

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织  
国际局

(43) 国际公布日  
2017年7月13日 (13.07.2017)



(10) 国际公布号  
WO 2017/118427 A1

- (51) 国际专利分类号:  
G06F 17/30 (2006.01)
- (21) 国际申请号:  
PCT/CN2017/070504
- (22) 国际申请日:  
2017年1月6日 (06.01.2017)
- (25) 申请语言:  
中文
- (26) 公布语言:  
中文
- (30) 优先权:  
201610008131.3 2016年1月7日 (07.01.2016) CN
- (71) 申请人: 腾讯科技(深圳)有限公司 (TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED) [CN/CN]; 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。
- (72) 发明人: 王忠存 (WANG, Zhongcun); 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。
- (74) 代理人: 北京派特恩知识产权代理有限公司 (CHINA PAT INTELLECTUAL PROPERTY OFFICE); 中国北京市海淀区海淀南路21号中关村知识产权大厦B座2层, Beijing 100080 (CN)。

- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:  
— 包括国际检索报告(条约第21条(3))。

(54) Title: WEBPAGE TRAINING METHOD AND DEVICE, AND SEARCH INTENTION IDENTIFICATION METHOD AND DEVICE

(54) 发明名称: 网页训练的方法和装置、搜索意图识别的方法和装置

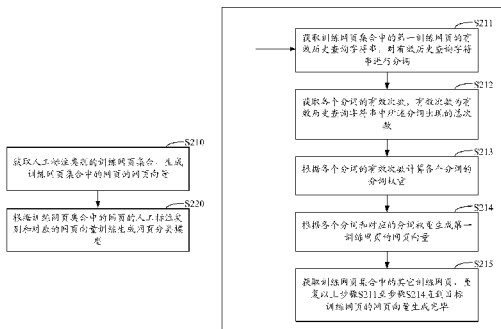


图 3

- S210 Acquiring a training webpage set of a manually annotated class, and generating a webpage vector of a webpage in the training webpage set
- S211 Acquiring an effective history query character string of a first training webpage in the training webpage set, and carrying out word segmentation on the effective history query character string
- S212 Acquiring an effective number of times of each segmented word, wherein the effective number of times is the total number of times the segmented word appears in the effective history query character string
- S213 According to the effective number of times of each segmented word, calculating a segmented word weight of each segmented word
- S214 Generating a webpage vector of the first training webpage according to each segmented word and the corresponding segmented word weight
- S215 Acquiring other training webpages in the training webpage set, and repeating steps S211 to S214 until a webpage vector of a target training webpage is completely generated
- S220 Generating a webpage classification model according to the manually annotated class of the webpage in the training webpage set and the corresponding webpage vector

(57) Abstract: Disclosed are a webpage training method and device, and a search intention identification method and device. The webpage training method comprises: acquiring a training webpage set of a manually annotated class, and generating a webpage vector of a webpage in the training webpage set (S210), specifically comprising: acquiring an effective history query character string of a first training webpage in the training webpage set, and carrying out word segmentation on the effective history query character string (S211); acquiring an effective number of times of each segmented word, wherein the effective number of times is the total number of times the segmented word appears in the effective history query character string (S212); according to the effective number of times of each segmented word, calculating a segmented word weight of each segmented word (S213); and generating a webpage vector of the first training webpage according to each segmented word and the corresponding segmented word weight (S214); and generating a webpage classification model according to the manually annotated class of the webpage in the training webpage set and the corresponding webpage vector (S220). The method and device are low in training cost and are highly efficient. Moreover, after a webpage classification model is generated, class annotation can be automatically carried out on a webpage, such that the accuracy of an identified intention is higher.

(57) 摘要:

[见续页]

WO 2017/118427 A1



---

一种网页训练的方法和装置，及一种搜索意图识别的方法和装置。所述网页训练的方法包括：获取人工标注类别的训练网页集合，生成训练网页集合中的网页的网页向量（S210），具体包括：获取训练网页集合中的第一训练网页的有效历史查询字符串，对有效历史查询字符串进行分词（S211）；获取各个分词的有效次数，有效次数为有效历史查询字符串中分词出现的总次数（S212）；根据各个分词的有效次数计算各个分词的分词权重（S213）；根据各个分词和对应的分词权重生成所述第一训练网页的网页向量（S214）；根据训练网页集合中的网页的人工标注类别和对应的网页向量，生成网页分类模型（S220）。所述方法和装置训练成本低，效率高，且生成网页分类模型后可自动对网页进行标注类别，使识别出的意图正确率更高。

## 网页训练的方法和装置、搜索意图识别的方法和装置

本专利申请要求 2016 年 01 月 07 日提交的中国专利申请号为 201610008131.3，申请人为腾讯科技（深圳）有限公司，发明名称为“网页训练的方法和装置、搜索意图识别的方法和装置”的优先权，该申请的全文以引用的方式并入本申请中。

### 技术领域

本发明涉及互联网技术领域，特别是涉及一种网页训练的方法和装置、搜索意图识别的方法和装置。

### 背景技术

10 随着互联网技术的发展，人们可以通过网络使用搜索引擎检索自己所需要的信息。如当用户在搜索引擎里输入“仙剑奇侠传”时，用户的意图较大可能是搜索电视剧或搜索游戏，搜索引擎需要先判断出用户搜索的意图，才能使返回的搜索结果更接近用户需要的内容。意图识别就是对于任意给定的查询字符串，判断该查询字符串属于的类别。

15 现有的搜索意图识别方法，往往使用人工标注的方法对网页标注类别，在进行意图识别时，需要使用人工标注的网页类别进行识别，需要人工标注每个类别的网页集合，成本太高，而且人工标注的结果往往数量有限，对于点击率少的网页很有可能网页的类别未知，导致意图识别的准确率不高。

### 发明内容

20 基于此，有必要针对上述技术问题，提供一种网页训练的方法和装置、搜索意图识别的方法和装置，提高搜索意图识别的准确率。

一种网页训练的方法，所述方法包括：

获取人工标注类别的训练网页集合，生成所述训练网页集合中的网页的网页向量，具体包括：

5 获取所述训练网页集合中的第一训练网页的有效历史查询字符串，对所述有效历史查询字符串进行分词；

获取各个分词的有效次数，所述有效次数为所述有效历史查询字符串中所述分词出现的总次数；

根据所述各个分词的有效次数计算各个分词的分词权重；

10 根据所述各个分词和对应的分词权重生成所述第一训练网页的网页向量；

根据所述训练网页集合中的网页的人工标注类别和对应的网页向量，生成网页分类模型。

一种网页训练的装置，所述装置包括：

15 网页向量生成模块，配置为获取人工标注类别的训练网页集合，生成所述训练网页集合中的网页的网页向量，所述网页向量生成模块包括：

分词单元，配置为获取所述训练网页集合中的第一训练网页的有效历史查询字符串，对所述有效历史查询字符串进行分词；

20 分词权重计算单元，配置为获取各个分词的有效次数，所述有效次数为所述有效历史查询字符串中所述分词出现的总次数，根据所述各个分词的有效次数计算各个分词的分词权重；

网页向量生成单元，配置为根据所述各个分词和对应的分词权重生成所述第一训练网页的网页向量；

网页分类模型生成模块，配置为根据所述训练网页集合中的网页的人工标注类别和对应的网页向量，生成网页分类模型。

25 上述网页训练的方法和装置，通过获取人工标注类别的训练网页集合，

生成训练网页集合中的网页的网页向量，具体包括：获取训练网页集合中的第一训练网页的有效历史查询字符串，对有效历史查询字符串进行分词，获取各个分词的有效次数，有效次数为有效历史查询字符串中分词出现的总次数，根据各个分词的有效次数计算各个分词的分词权重，根据各个分词和对应的分词权重生成第一训练网页的网页向量，根据训练网页集合中的网页的人工标注类别和对应的网页向量，生成网页分类模型，通过有效历史查询字符串分词后生成的网页向量进行训练，训练成本低，效率高，且生成网页分类模型后可自动对网页进行标注类别，使得中长尾网页也能自动得到类别，从而使得意图识别中网页类别的覆盖率高，识别出的意图正确率更高。

一种搜索意图识别的方法，所述方法包括：

获取待识别的查询字符串，获取所述查询字符串对应的历史网页集合，所述历史网页集合中包括历史通过所述查询字符串点击的各个网页；

获取通过上述的网页训练的方法生成的网页分类模型，根据所述网页分类模型得到所述历史网页集合中的网页的类别；

统计所述历史网页集合中的各个类别中的网页数量，根据所述各个类别中的网页数量和历史网页集合中网页的总数量计算得到所述查询字符串的意图分布；

根据所述意图分布得到所述查询字符串的意图识别结果。

一种搜索意图识别的装置，所述装置包括：

获取模块，配置为获取待识别的查询字符串，获取所述查询字符串对应的历史网页集合，所述历史网页集合中包括历史通过所述查询字符串点击的各个网页；

网页类别获取模块，配置为获取通过上述的网页训练的装置生成的网页分类模型，根据所述网页分类模型得到所述历史网页集合中的网页的类

别;

意图识别模块,配置为统计所述历史网页集合中的各个类别中的网页数量,根据所述各个类别中的网页数量和历史网页集合中网页的总数量计算得到所述查询字符串的意图分布,根据所述意图分布得到所述查询字符串的意图识别结果。

5

上述搜索意图识别的方法和装置,通过获取待识别的查询字符串,获取查询字符串对应的历史网页集合,历史网页集合中包括历史通过查询字符串点击的各个网页,获取通过上述实施例的网页训练的方法生成的网页分类模型,根据网页分类模型得到历史网页集合中的网页的类别,统计所述历史网页集合中的各个类别中的网页数量,根据各个类别中的网页数量和历史网页集合中网页的总数量计算得到查询字符串的意图分布,根据意图分布得到查询字符串的意图识别结果,在意图识别时根据网页分类模型自动对历史网页集合中的网页的类别进行识别,比人工标注的类别的网页覆盖率大,使得中长尾网页也能自动得到类别,识别出的意图正确率更高。

10

## 15 附图说明

图 1 为一个实施例中网页训练的方法、搜索意图识别的方法的应用环境图;

图 2 为一个实施例中图 1 中服务器的内部结构图;

图 3 为一个实施例中网页训练的方法的流程图;

20 图 4 为一个实施例中搜索意图识别的方法的流程图;

图 5 为一个实施例中生成字符串分类模型的流程图;

图 6 为一个实施例中网页训练的装置的结构框图;

图 7 为另一个实施例中网页训练的装置的结构框图;

图 8 为一个实施例中搜索意图识别的装置的结构框图;

25 图 9 为另一个实施例中搜索意图识别的装置的结构框图;

图 10 为再一个实施例中搜索意图识别的装置的结构框图。

### 具体实施方式

图 1 为一个实施例中网页训练的方法、搜索意图识别的方法运行的应用环境图。如图 1 所示，该应用环境包括终端 110、服务器 120，其中终端 5 110 和服务器 120 通过网络进行通信。

终端 110 可为智能手机、平板电脑、笔记本电脑、台式计算机等，但并不局限于此。终端 110 通过网络向服务器 120 发送查询字符串进行搜索，服务器 120 可以响应终端 110 发送的请求。

在一个实施例中，图 1 中的服务器 120 的内部结构如图 2 所示，该服 10 务器 120 包括通过系统总线连接的处理器、存储介质、内存和网络接口。其中，该服务器 120 的存储介质存储有操作系统、数据库和搜索意图识别的装置，其中搜索意图识别的装置中包括有网页训练的装置，数据库用于存储数据，搜索意图识别的装置用于实现一种适用于服务器 120 的搜索意图识别的方法，网页训练的装置用于实现一种适用于服务器 120 的网页训 15 练的方法。该服务器 120 的处理器用于提供计算和控制能力，支撑整个服务器 120 的运行。该服务器 120 的内存为存储介质中的搜索意图识别的装置的运行提供环境。该服务器 120 的网络接口用于与外部的终端 110 通过网络连接通信，比如接收终端 110 发送的搜索请求以及向终端 110 返回数据等。

20 如图 3 所示，在一个实施例中，提供了一种网页训练的方法，以应用于上述应用环境中的服务器来举例说明，包括如下步骤：

步骤 S210，获取人工标注类别的训练网页集合，生成训练网页集合中的网页的网页向量。

具体的，训练网页集合中网页的数量可以根据需要自定义，为了使训 25 练出的网页分类模型更准确，训练网页集合中网页的数量要足够多并且属

于不同的种类，种类的数量也要足够多。训练网页集合中的网页都通过人工标注了类别，如 mp3.baidu.com 被人工标记为音乐类，youku.com 被人工标记为视频类。生成训练网页集合中的网页的网页向量，可以将训练网页集合中的所有网页都生成网页向量，也可以根据预设条件选择部分网页生成对应的网页向量，如选择人工标注的不同的类别，从每个类别中选择预设数量的网页生成对应的网页向量。

生成训练网页集合中的网页的网页向量的步骤具体包括：

步骤 S211，获取训练网页集合中的第一训练网页的有效历史查询字符串，对有效历史查询字符串进行分词。

10 具体的，如果第一训练网页作为第一查询字符串的搜索结果，被用户点击进入，则此第一查询字符串为第一训练网页的有效历史查询字符串，如果第一训练网页作为第二查询字符串的搜索结果，但是没有被用户点击进入，则第二查询字符串不是第一训练网页的有效历史查询字符串。第一训练网页的有效历史查询字符串的数量可根据需要自定义，但是为了使训练结果有效，需要足够多，如获取预设时间段内的第一训练网页的所有有效历史查询字符串，预设时间段可为距离当前时间较接近的时间段。对有效历史查询字符串进行分词，用各个分词表示此查询字符串，如将“周杰伦的歌”分词后得到“周杰伦”、“歌”，分词的目的是更好的表示网页，如果直接用查询字符串 query 表示网页，数据太稀疏，如查询字符串“周杰伦的歌”和“周杰伦的歌曲”为 2 个不同的查询字符串，但是将它进行分词后得到“周杰伦”、“歌”和“周杰伦”、“歌曲”，其中都包括分词“周杰伦”，增加了查询字符串的相似度。

步骤 S212，获取各个分词的有效次数，有效次数为有效历史查询字符串中所述分词出现的总次数。

25 具体的，如有效历史查询字符串进行分词后，有 30 个分词为“周杰伦”，

则“周杰伦”这个分词的有效次数为 30。分词的有效次数越大，表明通过包括此分词的查询字符串进入当前训练网页的次数越多。

步骤 S213，根据各个分词的有效次数计算各个分词的分词权重。

具体的，分词权重的大小与有效次数的大小成正比，具体的分词权重  
5 计算方法可根据需要自定义。

在一个实施例中，根据公式  $W(q_i) = \log(c_i + 1)$  计算分词  $q_i$  的分词权重  $W(q_i)$ ，其中  $i$  为分词的序号， $c_i$  为分词  $q_i$  的有效次数。

具体的， $\log$  函数比较平滑，且满足分词权重  $W(q_i)$  的大小与有效次数  $c_i$  的大小成正比的比列关系，能简单方便的得到各个分词的分词权重。

10 步骤 S214，根据各个分词和对应的分词权重生成第一训练网页的网页向量。

具体的，对于第一训练网页，如果其有效历史查询字符串生成的分词数量为  $m$  个，用  $q_i$  表示各个分词，其中  $1 \leq i \leq m$ ， $W(q_i)$  为分词  $q_i$  对应的分词权重，则第一训练网页的网页向量可表示为  $\{q_1:W(q_1), q_2:W(q_2), \dots \dots$   
15  $q_m:W(q_m)\}$ ，生成的网页向量表示第一训练网页的词袋特征。如对于训练网页 mp3.baidu.com，它的网页向量为 { 周杰伦:5.4, 歌曲:3.6, 蔡依林:3.0, tfboys:10}。可根据网页向量计算不同的网页之间的相似度，如果第一网页与第二网页的相似度满足预设条件，且第一网页的网页类别为第一类，则可以推出第二网页的网页类别也为第一类，如 mp3.baidu.com 的网页向量与  
20 y.qq.com 的网页向量的余弦函数 cosine 相似度大于预设阈值，则根据 mp3.baidu.com 为音乐类推出 y.qq.com 也为音乐类。

步骤 S215，获取训练网页集合中的其它训练网页，重复以上步骤 S211 至步骤 S214 直到目标训练网页的网页向量生成完毕。

具体的，目标训练网页的数量可根据需要自定义，目标训练网页可以  
25 是训练网页集合中通过预设规则筛选出来的训练网页。也可以直接将网页

集合中的全部训练网页作为目标训练网页。

步骤 S220, 根据训练网页集合中的网页的人工标注类别和对应的网页向量, 生成网页分类模型。

具体地, 将所述训练网页集合中的网页的人工标注类别和对应的网页向量代入逻辑回归模型中进行训练, 得到所述网页分类模型。本发明实施例中, 网页分类模型的训练采用逻辑回归法。逻辑回归(Logistic Regression, LR)模型在线性回归的基础上, 套用了—个逻辑函数, 训练出来的网页分类模型正确率高。

具体的, 网页分类模型是一种数学模型, 用于对网页进行分类, 可以采用不同的方法训练分类模型得到不同的网页分类模型。根据需要选择训练方法。

通过监督学习的办法离线训练得到网页分类模型后, 对网页进行在线类别预测时使用训练好的网页分类模型进行类别预测。本实施例中通过有限数量的人工标注类别的网页和生成的网页向量生成网页分类模型, 可通过网页分类模型实现网页类别自动标注。同时, 采用网页向量作为训练数据, 不需要爬取网页上所有的内容并词袋化, 进行训练的数据成本低, 训练效率高。

本实施例中, 通过获取人工标注类别的训练网页集合, 生成训练网页集合中的网页的网页向量, 具体包括: 获取训练网页集合中的第一训练网页的有效历史查询字符串, 对有效历史查询字符串进行分词, 获取各个分词的有效次数, 有效次数为有效历史查询字符串中分词出现的总次数, 根据各个分词的有效次数计算各个分词的分词权重, 根据各个分词和对应的分词权重生成第一训练网页的网页向量, 根据训练网页集合中的网页的人工标注类别和对应的网页向量, 生成网页分类模型, 通过有效历史查询字符串分词后生成的网页向量进行训练, 训练成本低, 效率高, 且生成网页

分类模型后可自动对网页进行标注类别，使得中长尾网页也能自动得到类别，从而使得意图识别中网页类别的覆盖率高，识别出的意图正确率更高。

在一个实施例中，步骤 S220 之前，还包括：获取训练网页集合中的网页的 LDA 特征。

5 具体的，LDA (Latent Dirichlet Allocation, 文档主题生成模型) 用于对文本进行主题聚类，网页的 LDA 特征可通过对网页文本输入 LDA 模型得到。

步骤 S220 为：根据网页的 LDA 特征、人工标注类别和对应的网页向量，生成网页分类模型。

10 具体地，将网页的 LDA 特征、人工标注类别和对应的网页向量代入逻辑回归模型中进行训练，得到所述网页分类模型。本发明实施例中，网页分类模型的训练采用逻辑回归法。逻辑回归(Logistic Regression, LR)模型在线性回归的基础上，套用了一个逻辑函数，训练出来的网页分类模型正确率高。

15 具体的，训练网页分类模型的训练数据中增加了网页的 LDA 特征，LDA 特征反映了网页的主题，使得训练出的网页分类模型更能准确的对网页进行类别标注。

表 1 展示了采用不同的模型和方法进行训练得到的网页分类模型对网页进行分类的准确率和召回率，其只展示了对于小说类别和各个类别综合  
20 进行分类时的准确率和召回率以及对于准确率和召回率综合得到的 F1，其中  $F1 = 2 \times \text{准确率} / (\text{准确率} + \text{召回率})$ 。表格中 LDA 表示文档主题生成模型，LR + LDA 表示同时采用 LR(logistic regression)逻辑回归模型和 LDA 特征，LR + BOW + LDA 表示同时采用 LR 模型、LDA 特征和网页向量 BOW(Bag of Words)词袋特征进行训练。这里，准确率就是检索出来的条目  
25 (比如：文档、网页等)有多少是准确的；召回率就是所有准确的条目有

多少被检索出来了。准确率=提取出的正确信息条数/提取出的信息条数；召回率=提取出的正确信息条数/样本中的信息条数；F1 即为准确率和召回率的调和平均值。

表 1

|                       | 小说类  |     | 各个类别综合 |      |       |
|-----------------------|------|-----|--------|------|-------|
|                       | 准确率  | 召回率 | 准确率    | 召回率  | F1    |
| <b>LDA</b>            | 0.99 | 0.1 | 0.93   | 0.06 | 0.11  |
| <b>LR + LDA</b>       | 0.98 | 0   | 0.90   | 0.03 | 0.005 |
| <b>LR + BOW + LDA</b> | 0.97 | 0.3 | 0.96   | 0.66 | 0.77  |

5 从表格中可以看出基于网页向量采用逻辑回归法训练生成的网页分类模型对网页进行分类时，准确率和召回率大部分得到提高，并且对于准确率和召回率综合得到的 F1 比其它方法要高很多，效果很好。

在一个实施例中，如图 4 所示，提供了一种搜索意图识别的方法，包括：

10 步骤 S310，获取待识别的查询字符串，获取查询字符串对应的历史网页集合，历史网页集合中包括历史通过所述查询字符串点击的各个网页。

具体的，待识别的查询字符串为终端在搜索引擎输入的查询字符串，获取历史搜索中通过此查询字符串点击的各个网页组成的历史网页集合。

15 步骤 S320，获取通过上述任一项实施例的网页训练的方法生成的网页分类模型，根据网页分类模型得到历史网页集合中的网页的类别。

具体的，通过上述实施例中的网页训练的方法生成的网页分类模型自动对历史网页集合中的网页进行分类。如历史网页集合为  $\{url_1, url_2, \dots, url_n\}$ ，其中  $url_i$  ( $1 \leq i \leq n$ ) 代表各个网页，得到各个网页的类别  $url_i \in d_1$ ，

$url_2 \in d_2, \dots \dots url_n \in d_s$ , 其中  $d_1, d_2, \dots, d_s$  表示类别,  $s$  为类别的总个数, 类别集合为  $\{d_1, d_2, \dots \dots d_s\}$ 。

步骤 S330, 统计所述历史网页集合中的各个类别中的网页数量, 根据各个类别中的网页数量和历史网页集合中网页的总数量计算得到查询字符串的意图分布。

具体的, 统计所述历史网页集合中的各个类别中的网页数量, 如类别  $d_1$  中包括  $t$  个网页, 则  $num_{d_1} = t$ , 统计历史网页集合中网页的总个数得到历史网页集合中网页的总数量, 如对于历史网页集合  $\{url_1, url_2, \dots \dots url_n\}$  其总数量  $total_{url} = n$ , 则待识别的查询字符串  $p\text{-query}$  属于类别  $d_1$  的概率  $p(d_1 / p\text{-query}) = num_{d_1} / total_{url}$ , 采用相同的方法计算得到  $p\text{-query}$  属于各个类别的概率  $p(d_i / p\text{-query})$  得到查询字符串的意图分布, 其中  $1 \leq i \leq s$ , 其中概率  $p(d_i / p\text{-query})$  的大小表示查询字符串属于类别  $d_i$  的可能性。

步骤 S340, 根据意图分布得到查询字符串的意图识别结果。

具体的, 可将意图分布中概率最大的类别作为查询字符串的意图识别结果, 或按概率从大到小的顺序取预设数目个类别作为查询字符串的意图识别结果, 或将概率大于预设阈值的类别作为查询字符串的意图识别结果。还可获取发送查询字符串的当前应用所对应的业务, 根据业务信息和意图分布得到查询字符串的意图识别结果, 如发送查询字符串的当前应用的业务信息为音乐业务, 则即使意图分布中概率最大的类别不为音乐, 也可将音乐类别作为意图识别的一个结果。

本实施例中, 通过获取待识别的查询字符串, 获取查询字符串对应的历史网页集合, 历史网页集合中包括历史通过查询字符串点击的各个网页, 获取通过上述实施例的网页训练的方法生成的网页分类模型, 根据网页分类模型得到历史网页集合中的网页的类别, 统计所述历史网页集合中的各

个类别中的网页数量，根据各个类别中的网页数量和历史网页集合中网页的总数量计算得到查询字符串的意图分布，根据意图分布得到查询字符串的意图识别结果，在意图识别时根据网页分类模型自动对历史网页集合中的网页的类别进行识别，比人工标注的类别的网页覆盖率大，使得中长尾网页也能自动得到类别，识别出的意图正确率更高。

在一个实施例中，步骤 S340 之前，还包括：获取字符串分类模型，根据字符串分类模型得到查询字符串的预测类别。

具体的，字符串分类模型是一种数学模型，用于对查询字符串进行分类，可以采用不同的方法训练分类模型得到不同的字符串分类模型，根据需要选择训练方法。通过监督学习的办法离线训练得到字符串分类模型后，对查询字符串进行意图识别时可使用训练好的字符串分类模型进行查询字符串的类别预测。查询字符串的预测类别可以在查询字符串的意图分布不明显时修正查询字符串的意图识别结果，如查询字符串的意图分布中类别多，且各个类别的概率都接近，且比较小，此时只根据查询字符串的意图分布进行识别往往结果不准确。

步骤 S340 为：根据意图分布和预测类别得到查询字符串的意图识别结果。

具体的，可根据意图分布中类别的多少和各个类别对应的概率，决定查询字符串的意图识别结果。如意图分布中类别多且各个类别对应的概率都比较小，可直接将预测类别作为查询字符串的意图识别结果，或将意图分布中概率最大的类别和预测类别组合形成查询字符串的意图识别结果，具体的得到意图识别结果的算法可根据需要自定义。在意图分布得不到的情况下，如查询字符串为一个罕见的字符串，其对应的历史网页集合中的网页数量为 0 或非常小，导致意图分布无法计算或得到的意图分布只有一个类别的概率，且为 100%很可能是错误的，此时也可直接将查询字符串的

预测类别作为查询字符串的意图识别结果。

在一个实施例中，如图所示，获取字符串分类模型的步骤之前，还包括：

5 步骤 S410，获取历史查询字符串对应的意图分布中意图概率最大的类别对应的查询字符串作为类别训练查询字符串，其中意图概率最大的类别包括多个不同类别。

具体的，对大量的历史查询字符串计算得到了意图分布，不同的查询字符串对应的意图分布中意图概率最大的类别可能不同。将意图分布中意图概率最大的类别对应的查询字符串作为类别训练查询字符串且意图概率  
10 最大的类别包括多个不同类别以保证训练数据的有效性。

步骤 S420，对不同类别对应的类别训练查询字符串提取基于词语和/或基于字符的  $n$  元语法特征， $n$  为大于 1 且小于  $M$  的整数， $M$  为当前提取的类别训练查询字符串的词语长度或字符长度。

具体的，如果直接用类别训练查询字符串训练模型，对于比较短的查询字符串，如长度在 4 个词语左右，这种情况下特征过于稀疏，训练模型  
15 不能得到很好的训练结果。提取基于词语和/或基于字符的  $n$  元语法特征，使得特征长度被扩充。对于同一查询字符串，可进行多次提取，每次提取的元数不同，这里，元数代表词语的数目，将每次提取的结果形成一个特征组合。如对于“周杰伦的歌曲”这个类别训练查询字符串，提取基于词  
20 语的 1-3 元语法特征分别得到如下：

1 元语法特征：周杰伦 的 歌曲

2 元语法特征：周杰伦的 的歌曲

3 元语法特征： 周杰伦的歌曲

提取基于字符的 1-3 元语法特征分别得到如下：

25 1 元语法特征:: 周 杰 伦 的 歌 曲

2 元语法特征:: 周杰 杰伦 伦的 的歌 歌曲

3 元语法特征:: 周杰伦 杰伦的 伦的歌 的歌曲

对于一个长度为 3 个词语的查询字符串,提取基于字符的 1-3 元语法特征后其特征长度被扩充为 15 维以上,有效的解决了特征稀疏的问题。同时  
5 因为训练数据足够大,具有很好的扩展性。

步骤 S430,将 n 元语法特征和对应的类别作为训练数据采用分类模型进行训练生成字符串分类模型。

具体地,将 n 元语法特征和对应的类别作为训练数据,代入分类模型中进行训练,得到字符串分类模型。

10 具体的,使用 n 元语法特征和对应的类别作为训练数据,训练数据从类别训练查询字符串进行了扩展,得到的字符串分类模型分类的准确性和覆盖率都能提高。在一个实施例中,可将训练特征映射到固定维度(例如 100 万维)的向量以提高训练的效率和减少无效的训练数据提高训练结果的准确性,或增加查询字符串点击的网页的类别比例特征等增加训练数据的  
15 覆盖率,这里,类别比例特征是指点击的各个网页类别占全部网页的比例,如点击的视频类网页占全部网页的比例。

表 2 展示了采用不同的模型和方法进行训练得到的字符串分类模型对查询字符串进行分类的准确率和召回率,以及对于准确率和召回率综合得到的 F1,其中  $F1 = 2 \times \text{准确率} / (\text{准确率} + \text{召回率})$ 。表格中 NB (Naïve  
20 Bayesian) 表示朴素贝叶斯模型,分词表示提取基于词语的 n 元语法特征,字符特征表示提取基于字符的 n 元语法特征,SVM(support vector machine) 表示支持向量机模型。

表 2

|         | 准确率  | 召回率 | F1   |
|---------|------|-----|------|
| NB + 分词 | 0.50 | 0.1 | 0.16 |

|                    |       |      |       |
|--------------------|-------|------|-------|
| NB + 字符特征          | 0.834 | 0.85 | 0.85  |
| SVM + 分词           | 0.51  | 0.11 | 0.18  |
| SVM + 字符特征 +<br>分词 | 0.887 | 0.88 | 0.883 |

从表格中可以看出采用提取基于字符的 n 元语法特征训练生成的字符串分类模型对查询字符串进行分类时正确率和召回率都很高，且同时采用提取基于字符的 n 元语法特征和基于词语的 n 元语法特征的正确率和召回率更高。使用了本方法的意图识别的整体准确率相比于未使用前可从 54.6% 5 提升至 85%，提升幅度达 60%。

在一个实施例中，如图 6 所示，提供了一种网页训练的装置，包括：

网页向量生成模块 510，配置为获取人工标注类别的训练网页集合，生成训练网页集合中的网页的网页向量，网页向量生成模块 510 包括：

10 分词单元 511，配置为获取训练网页集合中的第一训练网页的有效历史查询字符串，对有效历史查询字符串进行分词。

分词权重计算单元 512，配置为获取各个分词的有效次数，有效次数为有效历史查询字符串中分词出现的总次数，根据各个分词的有效次数计算各个分词的分词权重。

15 网页向量生成单元 513，配置为根据各个分词和对应的分词权重生成第一训练网页的网页向量。

网页分类模型生成模块 520，配置为根据训练网页集合中的网页的人工标注类别和对应的网页向量，生成网页分类模型。

在一个实施例中，如图 7 所示，装置还包括：

20 LDA 特征获取模块 530，配置为获取训练网页集合中的网页的 LDA 特征。

网页分类模型生成模块 520 还配置为根据网页的 LDA 特征、人工标注类别和对应的网页向量，生成网页分类模型。

5 在一个实施例中，网页分类模型生成模块 520 还配置为将所述训练网页集合中的网页的人工标注类别和对应的网页向量代入逻辑回归模型中进行训练，得到所述网页分类模型。

在一个实施例中，分词权重计算单元 511 还配置为根据公式  $W(q_i) = \log(c_i + 1)$  计算分词  $q_i$  的分词权重  $W(q_i)$ ，其中  $i$  为分词的序号， $c_i$  为分词  $q_i$  的有效次数。

10 在一个实施例中，如图 8 所示，提供了一种搜索意图识别的装置，包括：

获取模块 610，配置为获取待识别的查询字符串，获取查询字符串对应的历史网页集合，历史网页集合中包括历史通过查询字符串点击的各个网页。

15 网页类别获取模块 620，配置为获取通过上述任一实施例的网页训练的装置生成的网页分类模型，根据网页分类模型得到历史网页集合中的网页的类别。

20 意图识别模块 630，配置为统计所述历史网页集合中的各个类别中的网页数量，根据各个类别中的网页数量和历史网页集合中网页的总数量计算得到查询字符串的意图分布，根据意图分布得到查询字符串的意图识别结果。

在一个实施例中，如图 9 所示，装置还包括：

预测类别模块 640，配置为获取字符串分类模型，根据字符串分类模型得到查询字符串的预测类别。

25 意图识别模块 630 还配置为根据意图分布和预测类别得到查询字符串

的意图识别结果。

在一个实施例中，如图 10 所示，装置还包括：

字符串分类模型生成模块 650，配置为获取历史查询字符串对应的意图分布中意图概率最大的类别对应的查询字符串作为类别训练查询字符串，  
5 其中意图概率最大的类别包括多个不同类别，对不同类别对应的类别训练查询字符串提取基于词语和/或基于字符的  $n$  元语法特征， $n$  为大于 1 且小于当前提取的查询字符串词语长度或字符长度的整数，将  $n$  元语法特征和对应的类别作为训练数据采用分类模型进行训练生成字符串分类模型。

10 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机程序来指令相关的硬件来完成，所述程序可存储于一计算机可读取存储介质中，如本发明实施例中，该程序可存储于计算机系统的存储介质中，并被该计算机系统至少一个处理器执行，以实现包括如上述各方法的实施例的流程。其中，所述存储介质可为磁碟、光盘、  
15 只读存储记忆体（Read-Only Memory, ROM）或随机存储记忆体（Random Access Memory, RAM）等。

相应地，本发明实施例还提供一种计算机存储介质，其中存储有计算机程序，该计算机程序用于执行本发明实施例的网页训练的方法或者搜索意图识别的方法。

20 以上所述实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

以上所述实施例仅表达了本发明的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对发明专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本发明构思的前提下，还可以做出  
25 若干变形和改进，这些都属于本发明的保护范围。因此，本发明的

保护范围应以所附权利要求为准。

## 权利要求书

1、一种网页训练的方法，所述方法包括：

获取人工标注类别的训练网页集合，生成所述训练网页集合中的网页的网页向量，具体包括：

5 获取所述训练网页集合中的第一训练网页的有效历史查询字符串，对所述有效历史查询字符串进行分词；

获取各个分词的有效次数，所述有效次数为所述有效历史查询字符串中所述分词出现的总次数；

根据所述各个分词的有效次数计算各个分词的分词权重；

10 根据所述各个分词和对应的分词权重生成所述第一训练网页的网页向量；

根据所述训练网页集合中的网页的人工标注类别和对应的网页向量，生成网页分类模型。

2、根据权利要求1所述的方法，其中，所述根据所述训练网页集合中的网页的人工标注类别和对应的网页向量，生成网页分类模型的步骤之前，还包括：

获取所述训练网页集合中的网页的LDA特征；

所述根据所述训练网页集合中的网页的人工标注类别和对应的网页向量，生成网页分类模型的步骤为：

20 根据所述网页的LDA特征、人工标注类别和对应的网页向量，生成网页分类模型。

3、根据权利要求1所述的方法，其中，所述根据所述训练网页集合中的网页的人工标注类别和对应的网页向量，生成网页分类模型，包括：

25 将所述训练网页集合中的网页的人工标注类别和对应的网页向量代入逻辑回归模型中进行训练，得到所述网页分类模型。

4、根据权利要求1所述的方法，其中，所述根据所述各个分词的有效次数计算各个分词的分词权重的步骤包括：

根据公式  $W(q_i) = \log(c_i + 1)$  计算分词  $q_i$  的分词权重  $W(q_i)$ ，其中  $i$  为分词的序号， $c_i$  为分词  $q_i$  的有效次数。

5 5、一种搜索意图识别的方法，所述方法包括：

获取待识别的查询字符串，获取所述查询字符串对应的历史网页集合，所述历史网页集合中包括历史通过所述查询字符串点击的各个网页；

获取通过所述权利要求1至4中任一项所述的网页训练的方法生成的网页分类模型，根据所述网页分类模型得到所述历史网页集合中的网页的类别；

统计所述历史网页集合中的各个类别中的网页数量，根据所述各个类别中的网页数量和历史网页集合中网页的总数量计算得到所述查询字符串的意图分布；

根据所述意图分布得到所述查询字符串的意图识别结果。

15 6、根据权利要求5所述的方法，其中，在所述根据所述意图分布得到所述查询字符串的意图识别结果的步骤之前，还包括：

获取字符串分类模型，根据所述字符串分类模型得到所述查询字符串的预测类别；

所述根据所述意图分布得到所述查询字符串的意图识别结果的步骤为：

根据所述意图分布和预测类别得到所述查询字符串的意图识别结果。

7、根据权利要求6所述的方法，其中，所述获取字符串分类模型的步骤之前，还包括：

25 获取历史查询字符串对应的意图分布中意图概率最大的类别对应的查询字符串作为类别训练查询字符串，其中所述意图概率最大的类别包括多

个不同类别；

对所述不同类别对应的类别训练查询字符串提取基于词语和/或基于字符的  $n$  元语法特征，所述  $n$  为大于 1 且小于当前提取的查询字符串词语长度或字符长度的整数；

5 将所述  $n$  元语法特征和对应的类别作为训练数据采用分类模型进行训练生成所述字符串分类模型。

8、一种网页训练的装置，所述装置包括：

网页向量生成模块，配置为获取人工标注类别的训练网页集合，生成所述训练网页集合中的网页的网页向量，所述网页向量生成模块包括：

10 分词单元，配置为获取所述训练网页集合中的第一训练网页的有效历史查询字符串，对所述有效历史查询字符串进行分词；

分词权重计算单元，配置为获取各个分词的有效次数，所述有效次数为所述有效历史查询字符串中所述分词出现的总次数，根据所述各个分词的有效次数计算各个分词的分词权重；

15 网页向量生成单元，配置为根据所述各个分词和对应的分词权重生成所述第一训练网页的网页向量；

网页分类模型生成模块，配置为根据所述训练网页集合中的网页的人工标注类别和对应的网页向量，生成网页分类模型。

9、根据权利要求 8 所述的装置，其中，所述装置还包括：

20 LDA 特征获取模块，配置为获取所述训练网页集合中的网页的 LDA 特征；

所述网页分类模型生成模块还配置为根据所述网页的 LDA 特征、人工标注类别和对应的网页向量，生成网页分类模型。

25 10、根据权利要求 8 所述的装置，其中，网页分类模型生成模块，还配置为将所述训练网页集合中的网页的人工标注类别和对应的网页向量代

入逻辑回归模型中进行训练，得到所述网页分类模型。

11、根据权利要求 8 所述的装置，其中，所述分词权重计算单元还配置为根据公式  $W(q_i) = \log(c_i + 1)$  计算分词  $q_i$  的分词权重  $W(q_i)$ ，其中  $i$  为分词的序号， $c_i$  为分词  $q_i$  的有效次数。

5 12、一种搜索意图识别的装置，所述装置包括：

获取模块，配置为获取待识别的查询字符串，获取所述查询字符串对应的历史网页集合，所述历史网页集合中包括历史通过所述查询字符串点击的各个网页；

10 网页类别获取模块，配置为获取通过所述权利要求 8 至 11 中任一项所述的网页训练的装置生成的网页分类模型，根据所述网页分类模型得到所述历史网页集合中的网页的类别；

意图识别模块，配置为统计所述历史网页集合中的各个类别中的网页数量，根据所述各个类别中的网页数量和历史网页集合中网页的总数量计算得到所述查询字符串的意图分布，根据所述意图分布得到所述查询字符串的意图识别结果。

13、根据权利要求 12 所述的装置，其中，所述装置还包括：

预测类别模块，配置为获取字符串分类模型，根据所述字符串分类模型得到所述查询字符串的预测类别；

20 所述意图识别模块还配置为根据所述意图分布和预测类别得到所述查询字符串的意图识别结果。

14、根据权利要求 13 所述的装置，其中，所述装置还包括：

25 字符串分类模型生成模块，配置为获取历史查询字符串对应的意图分布中意图概率最大的类别对应的查询字符串作为类别训练查询字符串，其中所述意图概率最大的类别包括多个不同类别，对所述不同类别对应的类别训练查询字符串提取基于词语和/或基于字符的  $n$  元语法特征，所述  $n$  为

大于 1 且小于当前提取的查询字符串词语长度或字符长度的整数, 将所述  $n$  元语法特征和对应的类别作为训练数据采用分类模型进行训练生成所述字符串分类模型。

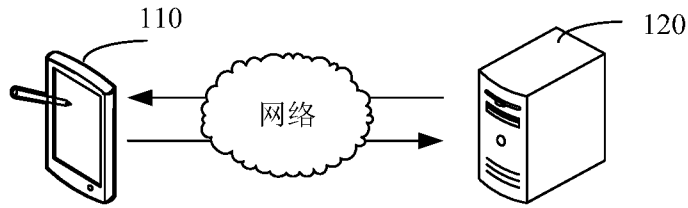


图 1

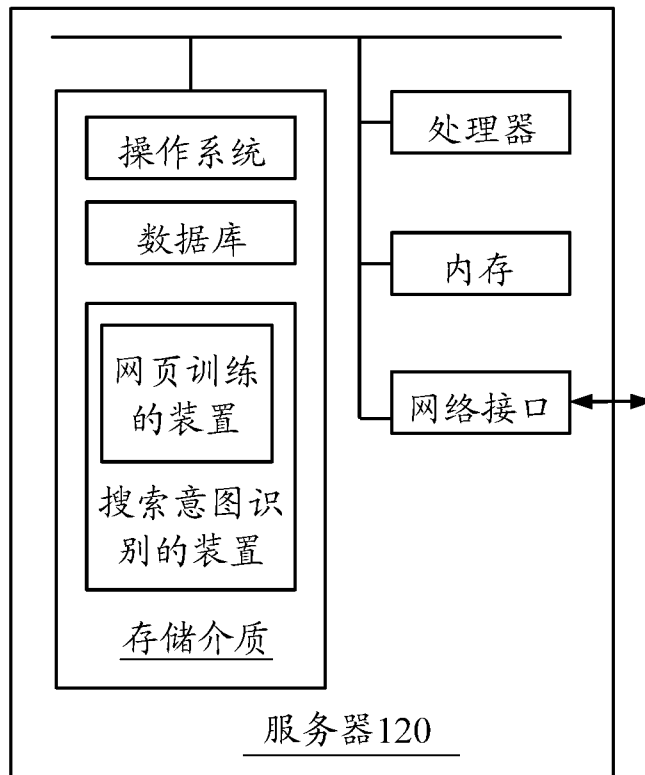


图 2

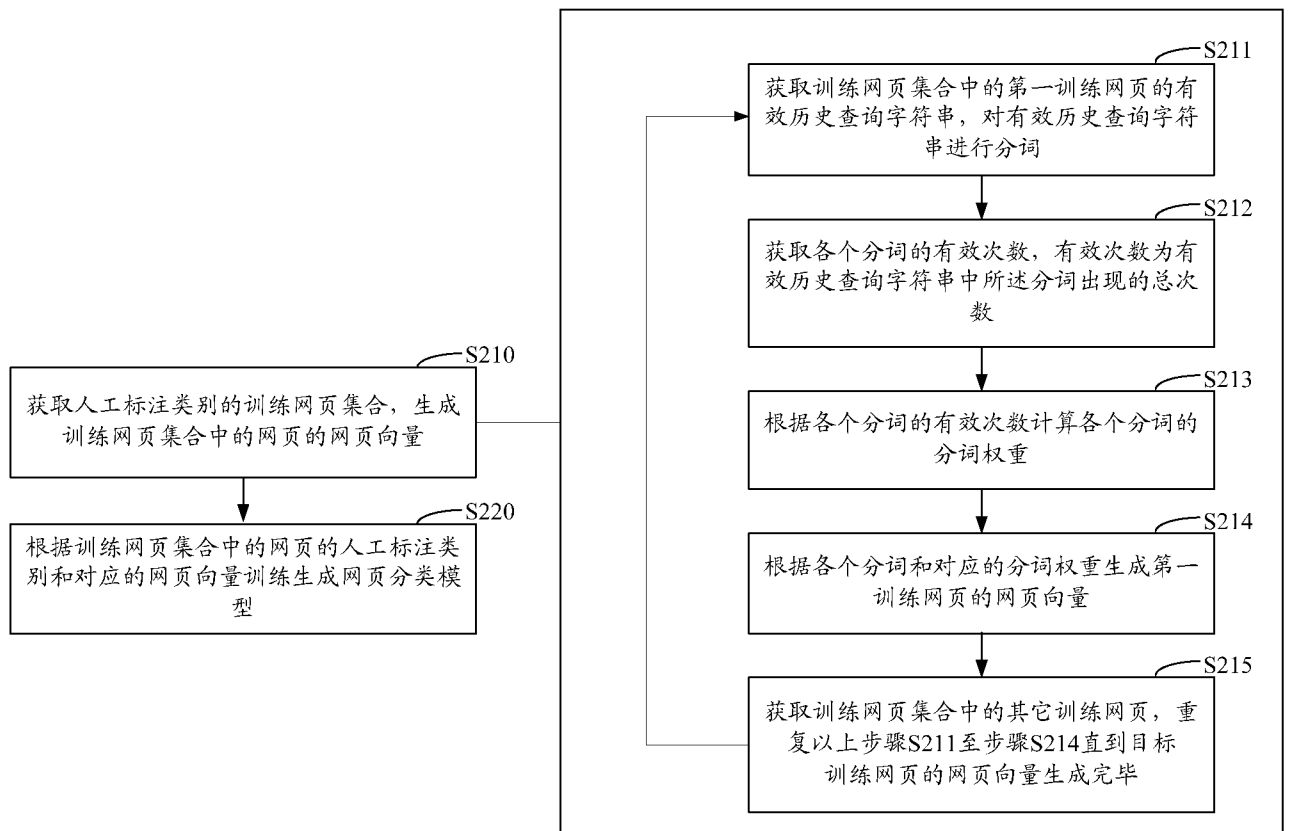


图 3

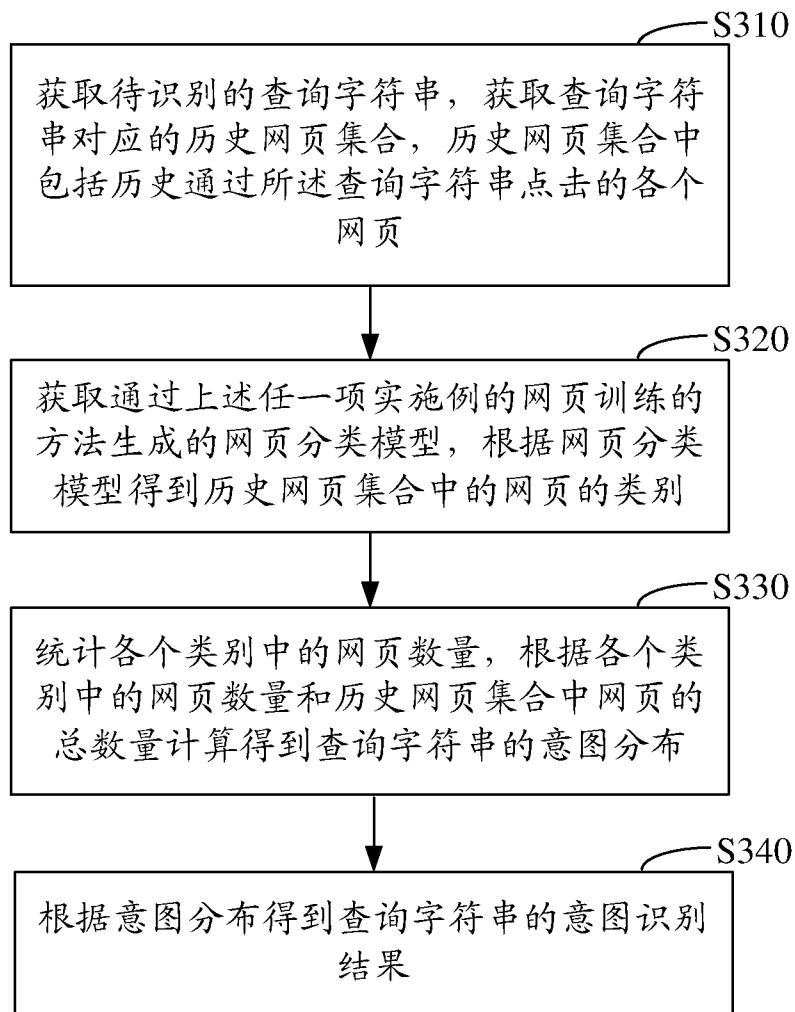


图 4

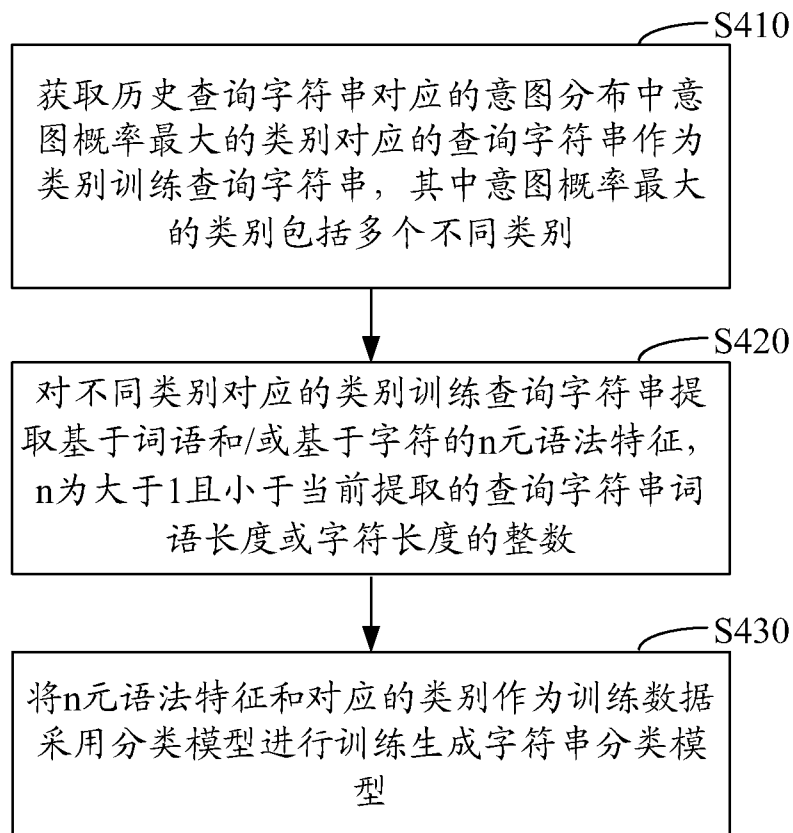


图 5

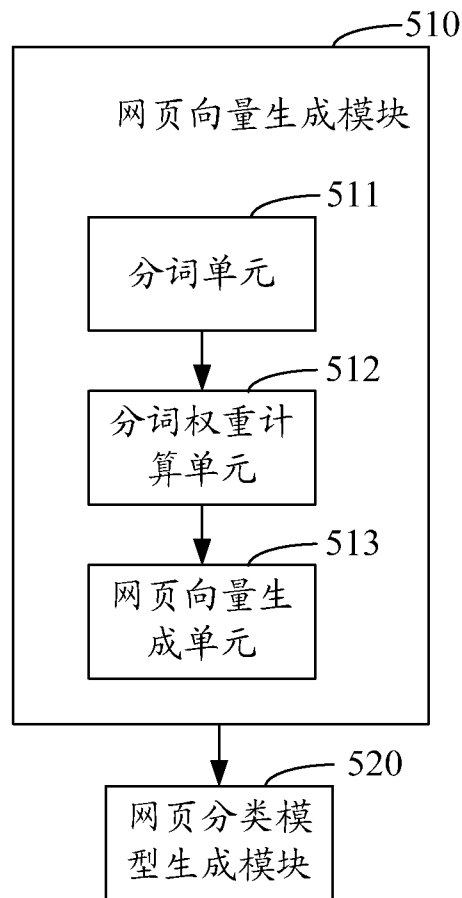


图 6

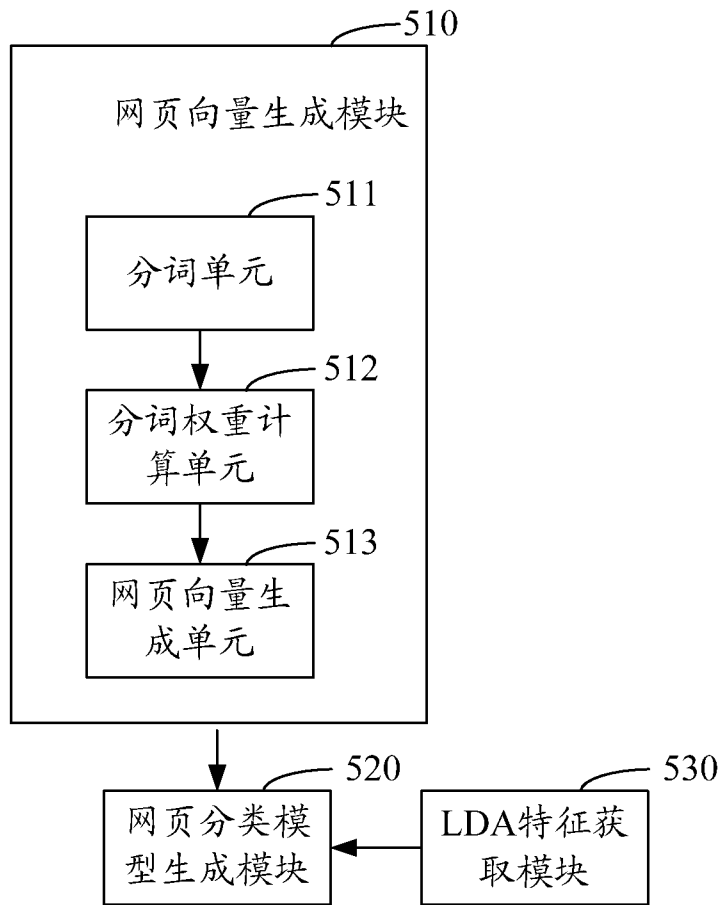


图 7

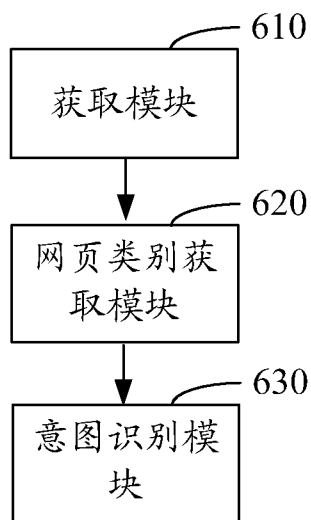


图 8

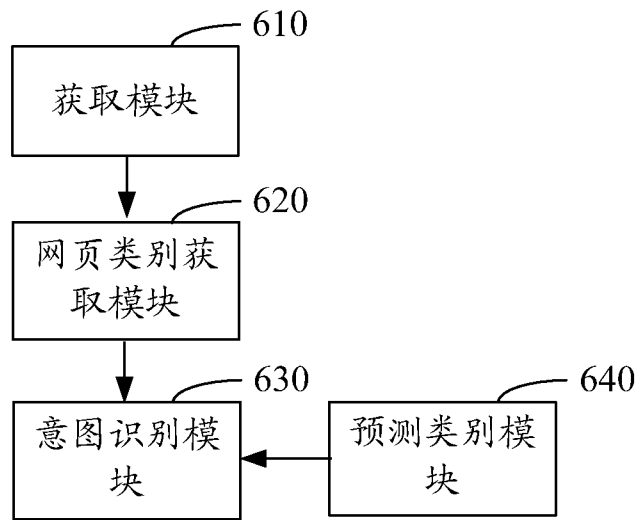


图 9

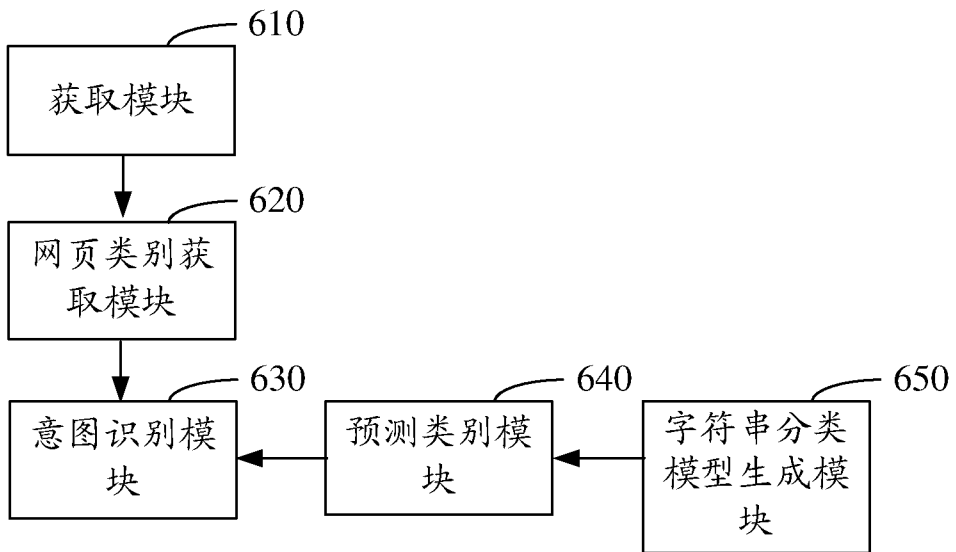


图 10

# INTERNATIONAL SEARCH REPORT

International application No.

**PCT/CN2017/070504**

**A. CLASSIFICATION OF SUBJECT MATTER**

G06F 17/30 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, WPI, EPODOC, CNKI: web page training, web page vector, word segmentation, effective times, word frequency, web page classification model, collection of web pages, intention recognition, web page, training, query history, character string, segment, times, weights, classification model, intention

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages   | Relevant to claim No. |
|-----------|--|-----------------------|
| A         | CN 104834640 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.), 12 August 2015 (12.08.2015), description, paragraphs [0028]-[0035] | 1-14                  |
| A         | CN 101609450 A (NANJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS), 23 December 2009 (23.12.2009), the whole document     | 1-14                  |
| A         | CN 101261629 A (SHANGHAI UNIVERSITY), 10 September 2008 (10.09.2008), the whole document                                   | 1-14                  |
| A         | US 2013144860 A1 (XU, C. et al.), 06 June 2013 (06.06.2013), the whole document  | 1-14                  |
| A         | US 2015248715 A1 (EBAY INC.), 03 September 2015 (03.09.2015), the whole document   | 1-14                  |

Further documents are listed in the continuation of Box C.

See patent family annex.

|   |   |
|---|---|
| <p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p> | <p>“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>“&amp;” document member of the same patent family</p> |
|---|---|

Date of the actual completion of the international search  
21 March 2017 (21.03.2017)

Date of mailing of the international search report  
**17 April 2017 (17.04.2017)**

Name and mailing address of the ISA/CN:  
State Intellectual Property Office of the P. R. China  
No. 6, Xitucheng Road, Jimenqiao  
Haidian District, Beijing 100088, China  
Facsimile No.: (86-10) 62019451

Authorized officer  
**MA, Jintao**  
Telephone No.: (86-10) **62414067**

**INTERNATIONAL SEARCH REPORT**  
Information on patent family members

International application No.

**PCT/CN2017/070504**

| Patent Documents referred<br>in the Report | Publication Date  | Patent Family                                   | Publication Date                                   |
|--|-------------------|---|--|
| CN 104834640 A                             | 12 August 2015    | US 2016314207 A1<br>WO 2015117560 A1            | 27 October 2016<br>13 August 2015                  |
| CN 101609450 A                             | 23 December 2009  | None  |  |
| CN 101261629 A                             | 10 September 2008 | None  |  |
| US 2013144860 A1                           | 06 June 2013      | US 8682882 B2<br>US 8682881 B1<br>US 8380693 B1 | 25 March 2014<br>25 March 2014<br>19 February 2013 |
| US 2015248715 A1                           | 03 September 2015 | None  |  |

国际检索报告

国际申请号

PCT/CN2017/070504

| <p>A. 主题的分类</p> <p>G06F 17/30(2006.01)i</p> <p>按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类</p>  |   |         |     |                   |         |   |   |      |   |  |      |   |   |      |   |   |      |   |   |      |
|---|---|---------|-----|-------------------|---------|---|---|------|---|--|------|---|---|------|---|---|------|---|---|------|
| <p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>G06F</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))</p> <p>CNPAT, WPI, EPODOC, CNKI: 网页训练, 网页向量, 历史查询, 字符串, 分词, 切词, 有效次数, 词频, 权重, 网页分类模型, 网页集合, 意图识别, web page, training, query history, character string, segment, times, weights, classification model, intention</p>  |   |         |     |                   |         |   |   |      |   |  |      |   |   |      |   |   |      |   |   |      |
| <p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>CN 104834640 A (腾讯科技深圳有限公司) 2015年 8月 12日 (2015 - 08 - 12)<br/>说明书第[0028]-[0035]段</td> <td>1-14</td> </tr> <tr> <td>A</td> <td>CN 101609450 A (南京邮电大学) 2009年 12月 23日 (2009 - 12 - 23)<br/>全文</td> <td>1-14</td> </tr> <tr> <td>A</td> <td>CN 101261629 A (上海大学) 2008年 9月 10日 (2008 - 09 - 10)<br/>全文</td> <td>1-14</td> </tr> <tr> <td>A</td> <td>US 2013144860 A1 (XU, CHENG 等) 2013年 6月 6日 (2013 - 06 - 06)<br/>全文</td> <td>1-14</td> </tr> <tr> <td>A</td> <td>US 2015248715 A1 (EBAY INC.) 2015年 9月 3日 (2015 - 09 - 03)<br/>全文</td> <td>1-14</td> </tr> </tbody> </table> |   |         | 类型* | 引用文件, 必要时, 指明相关段落 | 相关的权利要求 | A | CN 104834640 A (腾讯科技深圳有限公司) 2015年 8月 12日 (2015 - 08 - 12)<br>说明书第[0028]-[0035]段 | 1-14 | A | CN 101609450 A (南京邮电大学) 2009年 12月 23日 (2009 - 12 - 23)<br>全文 | 1-14 | A | CN 101261629 A (上海大学) 2008年 9月 10日 (2008 - 09 - 10)<br>全文 | 1-14 | A | US 2013144860 A1 (XU, CHENG 等) 2013年 6月 6日 (2013 - 06 - 06)<br>全文 | 1-14 | A | US 2015248715 A1 (EBAY INC.) 2015年 9月 3日 (2015 - 09 - 03)<br>全文 | 1-14 |
| 类型*   | 引用文件, 必要时, 指明相关段落   | 相关的权利要求 |     |                   |         |   |   |      |   |  |      |   |   |      |   |   |      |   |   |      |
| A   | CN 104834640 A (腾讯科技深圳有限公司) 2015年 8月 12日 (2015 - 08 - 12)<br>说明书第[0028]-[0035]段 | 1-14    |     |                   |         |   |   |      |   |  |      |   |   |      |   |   |      |   |   |      |
| A   | CN 101609450 A (南京邮电大学) 2009年 12月 23日 (2009 - 12 - 23)<br>全文                    | 1-14    |     |                   |         |   |   |      |   |  |      |   |   |      |   |   |      |   |   |      |
| A   | CN 101261629 A (上海大学) 2008年 9月 10日 (2008 - 09 - 10)<br>全文                       | 1-14    |     |                   |         |   |   |      |   |  |      |   |   |      |   |   |      |   |   |      |
| A   | US 2013144860 A1 (XU, CHENG 等) 2013年 6月 6日 (2013 - 06 - 06)<br>全文               | 1-14    |     |                   |         |   |   |      |   |  |      |   |   |      |   |   |      |   |   |      |
| A   | US 2015248715 A1 (EBAY INC.) 2015年 9月 3日 (2015 - 09 - 03)<br>全文                 | 1-14    |     |                   |         |   |   |      |   |  |      |   |   |      |   |   |      |   |   |      |
| <p><input type="checkbox"/> 其余文件在C栏的续页中列出。</p> <p><input checked="" type="checkbox"/> 见同族专利附件。</p>  |   |         |     |                   |         |   |   |      |   |  |      |   |   |      |   |   |      |   |   |      |
| <p>* 引用文件的具体类型:</p> <p>“A” 认为不特别相关的表示了现有技术一般状态的文件</p> <p>“E” 在国际申请日的当天或之后公布的在先申请或专利</p> <p>“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)</p> <p>“O” 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>“P” 公布日先于国际申请日但迟于所要求的优先权日的文件</p> <p>“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件</p> <p>“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性</p> <p>“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性</p> <p>“&amp;” 同族专利的文件</p>   |   |         |     |                   |         |   |   |      |   |  |      |   |   |      |   |   |      |   |   |      |
| <p>国际检索实际完成的日期</p> <p>2017年 3月 21日</p>  | <p>国际检索报告邮寄日期</p> <p>2017年 4月 17日</p>   |         |     |                   |         |   |   |      |   |  |      |   |   |      |   |   |      |   |   |      |
| <p>ISA/CN的名称和邮寄地址</p> <p>中华人民共和国国家知识产权局(ISA/CN)<br/>中国北京市海淀区蓟门桥西土城路6号 100088</p> <p>传真号 (86-10)62019451</p>   | <p>受权官员</p> <p>马晋涛</p> <p>电话号码 (86-10)62414067</p>                              |         |     |                   |         |   |   |      |   |  |      |   |   |      |   |   |      |   |   |      |

国际检索报告  
关于同族专利的信息

国际申请号

PCT/CN2017/070504

| 检索报告引用的专利文件 |            |    | 公布日<br>(年/月/日) | 同族专利 |            |    | 公布日<br>(年/月/日) |
|-------------|------------|----|----------------|------|------------|----|----------------|
| CN          | 104834640  | A  | 2015年 8月 12日   | US   | 2016314207 | A1 | 2016年 10月 27日  |
|             |            |    |                | WO   | 2015117560 | A1 | 2015年 8月 13日   |
| CN          | 101609450  | A  | 2009年 12月 23日  | 无    |            |    |                |
| CN          | 101261629  | A  | 2008年 9月 10日   | 无    |            |    |                |
| US          | 2013144860 | A1 | 2013年 6月 6日    | US   | 8682882    | B2 | 2014年 3月 25日   |
|             |            |    |                | US   | 8682881    | B1 | 2014年 3月 25日   |
|             |            |    |                | US   | 8380693    | B1 | 2013年 2月 19日   |
| US          | 2015248715 | A1 | 2015年 9月 3日    | 无    |            |    |                |

表 PCT/ISA/210 (同族专利附件) (2009年7月)