种基于检索日志的POI常用检索词分类方法、装置。

背景技术

[0002] 在利用地图查询地理位置的时候,用户首先向地图搜索引擎输入一个检索词,搜索引擎从兴趣点(POI,Point of Interest)数据库(以下简称为母库)中搜索出与该检索词匹配的一些POI数据,用户从中选择满足自己意图的那条POI结果。作为用户,通常都希望满足自己意图的POI数据在结果排序中处于靠前的位置,对此,搜索引擎做出了一些相应的排序优化。一种优化排序的思路是:构造一个POI常用检索词数据库(以下简称为词库),用于记录常用检索词与POI数据的对应关系。当用户输入一个检索词时,搜索引擎同时从母库和词库中搜索POI数据,其中,母库的搜索方式不变,而词库的搜索要求用户输入的检索词与词库中的检索词完全匹配。这样,从词库获得的搜索结果就是精确匹配用户的搜索需求的,排序时优先于母库的搜索结果。

[0003] 以上方法的核心是构建和更新POI常用检索词库,目前主要有以下两种方法。

[0004] 第一种方法是自造:对母库中的POI数据进行分词处理,再根据词频、词的类别、词的角色等信息获得词的权重,选取权重最高的一个关键词作为检索词,而该POI数据就是这个检索词对应的POI数据。当母库中增加了新的POI数据时,按照构造词库的方法对词库进行更新。

[0005] 第二种方法是人工日志分析:人工对日志进行统计分析,找出用户经常输入的检索词和与之匹配的搜索结果,构造常用检索词库。当搜索日志更新时,筛选出新增检索词,按照构造词库的方法对词库进行更新。

[0006] 以上两种方法存在如下问题:

[0007] 1)自造法是一种理论上可行的方法,但由于其检索词并非来自用户输入,因而不能完全代表真实的搜索情况;另外,由于母库中的全部POI数据都参与词库的更新,造成词库数据和母库数据一样庞大,不但不能利用常用的检索词来优化排序,而且造成了搜索过程的重复,降低了搜索效率。

[0008] 2)人工日志分析法利用了真实的搜索场景,且词库中的检索词都是用户曾经搜索过的词,因而是常用词,有实际的指导意义。但是采用人工分析的方法,需要耗费大量人力,成本高、时间效率低,并且人的主观意愿会对结果有一定的干扰。

发明内容

[0009] 为解决上述技术问题,本发明实施例提供了一种基于检索日志的POI常用检索词分类方法、装置。

[0010] 本发明实施例提供的基于检索日志的POI常用检索词分类方法,包括:

[0011] 从日志中提取出新增检索词;

- [0012] 对所述新增检索词进行多种类别的筛选,以过滤掉所述新增检索词中的垃圾词;
- [0013] 利用多种搜索引擎对过滤后的所述新增检索词进行POI检索,合并检索结果形成所述新增检索词对应的POI数据;
- [0014] 计算所述检索词与所述POI数据的匹配关系,并依据所述匹配关系对所述检索词进行分类;
- [0015] 选择相应类别的检索词以更新POI数据分类运营表。
- [0016] 本发明实施例中,所述对所述新增检索词进行多种类别的筛选,以过滤掉所述新增检索词中的垃圾词,包括:
- [0017] 将所述新增检索词逐条与POI信息库分类表中的各个分类名称进行字符对比;若检索词与分类名称完全匹配,则将所述检索词从新增检索词中删除;和/或,
- [0018] 逐条验证所述新增检索词是否为预设的公交车线路/地铁线路或预设的公交车站点/地铁站点;是时,将所述检索词从所述新增检索词中删除;和/或,
- [0019] 逐条验证所述新增检索词是否指示路线查询;是时,将所述检索词从所述新增检索词中删除;和/或,
- [0020] 逐条验证所述新增检索词是否由区域和检索词组成;是时,将所述检索词从所述新增检索词中删除。
- [0021] 本发明实施例中,所述利用多种搜索引擎对过滤后的所述新增检索词进行POI检索,合并检索结果形成所述新增检索词对应的POI数据,包括:
- [0022] 将过滤后的所述新增检索词逐条进入POI信息库或第一搜索引擎中搜索,获取结果列表;将结果列表为空的检索词汇总作为第二引擎搜索的输入;将结果列表不为空的检索词汇总以待检索词合并;
- [0023] 将所述输入的检索词逐条进入第二搜索引擎中搜索,获取结果列表;对于结果列表为空的检索词,提示其他搜索城市选择的检索词汇总生成非归属地数据;将未提示其他搜索城市选择的检索词汇总生成高额赏金任务;对于结果列表不为空的检索词,将首条结果的POI名称作为替换检索词,并与原检索词建立关联关系;
- [0024] 将替换后检索词逐条进入所述POI信息库的第三搜索引擎中搜索,获取结果列表;将结果列表为空的检索词汇总生成查无数据;将结果列表不为空的检索词汇总以待检索词合并;
- [0025] 将以待合并的检索词进行合并,生成待匹配检索词。
- [0026] 本发明实施例中,所述计算所述检索词与所述POI数据的匹配关系,并依据所述匹配关系对所述检索词进行分类,包括:
- [0027] 逐条判断所述待匹配检索词搜索结果中,是否存在任意一条poi名称包含完整的检索词;若不存在,保留全部搜索结果;
- [0028] 若存在,提取出poi名称包含完整的检索词类数据;若包含完整的检索词的poi名称只有一条,更新POI常用检索词库,并将其匹配状态指示为一对一类成功匹配,同时更新一对一匹配文件;若包含完整的检索词的poi名称有多条,且多条数据类别属于同一个分类时,更新POI常用检索词库,并将其匹配状态指示为一对多类成功匹配,同时更新一对多匹配文件;若多条数据类别属于多个分类,则判断各个分类的占比,若存在分类占比大于等于第一阈值时,则将该分类数据与检索词匹配,更新POI常用检索词库,并将其匹配状态指示

为一对多类成功匹配,同时更新一对多匹配文件;否则将数据放入无法识别类文件中;

[0029] 计算剩余的待匹配检索词与其对应的搜索结果的匹配度,若检索词与所有结果的匹配度均低于第二阈值,则此检索词属于查无类,将所述检索词添加至查无数据文件中;若检索词与结果的匹配度高于第二阈值,将所述检索词与所述结果匹配,添加至人工待处理文件中。

[0030] 本发明实施例中,所述选择相应类别的检索词以更新POI数据分类运营表,包括:

[0031] 对于一对多匹配文件中的检索词,逐条提取匹配结果在POI信息库中的数据类别,生成待分类运营文件;

[0032] 根据待分类运营文件中的数据类别,在POI信息库分类运营表中确定对应的类别运营种类,并汇总形成自然增加类别的检索词数据文件、根据条件增加类别的检索词数据文件和独立词数据文件;

[0033] 根据预制条件把条件增加类文件中的检索词分别添加到自然增加类文件和独立词数据文件中;

[0034] 根据自然增加类别文件中的检索词,自然增加类别,更新POI信息库分类运营表;并根据新增分类情况,更新POI信息库中的数据类别;以及根据独立词数据文件中的检索词,更新独立词词库。

[0035] 本发明实施例提供的基于检索日志的POI常用检索词分类装置,包括:

[0036] 检索词提取模块,用于从日志中提取出新增检索词;

[0037] 检索词过滤模块,用于对所述新增检索词进行多种类别的筛选,以过滤掉所述新增检索词中的垃圾词;

[0038] 多引擎搜索模块,用于利用多种搜索引擎对过滤后的所述新增检索词进行POI检索,合并检索结果形成所述新增检索词对应的POI数据;

[0039] 类型匹配模块,用于计算所述检索词与所述POI数据的匹配关系,并依据所述匹配关系对所述检索词进行分类;

[0040] 分类运营模块,用于选择相应类别的检索词以更新POI数据分类运营表。

[0041] 本发明实施例中,所述检索词过滤模块,还用于将所述新增检索词逐条与POI信息库分类表中的各个分类名称进行字符对比;若检索词与分类名称完全匹配,则将所述检索词从新增检索词中删除;和/或,逐条验证所述新增检索词是否为预设的公交车线路/地铁线路或预设的公交车站点/地铁站点;是时,将所述检索词从所述新增检索词中删除;和/或,逐条验证所述新增检索词是否指示路线查询;是时,将所述检索词从所述新增检索词中删除;和/或,逐条验证所述新增检索词是否由区域和检索词组成;是时,将所述检索词从所述新增检索词中删除。

[0042] 本发明实施例中,所述多引擎搜索模块,还用于将过滤后的所述新增检索词逐条进入POI信息库或第一搜索引擎中搜索,获取结果列表;将结果列表为空的检索词汇总作为第二引擎搜索的输入;将结果列表不为空的检索词汇总以待检索词合并;将所述输入的检索词逐条进入第二搜索引擎中搜索,获取结果列表;对于结果列表为空的检索词,提示其他搜索城市选择的检索词汇总生成非归属地数据;将未提示其他搜索城市选择的检索词汇总生成高额赏金任务;对于结果列表不为空的检索词,将首条结果的POI名称作为替换检索词,并与原检索词建立关联关系;将替换后检索词逐条进入所述POI信息库的第三搜索引擎

中搜索,获取结果列表;将结果列表为空的检索词汇生成查无数据;将结果列表不为空的检索词汇以待检索词合并;将以待合并的检索词进行合并,生成待匹配检索词。

[0043] 本发明实施例中,所述类型匹配模块,还用于逐条判断所述待匹配检索词搜索结果中,是否存在任意一条poi名称包含完整的检索词;若不存在,保留全部搜索结果;若存在,提取出poi名称包含完整的检索词类数据;若包含完整的检索词的poi名称只有一条,更新POI常用检索词库,并将其匹配状态指示为一对一类成功匹配,同时更新一对一匹配文件;若包含完整的检索词的poi名称有多条,且多条数据类别属于同一个分类时,更新POI常用检索词库,并将其匹配状态指示为一对多类成功匹配,同时更新一对多匹配文件;若多条数据类别属于多个分类,则判断各个分类的占比,若存在分类占比大于等于第一阈值时,则将该分类数据与检索词匹配,更新POI常用检索词库,并将其匹配状态指示为一对多类成功匹配,同时更新一对多匹配文件;否则将数据放入无法识别类文件中;计算剩余的待匹配检索词与其对应的搜索结果的匹配度,若检索词与所有结果的匹配度均低于第二阈值,则此检索词属于查无类,将所述检索词添加至查无数据文件中;若检索词与结果的匹配度高于第二阈值,将所述检索词与所述结果匹配,添加至人工待处理文件中。

[0044] 本发明实施例中,所述分类运营模块,还用于对于一对多匹配文件中的检索词,逐条提取匹配结果在POI信息库中的数据类别,生成待分类运营文件;根据待分类运营文件中的数据类别,在POI信息库分类运营表中确定对应的类别运营种类,并汇总形成自然增加类别的检索词数据文件、根据条件增加类别的检索词数据文件和独立词数据文件;根据预制条件把条件增加类文件中的检索词分别添加到自然增加类文件和独立词数据文件中;根据自然增加类别文件中的检索词,自然增加类别,更新POI信息库分类运营表;并根据新增分类情况,更新POI信息库中的数据类别;以及根据独立词数据文件中的检索词,更新独立词词库。

[0045] 本发明实施例的技术方案中,从日志中提取出新增检索词;对所述新增检索词进行多种类别的筛选,以过滤掉所述新增检索词中的垃圾词;利用多种搜索引擎对过滤后的所述新增检索词进行POI检索,合并检索结果形成所述新增检索词对应的POI数据;计算所述检索词与所述POI数据的匹配关系,并依据所述匹配关系对所述检索词进行分类;选择相应类别的检索词以更新POI数据分类运营表。可见,本发明实施例基于人工智能的思想,构建了一套完整的常用检索词分析和运营系统。从检索词提取、检索词过滤、多引擎搜索、类型匹配到分类运营,整个过程自动地进行,不再需要人工的参与,大大提高了时间效率。由于不再需要人工的参与,本发明能大大节省人力成本。使用者可根据实际场景增加或替换相应的词库、搜索引擎、匹配算法、词的类别等,扩展灵活,实用性强。

附图说明

[0046] 图1为本发明实施例的基于检索日志的POI常用检索词分类方法的流程示意图;

[0047] 图2为本发明实施例的多引擎搜索模块的处理流程图;

[0048] 图3为本发明实施例的类型匹配模块的处理流程图;

[0049] 图4为本发明实施例的分类运营模块的处理流程图;

[0050] 图5为本发明实施例的基于检索日志的POI常用检索词分类装置中各个模块的处理流程图;

[0051] 图6为本发明实施例的基于检索日志的POI常用检索词分类装置的结构组成示意图。

具体实施方式

[0052] 为了能够更加详尽地了解本发明的特点与技术内容,下面结合附图对本发明的实现进行详细阐述,所附附图仅供参考说明之用,并非用来限定本发明。

[0053] 本发明实施例采用自动化的常用检索词库构造和更新方法,至少解决了如下技术问题:如何从日志中提取出新增检索词。如何过滤掉新增检索词中的垃圾词。如何获取与检索词对应的POI数据。如何计算检索词与POI数据的匹配关系,并依据匹配关系对检索词进行分类。如何对常用检索词进行分类运营,更新POI常用检索词库。

[0054] 本发明实施例的基于检索日志的POI常用检索词分类装置由检索词提取模块、检索词过滤模块、多引擎搜索模块、类型匹配模块、分类运营模块五大模块构成,依次对上述五个问题进行了解决。通过所述基于检索日志的POI常用检索词分类装置能够执行基于检索日志的POI常用检索词分类方法。

[0055] 为了便于理解本发明实施例,现将与本发明实施例相关的术语做以下描述:

[0056] POI信息库:记录POI数据的详细信息,包括id、地理位置、分类等,简称母库。

[0057] POI信息库分类运营表:记录POI数据的分类信息,包括大、中、小三层分类。

[0058] 历史检索词文件:记录所有的历史检索词。

[0059] POI常用检索词库:记录常用检索词与POI数据的对应关系。

[0060] 图1为本发明实施例的基于检索日志的POI常用检索词分类方法的流程示意图,如图1所示,所述基于检索日志的POI常用检索词分类方法包括以下步骤:

[0061] 步骤101:从日志中提取出新增检索词。

[0062] 检索词提取模块从日志中提取检索词信息,并和历史数据去重形成新增检索词数据。检索词提取模块为检索词过滤模块提供输入数据。

[0063] 1)日志提取

[0064] 日志分析系统设置自动触发任务,每天固定的时间开始从日志文件中全量提取当天的用户检索词,并按照城市进行分组,分组后每个城市的检索词均进行去重及次数统计,按城市生成包括“检索词、搜索次数”两个字段的中间文件。

[0065] 2)新增检索词提取

[0066] 以城市为单位,将当天形成的中间过程文件与历史检索词文件中的检索词进行对比:对于历史检索词文件中已经存在的记录,直接在历史检索词文件中进行搜索次数累计,并将该检索词从中间过程文件中删除;对于历史检索词文件中不存在的记录,将其加入到历史检索词文件中,在中间过程文件中保留。完成所有检索词的比对后,中间过程文件剩余内容即为该城市的新增检索词,该文件称为新增检索词文件。

[0067] 步骤102:对所述新增检索词进行多种类别的筛选,以过滤掉所述新增检索词中的垃圾词。

[0068] 检索词过滤模块对新增检索词进行多种类别的筛选,过滤掉检索词中的垃圾词。检索词过滤模块为多引擎搜索模块提供数据。将所述新增检索词逐条与POI信息库分类表中的各个分类名称进行字符对比;若检索词与分类名称完全匹配,则将所述检索词从新增

检索词中删除;和/或,逐条验证所述新增检索词是否为预设的公交车线路/地铁线路或预设的公交车站点/地铁站;是时,将所述检索词从所述新增检索词中删除;和/或,逐条验证所述新增检索词是否指示路线查询;是时,将所述检索词从所述新增检索词中删除;和/或,逐条验证所述新增检索词是否由区域和检索词组成;是时,将所述检索词从所述新增检索词中删除。

[0069] 1)分类名称排重

[0070] 将新增检索词逐条与POI信息库分类表中的各大、中、小分类名称进行字符对比:若某个检索词与分类名称完全匹配,则将其从新增检索词文件中删除。

[0071] 2)垃圾词-公交线路类筛选

[0072] 逐条验证新增检索词是否符合“垃圾词-公交线路类”的特征:检索词是某一路公交车/地铁线路或具体某公交车/地铁站。将符合“垃圾词-公交线路类”特征的检索词全部从新增检索词文件中删除。

[0073] 3)垃圾词-拆词类筛选

[0074] 逐条验证新增检索词是否符合“垃圾词-拆词类”的特征:用户的查询意图为路线查询情况。如,A地点到B地点,A地点至B地点,A地点去B地点,到A地点,至A地点去A地点等等。将符合“垃圾词-拆词类”特征的检索词全部从新增检索词文件中删除。

[0075] 4)垃圾词-包含类筛选

[0076] 逐条验证新增检索词是否符合“垃圾词-包含类”的特征:检索词由区域+检索词组成。将符合“垃圾词-包含类”特征的检索词全部从新增检索词文件中删除。此时的新增检索词文件为已过滤的检索词文件。

[0077] 步骤103:利用多种搜索引擎对过滤后的所述新增检索词进行POI检索,合并检索结果形成所述新增检索词对应的POI数据。

[0078] 多引擎搜索模块同时利用系统的POI搜索引擎、百度地图、高德地图等多种搜索引擎进行搜索,合并结果形成新增检索词对应的POI数据。多引擎搜索模块为类型匹配模块提供输入数据。将过滤后的所述新增检索词逐条进入POI信息库或第一搜索引擎中搜索,获取结果列表;将结果列表为空的检索词汇总作为第二搜索引擎的输入;将结果列表不为空的检索词汇总以待检索词合并;将所述输入的检索词逐条进入第二搜索引擎中搜索,获取结果列表;对于结果列表为空的检索词,提示其他搜索城市选择的检索词汇总生成非归属地数据;将未提示其他搜索城市选择的检索词汇总生成高额赏金任务;对于结果列表不为空的检索词,将首条结果的POI名称作为替换检索词,并与原检索词建立关联关系;将替换后检索词逐条进入所述POI信息库的第三搜索引擎中搜索,获取结果列表;将结果列表为空的检索词汇总生成查无数据;将结果列表不为空的检索词汇总以待检索词合并;将以待合并的检索词进行合并,生成待匹配检索词。

[0079] 详细的流程参照图2所示,所述流程包括:

[0080] 1)过滤后POI/高德搜索引擎搜索

[0081] 将已过滤的检索词逐条进入本系统的POI信息库或高德的模糊搜索引擎中搜索,获取结果列表;将结果列表为空的检索词汇总作为百度地图引擎搜索的输入;将结果列表不为空的检索词汇总作为检索词合并的输入。

[0082] 2)百度地图引擎搜索

[0083] 将输入的检索词逐条进入百度地图的搜索引擎中搜索,获取结果列表。对于百度搜索结果列表为空的检索词:将提示其他搜索城市选择的检索词汇生成非归属地数据文件;将未提示其他搜索城市选择的检索词汇生成高额赏金任务文件。对于百度搜索后有结果的检索词,将首条结果的POI名称作为替换检索词,并与原检索词建立关联关系。此时的检索词文件成为替换后检索词文件,包含原检索词、替换检索词、搜索次数三个字段。

[0084] 3)替换后POI搜索引擎搜索

[0085] 将替换后检索词逐条进入本系统的POI信息库的模糊搜索引擎中搜索,获取结果列表:将结果列表为空的检索词汇生成查无数据文件1;将结果列表不为空的检索词汇总作为检索词合并的输入。

[0086] 4)检索词合并

[0087] 将1)和3)输出的检索词文件进行合并,生成待匹配检索词文件,此文件包含原检索词、替换检索词、搜索次数三个字段。

[0088] 步骤104:计算所述检索词与所述POI数据的匹配关系,并依据所述匹配关系对所述检索词进行分类。

[0089] 类型匹配模块将待匹配的检索词分成一对一匹配类、一对多匹配类、无法识别类、查无类四大类,形成检索词库数据,更新POI常用检索词库。四类数据的含义如下:

[0090] 一对一匹配:有且仅有一条POI数据与该检索词匹配。

[0091] 一对多匹配:有多条POI数据与该检索词匹配。经算法判断,当搜索结果中可建立匹配关系的POI名称为特指提供某类服务或具备相同职能的某一类POI的名称,或当检索词为某一品牌名称时,可建立一对多匹配关系。

[0092] 查无:该检索词搜索不到结果或者搜索到的结果与检索词不匹配

[0093] 无法识别:检索词的指向不明,可以搜索出多条完全包含检索词的POI数据,但是这些POI数据分散在各个不同的分类中。

[0094] 逐条判断所述待匹配检索词搜索结果中,是否存在任意一条poi名称包含完整的检索词;若不存在,保留全部搜索结果;

[0095] 若存在,提取出poi名称包含完整的检索词类数据:若包含完整的检索词的poi名称只有一条,更新POI常用检索词库,并将其匹配状态指示为一对一类成功匹配,同时更新一对一匹配文件;若包含完整的检索词的poi名称有多条,且多条数据类别属于同一个分类时,更新POI常用检索词库,并将其匹配状态指示为一对多类成功匹配,同时更新一对多匹配文件;若多条数据类别属于多个分类,则判断各个分类的占比,若存在分类占比大于等于第一阈值时,则将该分类数据与检索词匹配,更新POI常用检索词库,并将其匹配状态指示为一对多类成功匹配,同时更新一对多匹配文件;否则将数据放入无法识别类文件中;

[0096] 计算剩余的待匹配检索词与其对应的搜索结果的匹配度,若检索词与所有结果的匹配度均低于第二阈值,则此检索词属于查无类,将所述检索词添加至查无数据文件中;若检索词与结果的匹配度高于第二阈值,将所述检索词与所述结果匹配,添加至人工待处理文件中。

[0097] 具体地,1)包含完整输入判断

[0098] 逐条判断待匹配检索词搜索结果中,是否存在任意一条poi名称,包含完整的检索词(原始内容,未分词状态),如检索词为东北大学,所有“%东北大学%”都算,但是“东北%

大学”，不算。若不存在，保留全部搜索结果进入下一步。若存在，提取出“%东北大学%”类数据：若“%东北大学%”只有一条，更新POI常用检索词库，并将其匹配状态确定为“成功匹配-一对一类”，同时更新一对一匹配文件；若“%东北大学%”有多条，判断数据类别，属于同一个分类，更新POI常用检索词库，并将其匹配状态确定为“成功匹配-一对多类”，同时更新一对多匹配文件；若属于多个分类，判断分类占比，若存在某个分类占比大于等于阈值（可根据实际情况设定），则将此分类数据与检索词匹配，更新POI常用检索词库，并将其匹配状态确定为“成功匹配-一对多类”，同时更新一对多匹配文件；否则数据放入无法识别类文件中。

[0099] 处理完之后，将包含完整检索词类数据从待匹配检索词文件中删除。

[0100] 2)计算检索词与所有结果的匹配度

[0101] 计算经过1)剩余的待匹配检索词与其对应的搜索结果的匹配度，其中匹配度可采用编辑距离、余弦相似度、Jaccard相似度等多种方法来衡量。若检索词与所有结果的匹配度均低于N%（可根据实际情况设定），则此检索词属于查无类，加入到查无数据文件2中；若检索词与某些结果的匹配度高于N%，将检索词与这部分结果匹配，放入到人工待处理文件中。

[0102] 步骤105:选择相应类别的检索词以更新POI数据分类运营表。

[0103] 分类运营模块用识别为一对多的数据来更新POI信息库分类运营表。本模块的数据来源于类型匹配模块中的一对多匹配文件。

[0104] 对于一对多匹配文件中的检索词，逐条提取匹配结果在POI信息库中的数据类别，生成待分类运营文件；

[0105] 根据待分类运营文件中的数据类别，在POI信息库分类运营表中确定对应的类别运营种类，并汇总形成自然增加类别的检索词数据文件、根据条件增加类别的检索词数据文件和独立词数据文件；

[0106] 根据预制条件把条件增加类文件中的检索词分别添加到自然增加类文件和独立词数据文件中；

[0107] 根据自然增加类别文件中的检索词，自然增加类别，更新POI信息库分类运营表；并根据新增分类情况，更新POI信息库中的数据类别；以及根据独立词数据文件中的检索词，更新独立词词库。

[0108] 具体地，1)数据类别提取

[0109] 对于一对多匹配文件中的检索词，逐条提取它的匹配结果在POI信息库中的数据类别（即分类ID），生成待分类运营文件，此文件包含原检索词、替换检索词、搜索次数、匹配ID、匹配状态、匹配分类ID等字段。

[0110] 2)分类运营状态判断

[0111] 根据待分类运营文件中的数据类别，在POI信息库分类运营表中确定对应的类别运营种类（自然增加/根据条件增加/独立词运营）。根据以上条件判断，汇总形成自然增加类别的检索词数据文件、根据条件增加类别的检索词数据文件和独立词数据文件。文件包含的字段有原检索词、替换检索词、搜索次数、匹配ID、匹配状态、匹配分类ID、分类运营状态。

[0112] 3)条件增加类别文件处理

[0113] 根据预制条件把条件增加类文件中的检索词分别添加到自然增加类文件和独立词数据文件中。预制条件可以根据实际情况进行设定,例如把预制条件设定为检索词的搜索次数,当搜索次数超过某个阈值时就把该检索词添加到自然增加类文件,低于这个阈值则添加到独立词文件。

[0114] 4)自然增加类别文件处理

[0115] 根据自然增加类别文件中的检索词,自然增加类别,更新POI信息库分类运营表;并根据新增分类情况,更新POI信息库中的数据类别。

[0116] 5)独立词数据文件处理

[0117] 根据独立词数据文件中的检索词,更新独立词词库。

[0118] 图6为本发明实施例的基于检索日志的POI常用检索词分类装置的结构组成示意图,所述装置包括:

[0119] 检索词提取模块61,用于从日志中提取出新增检索词;

[0120] 检索词过滤模块62,用于对所述新增检索词进行多种类别的筛选,以过滤掉所述新增检索词中的垃圾词;

[0121] 多引擎搜索模块63,用于利用多种搜索引擎对过滤后的所述新增检索词进行POI检索,合并检索结果形成所述新增检索词对应的POI数据;

[0122] 类型匹配模块64,用于计算所述检索词与所述POI数据的匹配关系,并依据所述匹配关系对所述检索词进行分类;

[0123] 分类运营模块65,用于选择相应类别的检索词以更新POI数据分类运营表。

[0124] 所述检索词过滤模块62,还用于将所述新增检索词逐条与POI信息库分类表中的各个分类名称进行字符对比;若检索词与分类名称完全匹配,则将所述检索词从新增检索词中删除;和/或,逐条验证所述新增检索词是否为预设的公交车线路/地铁线路或预设的公交车站点/地铁站点;是时,将所述检索词从所述新增检索词中删除;和/或,逐条验证所述新增检索词是否指示路线查询;是时,将所述检索词从所述新增检索词中删除;和/或,逐条验证所述新增检索词是否由区域和检索词组成;是时,将所述检索词从所述新增检索词中删除。

[0125] 所述多引擎搜索模块63,还用于将过滤后的所述新增检索词逐条进入POI信息库或第一搜索引擎中搜索,获取结果列表;将结果列表为空的检索词汇总作为第二引擎搜索的输入;将结果列表不为空的检索词汇总以待检索词合并;将所述输入的检索词逐条进入第二搜索引擎中搜索,获取结果列表;对于结果列表为空的检索词,提示其他搜索城市选择的检索词汇总生成非归属地数据;将未提示其他搜索城市选择的检索词汇总生成高额赏金任务;对于结果列表不为空的检索词,将首条结果的POI名称作为替换检索词,并与原检索词建立关联关系;将替换后检索词逐条进入所述POI信息库的第三搜索引擎中搜索,获取结果列表;将结果列表为空的检索词汇总生成查无数据;将结果列表不为空的检索词汇总以待检索词合并;将以待合并的检索词进行合并,生成待匹配检索词。

[0126] 所述类型匹配模块64,还用于逐条判断所述待匹配检索词搜索结果中,是否存在任意一条poi名称包含完整的检索词;若不存在,保留全部搜索结果;若存在,提取出poi名称包含完整的检索词类数据;若包含完整的检索词的poi名称只有一条,更新POI常用检索词库,并将其匹配状态指示为一对一类成功匹配,同时更新一对一匹配文件;若包含完整的

检索词的poi名称有多条,且多条数据类别属于同一个分类时,更新POI常用检索词库,并将其匹配状态指示为一对多类成功匹配,同时更新一对多匹配文件;若多条数据类别属于多个分类,则判断各个分类的占比,若存在分类占比大于等于第一阈值时,则将该分类数据与检索词匹配,更新POI常用检索词库,并将其匹配状态指示为一对多类成功匹配,同时更新一对多匹配文件;否则将数据放入无法识别类文件中;计算剩余的待匹配检索词与其对应的搜索结果的匹配度,若检索词与所有结果的匹配度均低于第二阈值,则此检索词属于查无类,将所述检索词添加至查无数据文件中;若检索词与结果的匹配度高于第二阈值,将所述检索词与所述结果匹配,添加至人工待处理文件中。

[0127] 所述分类运营模块65,还用于对于一对多匹配文件中的检索词,逐条提取匹配结果在POI信息库中的数据类别,生成待分类运营文件;根据待分类运营文件中的数据类别,在POI信息库分类运营表中确定对应的类别运营种类,并汇总形成自然增加类别的检索词数据文件、根据条件增加类别的检索词数据文件和独立词数据文件;根据预制条件把条件增加类文件中的检索词分别添加到自然增加类文件和独立词数据文件中;根据自然增加类别文件中的检索词,自然增加类别,更新POI信息库分类运营表;并根据新增分类情况,更新POI信息库中的数据类别;以及根据独立词数据文件中的检索词,更新独立词词库。

[0128] 本发明实施例采用了自动分析的方法,至少包括如下优点:高效率:基于人工智能的思想,构建了一套完整的常用检索词分析和运营系统。从检索词提取、检索词过滤、多引擎搜索、类型匹配到分类运营,整个过程自动地进行,不再需要人工的参与,大大提高了时间效率。低成本:由于不再需要人工的参与,本发明能大大节省人力成本。本发明在辽宁位置基地POI搜索系统中的应用,帮助整个项目节省了40%的人力成本。扩展灵活:提供了一种基于日志的检索词分类运营的方法和系统,使用者可根据实际场景增加或替换相应的词库、搜索引擎、匹配算法、词的类别等,扩展灵活,实用性强。

[0129] 参照图5,图5为本发明实施例基于检索日志的POI常用检索词分类装置中各个模块的处理流程示意图,包括:检索词提取模块M1、检索词过滤模块M2、多引擎搜索模块M3、类型匹配模块M4、分类运营模块M5。

[0130] 检索词提取模块M1

[0131] 检索词提取模块从日志中提取检索词信息,并和历史数据去重形成新增检索词数据。检索词提取模块为检索词过滤模块提供输入数据。

[0132] (1)日志提取

[0133] 日志分析系统设置自动触发任务,每天凌晨开始从日志文件中,全量提取当天的用户检索词,并按照城市进行分组,分组后每个城市的检索词均进行排重及次数统计。本部分输入日志文件,输出按城市生成的中间文件D1,D1包括“检索词、搜索次数”两个字段。

[0134] (2)新增检索词提取

[0135] 以城市为单位,将当天形成的中间过程文件D1与历史检索词文件H1进行对比:对于历史检索词文件中已经存在的记录,直接在历史检索词文件中进行搜索次数累计,并将该检索词从中间过程文件中删除;对于历史检索词文件中不存在的记录,将其加入到历史检索词文件中,在中间过程文件中保留。完成所有检索词的比对后,中间过程文件剩余内容D2即为该城市的新增检索词。本部分输入历史检索词文件H1和中间文件D1,输出更新后的历史检索词文件H2,以城市为单位的新增检索词文件D2,D2包括“检索词、搜索次数”两个字

段。

[0136] 检索词过滤模块M2

[0137] 检索词过滤模块对检索词进行多种类别的筛选,过滤出后续进行检索词匹配操作的数据。检索词过滤模块为多引擎搜索模块提供数据。

[0138] (1)分类名称排重

[0139] XX城市待匹配检索词逐条与POI信息库分类运营表中的各大、中、小分类名称进行字符对比:将字符100%相同的检索词与分类名称对应的分类代码建立匹配关系,形成文件1A,匹配状态确定为“成功匹配-可对照分类”,匹配结果为匹配上的分类代码;将字符不相同的检索词汇总,形成文件1B;本部分输入POI信息库分类运营表、某一城市待匹配检索词文件A0(一般是M1模块产生的D2文件),A0包括“检索词、搜索次数”两个字段;输出文件1A(检索词、搜索次数、匹配状态、匹配结果)和文件1B(检索词、搜索次数)。

[0140] (2)垃圾词-公交线路类筛选

[0141] 文件1B中的检索词逐条验证是否符合“垃圾词-公交线路类”的特征:“检索词是某一路公交车/地铁线路或具体某公交车/地铁站点”。将符合“垃圾词-公交线路类”特征的检索词全部筛选出来,汇总并将匹配状态确定为“垃圾词-公交线路类”,形成文件2A;将不符合“垃圾词-公交线路类”特征的检索词汇总,形成文件2B。本部分输入文件1B,输出文件2A(检索词、搜索次数、匹配状态)和文件2B(检索词、搜索次数)。

[0142] (3)垃圾词-拆词类筛选

[0143] 文件2B中的检索词逐条验证是否符合“垃圾词-拆词类”的特征:“用户的查询意图为路线查询情况”。如,A地点到B地点,A地点至B地点,A地点去B地点,到A地点,至A地点去A地点等等。将符合“垃圾词-拆词类”特征的检索词全部筛选出来,汇总并将匹配状态确定为“垃圾词-拆词类”,形成文件3A;将不符合“垃圾词-拆词类”特征的检索词汇总,形成文件3B。本部分输入文件2B,输出文件3A(检索词、搜索次数、匹配状态)和文件3B(检索词、搜索次数)。

[0144] (4)垃圾词-包含类筛选

[0145] 文件3B中的检索词逐条验证是否符合“垃圾词-包含类”的特征:“检索词由区域+检索词组成”。将符合“垃圾词-包含类”特征的检索词全部筛选出来,汇总并将匹配状态确定为“垃圾词-包含类”,形成文件4A;将不符合“垃圾词-包含类”特征的检索词汇总,形成文件4B。本部分输入文件3B,输出文件4A(检索词、搜索次数、匹配状态)和文件4B(检索词、搜索次数)。

[0146] 多引擎搜索模块M3

[0147] 多引擎搜索模块同时利用了本系统的POI搜索引擎、百度地图、高德地图的搜索引擎进行搜索,合并结果形成需要类型匹配的数据。多引擎搜索模块为类型匹配模块提供输入数据。多引擎搜索模块的处理流程还可以参照图2,包括:

[0148] (1)过滤后POI/高德搜索引擎搜索

[0149] 文件4B中的检索词逐条进入本系统的POI信息库或高德的模糊搜索引擎中搜索,获取结果列表:将结果列表为空的检索词汇总,形成5A;将结果列表不为空的检索词汇总,形成5B。本部分输入文件4B输出文件5A(检索词、搜索次数)和文件5B(检索词、搜索次数)。

[0150] (2)百度引擎搜索

[0151] 文件5A中的检索词逐条进入百度地图的搜索引擎中搜索,获取结果列表。将百度结果列表为空的检索词汇总,形成6A;将提示其他搜索城市选择的检索词汇总,形成7A,其中检索词的匹配状态逐条确定为“非归属地数据”;将未提示其他搜索城市选择的检索词汇总,形成7B,其中检索词的匹配状态逐条确定为“高额赏金任务”。将在百度搜索后有结果的结果列表,形成6B;将首条结果的POI名称作为替换检索词,并与原检索词建立关联关系,最后进行汇总,形成8A。本部分输入文件5A,输出文件6A(检索词、搜索次数)、文件6B(检索词、搜索次数)、非归属地数据文件7A(检索词、搜索次数、匹配状态、城市建议)、高额赏金任务文件7B(检索词、搜索次数、匹配状态)和替换后的文件8A(原检索词、替换检索词、搜索次数)。

[0152] (3)替换后POI搜索引擎搜索

[0153] 文件8A中的替换检索词逐条进入基地POI信息库的模糊搜索引擎中搜索,获取结果列表:将结果列表为空的检索词汇总,形成9A;将9A中的检索词的匹配状态逐条确定为“查无数据”;将结果列表不为空的检索词汇总,形成9B。本部分输入文件8A,输出文件9A(原检索词、替换检索词、搜索次数、匹配状态)和文件9B(原检索词、替换检索词、搜索次数)。

[0154] (4)检索词合并

[0155] 文件5B和文件9B中的检索词进行汇总,形成文件10B。用相同格式汇总文件,必须保留文件9B中原检索词和替换词两列数据。对替换词进行标识,在后续操作过程中,有替换词的使用替换词在基地POI模糊搜索引擎中搜索,但最终的匹配结果则使用原检索词进行匹配。本部分输入文件5B、文件9B,输出文件10B(原检索词、替换检索词、搜索次数)。

[0156] 类型匹配模块M4

[0157] 类型匹配模块将待匹配的检索词分成一对一匹配类、一对多匹配类、无法识别类、查无类四大类,形成检索词库数据,更新POI常用检索词库。类型匹配模块的处理流程还可以参照图3,包括:

[0158] (1)包含完整输入判断

[0159] 逐条判断检索词搜索结果中,是否存在任意一条poi名称,包含完整的检索词(原始内容,未分词状态),如检索词为东北大学,所有“%东北大学%”都算,但是“东北%大学”,不算。若不存在,保留全部搜索结果进入下一步。若存在,保留“%东北大学%”类数据:若“%东北大学%”只有一条,更新POI常用检索词库,并将其匹配状态确定为“成功匹配—一对一类”,同时加入到一对一匹配12A中;若“%东北大学%”有多条,判断数据类别,属于同一个分类,更新POI常用检索词库,并将其匹配状态确定为“成功匹配—一对多类”,并将其放入一对多匹配13A中;若属于多个分类,判断分类占比,若存在某个分类占比大于等于阈值(工具界面设定),则将此分类数据与检索词匹配,更新POI常用检索词库,将其匹配状态确定为“成功匹配—一对多类”,并将其放入一对多匹配13A中;否则数据放入无法识别类14A;

[0160] 处理完之后,将包含完整检索词类数据从待匹配检索词文件中删除。

[0161] (2)计算检索词与所有结果的匹配度

[0162] 计算检索词与所有搜索结果的匹配度,其中匹配度采用编辑距离来计算。若检索词与所有结果的匹配度均低于N%,则此检索词属于查无,与M3模块的查无数据汇总在一起;若检索词与某些结果的匹配度高于N%,将检索词与这部分结果匹配,放入到人工待处理文件14B中。

[0163] 分类运营模块M5

[0164] 分类运营模块用识别为一对多的数据来更新POI信息库分类运营表。本模块的数据来源于类型匹配模块中“成功匹配-一对多类”数据文件。分类运营模块的处理流程还可以参照图4,包括:

[0165] (1)数据类别提取

[0166] 文件13A中检索词的匹配结果逐条在POI信息库中提取该检索词匹配一对多的数据类别(即分类ID),生成文件15A。本部分输入文件13A,输出文件15A(原检索词、替换检索词、搜索次数、匹配ID、匹配状态、匹配分类ID)。

[0167] (2)分类运营状态判断

[0168] 根据文件15A中的数据类别,在《POI数据分类运营管理表》中确定对应的类别运营种类(自然增加/根据条件增加/独立词运营)。根据以上条件判断,汇总形成自然增加类别的检索词数据文件16A、根据条件增加类别的检索词数据文件17A和独立词数据文件18A。本部分输入文件15A和《POI数据分类运营管理表》,输出自然增加类别文件16A、条件增加类别文件17A和独立词数据文件18A(原检索词、替换检索词、搜索次数、匹配ID、匹配状态、匹配分类ID、分类运营状态)。

[0169] (3)条件增加类别文件处理

[0170] 根据预制条件判断17A中检索词,符合“增加分类”条件的检索词合并进入16A,不符合“增加分类”条件的检索词合并进入18A。本部分输入文件17A,输出更新后的文件16A和18A。

[0171] (4)自然增加类别文件处理

[0172] 根据16A文件中的检索词,自然增加类别,并更新POI数据分类表,形成19A;并根据新增分类情况,更新POI信息库中的数据类别。本部分输入文件16A,输出更新后的POI数据分类表19A和更新POI信息库中的数据类别。

[0173] (5)独立词数据文件处理

[0174] 根据18A文件中的检索词,汇总并更新独立词词库,形成文件20A。本部分输入文件18A,输出独立词词库文件20A。

[0175] 本发明实施例所记载的技术方案之间,在不冲突的情况下,可以任意组合。

[0176] 在本发明所提供的几个实施例中,应该理解到,所揭露的方法和智能设备,可以通过其它的方式实现。以上所描述的设备实施例仅仅是示意性的,例如,所述单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,如:多个单元或组件可以结合,或可以集成到另一个系统,或一些特征可以忽略,或不执行。另外,所显示或讨论的各组成部分相互之间的耦合、或直接耦合、或通信连接可以是通过一些接口,设备或单元的间接耦合或通信连接,可以是电性的、机械的或其它形式的。

[0177] 上述作为分离部件说明的单元可以是、或也可以不是物理上分开的,作为单元显示的部件可以是、或也可以不是物理单元,即可以位于一个地方,也可以分布到多个网络单元上;可以根据实际的需要选择其中的部分或全部单元来实现本实施例方案的目的。

[0178] 另外,在本发明各实施例中的各功能单元可以全部集成在一个第二处理单元中,也可以是各单元分别单独作为一个单元,也可以两个或两个以上单元集成在一个单元中;上述集成的单元既可以采用硬件的形式实现,也可以采用硬件加软件功能单元的形式实

现。

[0179] 以上所述,仅为本发明的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,可轻易想到变化或替换,都应涵盖在本发明的保护范围之内。

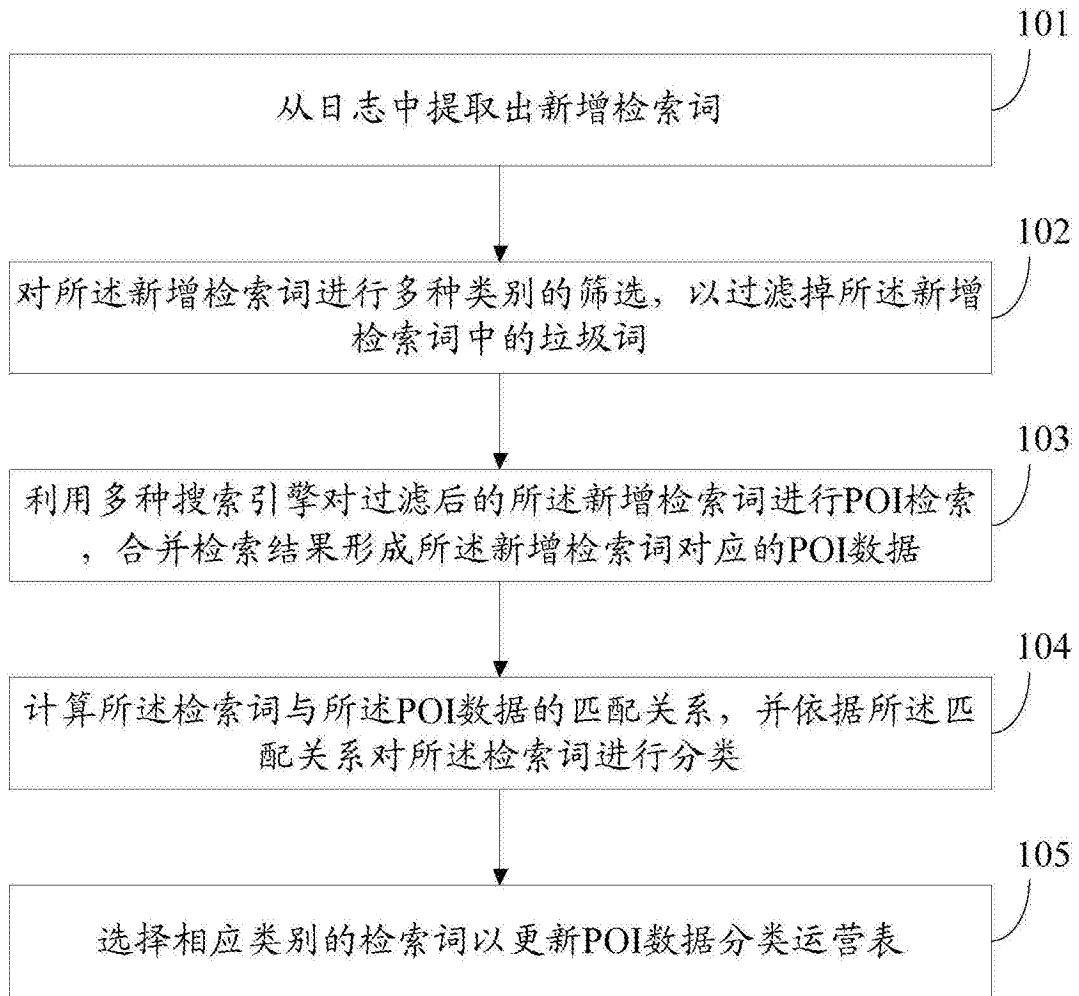


图1

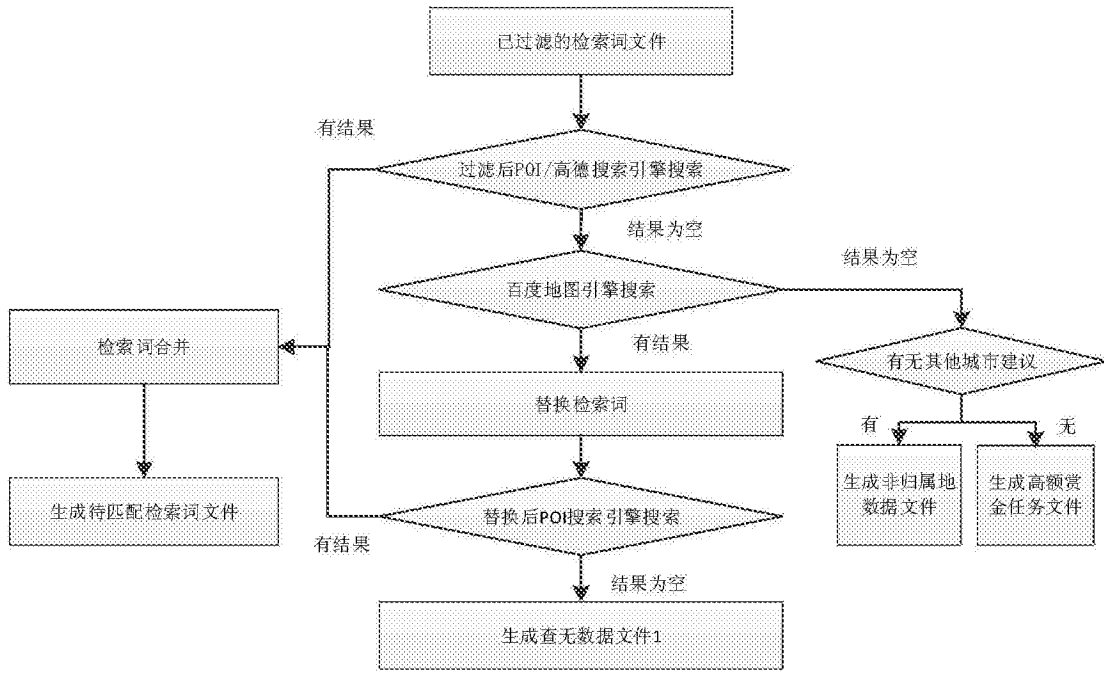


图2

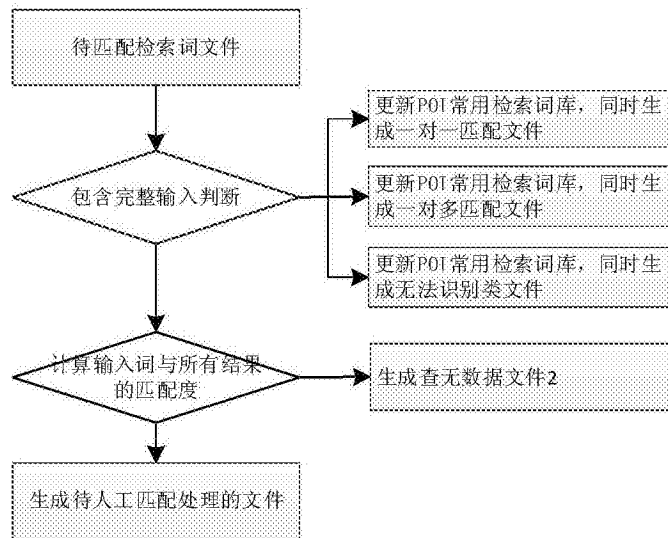


图3

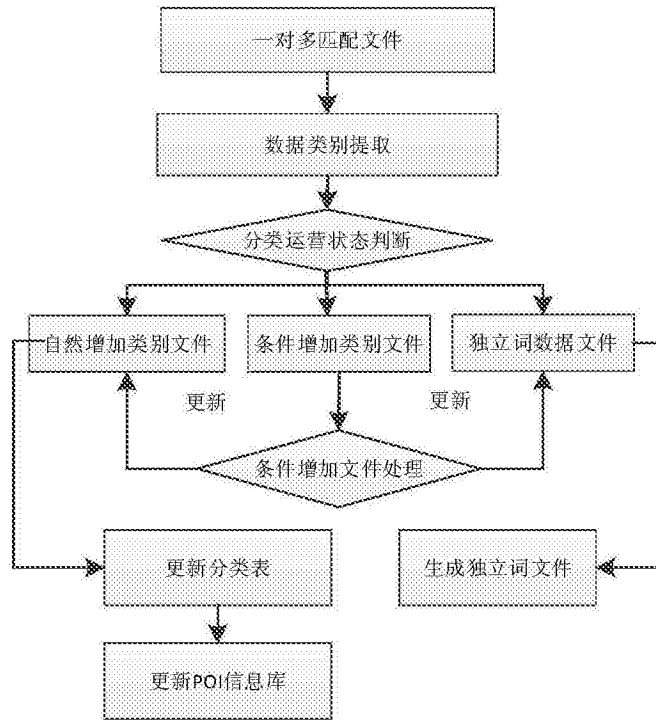


图4

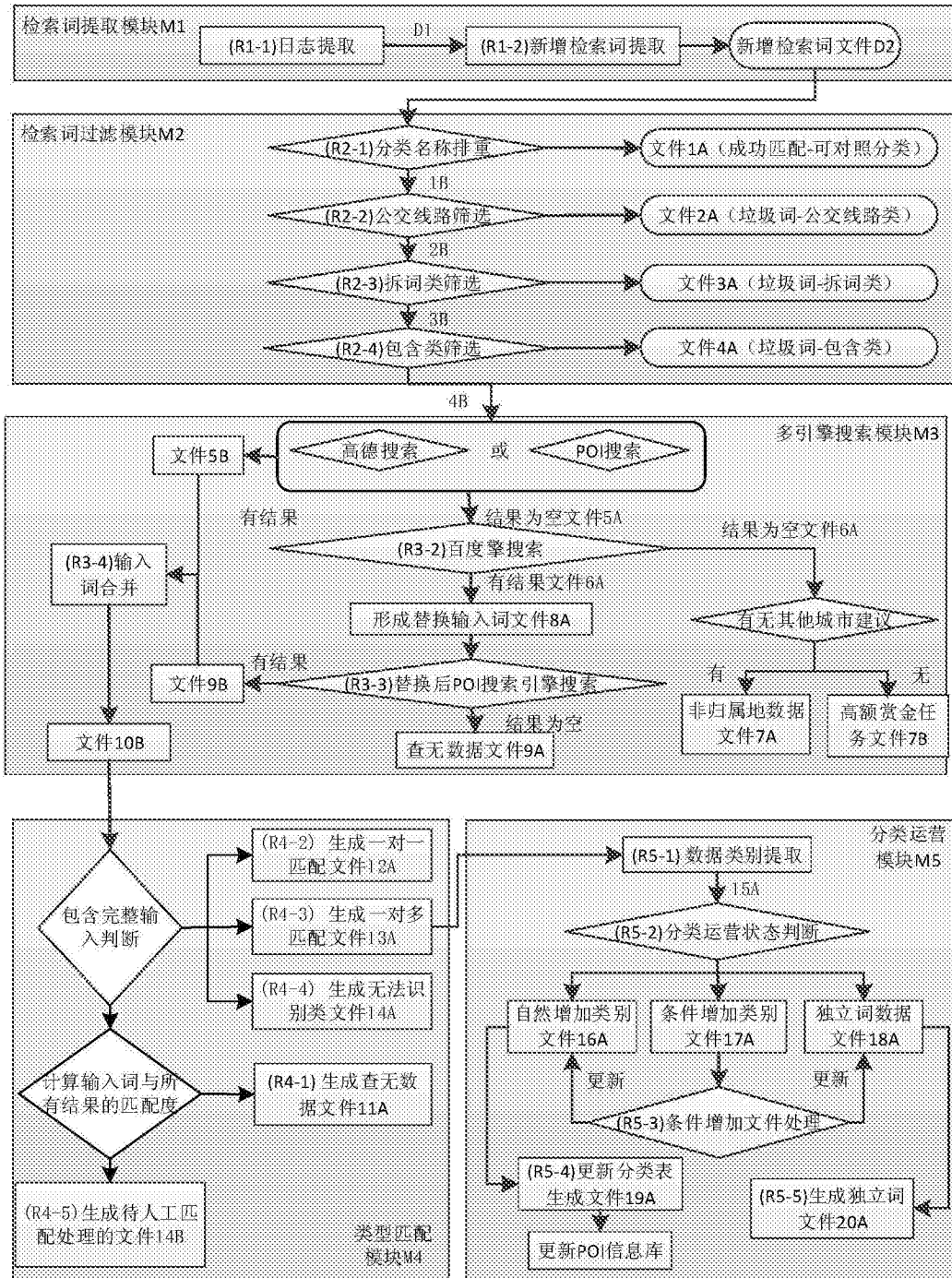


图5

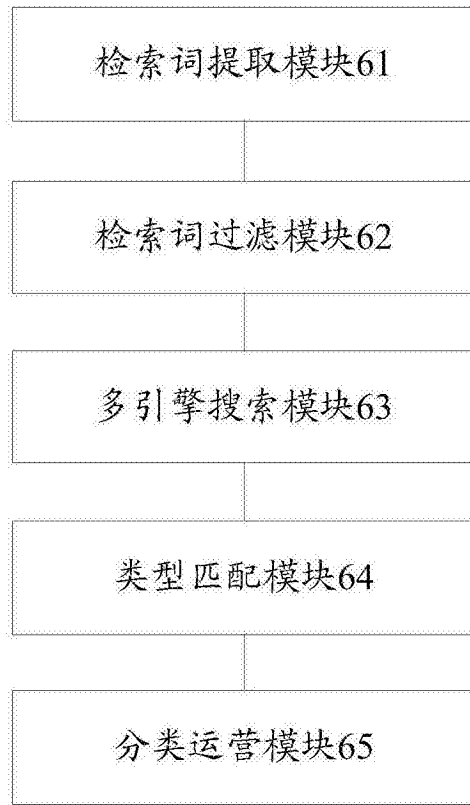


图6