

(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
2018年5月31日 (31.05.2018)



(10) 国际公布号
WO 2018/095281 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2017/111604
- (22) 国际申请日: 2017年11月17日 (17.11.2017)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201611055619.8 2016年11月25日 (25.11.2016) CN
- (71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 开曼群岛大开曼资本大厦一座四层847号邮箱, Grand Cayman (KY)。
- (72) 发明人; 及
- (71) 申请人 (仅对US): 孙清清 (SUN, Qingqing) [CN/CN]; 中国浙江省杭州市余杭区文一西路969号3号楼5楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。
- (74) 代理人: 北京国昊天诚知识产权代理有限公司 (CO-HORIZON INTELLECTUAL PROPERTY INC.); 中国北京市朝阳区小关北里甲2号渔阳置业大厦B座605, Beijing 100029 (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, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK,

(54) Title: NAME MATCHING METHOD AND APPARATUS

(54) 发明名称: 一种名称匹配方法及装置

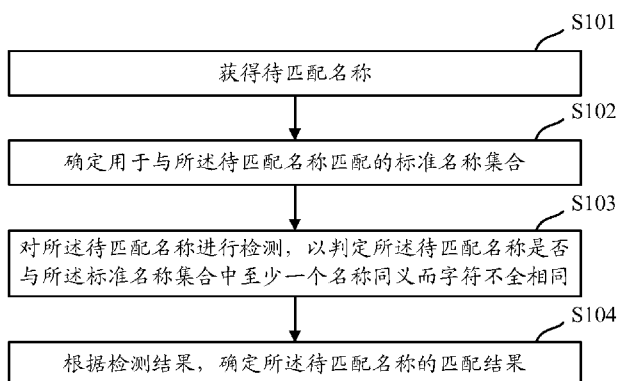


图 1

- S101 OBTAINING A NAME TO BE MATCHED
S102 DETERMINING A STANDARD NAME SET FOR MATCHING THE NAME TO BE MATCHED
S103 DETECTING THE NAME TO BE MATCHED, SO AS TO DETERMINE WHETHER THE NAME TO BE MATCHED IS SYNONYMOUS WITH AT LEAST ONE NAME IN THE STANDARD NAME SET AND IS NOT COMPLETELY IDENTICAL TO CHARACTERS THEREIN
S104 DETERMINING, ACCORDING TO A DETECTION RESULT, A MATCHING RESULT OF THE NAME TO BE MATCHED

(57) Abstract: A name matching method and apparatus. The method comprises: obtaining a name to be matched (S101); determining a standard name set for matching the name to be matched (S102); detecting the name to be matched, so as to determine whether the name to be matched is synonymous with at least one name in the standard name set and is not completely identical to characters therein (S103); and determining, according to a detection result, a matching result of the name to be matched (S104). By means of the method and apparatus, the matching precision can be improved, and the false alarm rate of a risk control system can be reduced.



WO 2018/095281 A1

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))。

(57) 摘要: 一种名称匹配方法及装置。所述方法包括: 获得待匹配名称 (S101); 确定用于与所述待匹配名称匹配的标准名称集合 (S102); 对所述待匹配名称进行检测, 以判定所述待匹配名称是否与所述标准名称集合中至少一个名称同义而字符不全相同 (S103); 根据检测结果, 确定所述待匹配名称的匹配结果 (S104)。利用本方法及装置, 可以提高匹配精确度, 降低风险控制系统的虚警率。

一种名称匹配方法及装置

技术领域

本申请涉及计算机软件技术领域，尤其涉及一种名称匹配方法及装置。

5 背景技术

人名匹配是风险控制领域非常重要的技术。例如，风险控制系统在黑名单中记录已确定的各非法用户的人名，之后在进行风险控制时，针对当前进行业务的每个用户，通过扫描将每个用户的人名与黑名单中的各人名进行匹配，若匹配成功，则可以认为该用户是非法用户并拒绝其业务，以防止风险。

10 人名匹配具体可以分为人名精确匹配和人名模糊匹配，相比而言，人名模糊匹配更具技术难点，因为难以把握合适的模糊程度。

在现有技术中，通常采用字符串匹配算法进行人名模糊匹配，则字符串匹配度阈值决定了模糊程度，而字符串匹配度阈值全凭经验设定，为了减少遗漏，字符串匹配度阈值往往设定得较低，从而容易导致匹配精确度较低，风险控制
15 系统的虚警率偏高。

发明内容

本申请实施例提供一种名称匹配方法及装置，用以解决如下技术问题：现有技术中采用字符串匹配算法进行人名模糊匹配的匹配精确度较低，系统虚警
20 率偏高。

为解决上述技术问题，本申请实施例是这样实现的：

本申请实施例提供的一种名称匹配方法，包括：

获得待匹配名称；

确定用于与所述待匹配名称匹配的标准名称集合；

25 对所述待匹配名称进行检测，以判定所述待匹配名称是否与所述标准名称集合中至少一个名称同义而字符不全相同；

根据检测结果，确定所述待匹配名称的匹配结果。

本申请实施例提供的一种名称匹配装置，包括：

获得模块，获得待匹配名称；

确定模块，确定用于与所述待匹配名称匹配的标准名称集合；

检测模块，对所述待匹配名称进行检测，以判定所述待匹配名称是否与所述标准名称集合中至少一个名称同义而字符不全相同；

5 匹配模块，根据检测结果，确定所述待匹配名称的匹配结果。

本申请实施例采用的上述至少一个技术方案能够达到以下有益效果：所述名称可以包括人名，实际应用中的待匹配人名可能由于数据录入时的时效性、不确定性以及人名的多变性，导致待匹配人名与实际人名有出入，这也是进行模糊匹配的原因，本申请的方案针对该原因，对待匹配人名进行检测，以判定
10 所述待匹配名称是否与所述标准名称集合中至少一个名称同义而字符不全相同，并根据检测结果确定人名匹配结果，相比于现有技术中仅局限于通过凭经验设定的字符串匹配度阈值控制模糊程度，更有利于提高对模糊程度的控制准确程度，可以提高匹配精确度，降低风险控制系统的虚警率，因此，可以部分或全部地减少现有技术中的问题。

15

附图说明

为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请中记载的一些实施例，对于本领域普通技术人员来讲，
20 在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

图 1 为本申请实施例提供的一种名称匹配方法的流程示意图；

图 2 为本申请实施例提供的一种实际应用场景下，名称匹配方法中初筛的一种具体实施方案的流程简图；

图 3 为本申请实施例提供的一种实际应用场景下，名称匹配方法中集成算法模糊匹配的一种具体实施方案的流程简图；
25

图 4 为本申请实施例提供的一种实际应用场景下，名称匹配方法的一种具体实施方案的流程简图；

图 5 为本申请实施例提供的对应于图 1 的一种名称匹配装置的结构示意图。

具体实施方式

本申请实施例提供一种名称匹配方法及装置。

为了使本技术领域的人员更好地理解本申请中的技术方案，下面将结合本
5 申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，
显然，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。基
于本申请中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获
得的所有其他实施例，都应当属于本申请保护的范围。

如前所述，实际应用中的待匹配人名可能由于数据录入时的时效性、不确
10 定性以及人名的多变性，导致待匹配人名与实际人名有出入（主要指人名的“形
（也即，字符）”发生变化）。为了便于理解，以英文人名为例，英文人名常见
变化类型及实例，如下表 1 所示。

表 1

英文人名常见变化类型	人名变化类型的实例
错拼	Jaxson、Jakson、Jaxon 本来是想表示同一个人名；SMYTH、SMITH 本来是想表示同一个人名；拼写错误
语音拼写差异	Michel、Michal、Miguel 本来是想表示同一个人名；发音类似导致的错拼
昵称	Mike, Mick, Mikey 表示同一人名的昵称
称谓词	Dr.、Mr. 等是对人的称谓词，和人名一起出现时不影响该人名的义
同义	Mohamed、Mohammed 和 Mohammad 出现在人名中时同义
缩写	人名 ROB 是 ROBBIN 缩写词，二者同义；人名 BOB 是 BOBBY 的缩写词，二者同义；
多国语言差异	俄语 BORISOV 与 BORISOVNA 不同义，西班牙语 JUANLTA 和 JUAN 同义

15 其中，若只采用字符串匹配的方法匹配人名，可能对于个别字母“错拼”
的变化类型尚能够尝试匹配（实际上也是靠误打误撞，没有过硬的根据），但是
对于其他变化类型，匹配精确度很低。

而本申请的方案针对上述其他变化类型，可以进行诸如缩写检测、称谓词检测、多语言检测、别称检测等特定检测，从而可以比较全面地考虑人名匹配时同义而字符不全相同（也即，同义而不同形）情况，进而可以提高匹配精确度。需要说明的是，“不同形”可以指由于错拼而导致“有误”的不同形，但在下面的实施例中更主要是指由于上述其他变化类型所导致的“合理无误”的不同形。

本申请的方案不仅适用于人名匹配，也适用于人名以外的其他名称的匹配，比如，地名、物名等。

下面对本申请的方案进行详细说明。

图 1 为本申请实施例提供的一种名称匹配方法的流程示意图。执行该流程的程序可搭载的设备包括但不限于：个人计算机、大中型计算机、计算机集群、手机、平板电脑、智能可穿戴设备、车机，等等。该流程通常可以在风险控制领域使用，具体由风险控制系统或相关系统执行。

图 1 中的流程可以包括以下步骤：

S101：获得待匹配名称。

在本申请实施例中，对待匹配名称所属的具体语言并不做限定，可以是英文、俄文、西班牙文等，也可以是中文等。为了便于描述，下面的实施例主要以待匹配名称所属的语言是英文为例进行说明。

S102：确定用于与所述待匹配名称匹配的标准名称集合。

在本申请实施例中，标准名称集合可以从一个更大的名称集合中筛选出来的子集，也可以直接是所述更大的名称集合本身，为了便于描述，可以将此处的筛选称为“初筛”。比如，在背景技术的场景下，所述更大的名称集合可以是风险控制系统所掌握的黑名单。

在前一种情况下，子集可以是只包含与待匹配名称相似的名称的集合，此处的“相似”可以不是那么严格，因为后续还有一系列进一步地确定相似性的操作，相当于后续还要进行“细筛”。通过初筛可以快速缩小匹配范围，进而可以减少后续细筛的工作量，提高细筛的针对性，有利于提高本申请的方案的效率。

S103：对所述待匹配名称进行检测，以判定所述待匹配名称是否与所述标

准名称集合中至少一个名称同义而字符不全相同。

在本申请实施例中，所要检测判定的“同义而字符不全相同”主要是由表1中的一种或多种变化类型导致的，所述检测具体可以包括：缩写检测、称谓检测、多语言检测、别称检测等检测中的至少一种，后面会一一详细说明。

5 在本申请实施例中，当所述检测包括多种时，可以按照一定的顺序依次执行多种检测，若在检测过程中已经可以确定待匹配名称的匹配结果，也可以不执行剩余的检测。当然，为了提高执行效率，也可以并行地执行多种检测，再汇总检测结果。

S104：根据检测结果，确定所述待匹配名称的匹配结果。

10 在本申请实施例中，通过执行步骤 S103，若判定待匹配名称与标准名称集合中至少一个名称同义而字符不全相同，则根据所述情况可以直接确定待匹配名称的匹配结果，在这种情况下，步骤 S103 中的检测过程实际上也就是对待匹配名称的整个匹配过程。

15 而若判定待匹配名称不满足：与所述标准名称集合中至少一个名称同义而字符不全相同；则可以进一步地采用其他匹配方式对待匹配名称进行匹配，以确定待匹配名称的匹配结果。

20 通过图1的方法，所述名称可以包括人名，实际应用中的待匹配人名可能由于数据录入时的时效性、不确定性以及人名的多变性，导致待匹配人名与实际人名有出入，这也是进行模糊匹配的原因，本申请的方案针对该原因，对待匹配人名进行检测，以判定所述待匹配名称是否与所述标准名称集合中至少一个名称同义而字符不全相同，并根据检测结果确定人名匹配结果，相比于现有技术中仅局限于通过凭经验设定的字符串匹配度阈值控制模糊程度，更有利于提高对模糊程度的控制准确程度，可以提高匹配精确度，降低风险控制系统的虚警率，因此，可以部分或全部地减少现有技术中的问题。

25 基于图1的方法，本申请实施例还提供了该方法的一些具体实施方案，以及扩展方案，下面进行说明。

在本申请实施例中，不同的待匹配名称的复杂程度可能不同，包含的信息也可能不同，对于某些信息量过于小或者信息特征过于简单的待匹配名称，即使对其进行匹配，所获得的匹配结果的价值也难以达到预期。比如，过于简单

通用的名称如英文人名“Jim”、“Jimmy”、“David”、“John”、“Mike”等，即使匹配成功也难以特定到一个人。

为了防止这种情况导致用于名称匹配的处理资源白白浪费，可以在获得待匹配名称后先对其进行一次过滤，以判决是否继续匹配。另外，若待匹配名称包含在白名单上，则也存在类似问题，也可以采用这种方式进行处理。

具体地，对于步骤 S101 和 S102，所述获得待匹配名称后，确定用于与所述待匹配名称匹配的标准名称集合前，还可以执行：获得预定的无需匹配名称集合；确定所述待匹配名称是否包含在所述无需匹配名单集合中；若是，则继续执行后续步骤；否则，可以不对所述待匹配名称进行匹配。

在本申请实施例中，以步骤 S102 的一种具体实施方案（对应于上述“初筛”的情况）为例展开说明。对于步骤 S102，所述确定用于与所述待匹配名称匹配的标准名称集合，具体可以包括：确定可用于与所述待匹配名称匹配的第一名称集合；通过将所述待匹配名称包含的各词与所述第一名称集合中名称包含的各词进行相似性匹配，确定用于与所述待匹配名称匹配的标准名称集合。

如何进行相似性匹配也有多种具体实施方案，可以进行对待匹配名称进行分词匹配，也可以对待匹配进行全文本匹配，等等。

以分词匹配为例，所述通过将所述待匹配名称包含的各词与所述第一名称集合中名称包含的各词进行相似性匹配，确定用于与所述待匹配名称匹配的标准名称集合，具体可以包括：获得所述第一名称集合包含的各名称的索引，所述名称的索引为该名称包含的任一词；以及对所述待匹配名称进行分词，获得所述待匹配名称包含的各词；分别将对所述待匹配名称包含的各词，与各所述索引进行相似性匹配，获得由匹配成功的各索引所索引的名称构成的所述第一名称集合的子集；根据各所述子集，确定用于与所述待匹配名称匹配的标准名称集合。

上例中的索引是预先建立的，基于索引进行分词匹配的优点是：可以有效加快在匹配过程中对集合中所需的名称的获取速度。若不基于索引，仍可以实现分词匹配（比如，在存储集合的数据表中直接用 Select 语句查询所需的名称以用于分词匹配等），只是效率上可能受到影响。

进一步地，所述分别将对所述待匹配名称包含的各词，与各所述索引进行

相似性匹配，具体可以包括：采用字符串匹配算法，分别将对所述待匹配名称包含的各词，与各所述索引进行相似性匹配；所述字符串匹配算法可以包括一种或多种，比如，前缀树匹配算法、字典树匹配算法、字符串相似度匹配算法、发音相似匹配算法等。这里，采用字符串匹配算法只是一种较优的方式，也可以采用文本匹配算法等其他可以用于实现相似性匹配的算法。

所述根据各所述子集确定用于与所述待匹配名称匹配的标准名称集合，也可以有多种具体实施方案。比如，若采用 N 种字符串匹配算法，对于分词得到的 M 个词中的每个词，分别采用每种字符串匹配算法将该词与各索引进行匹配，对应得到 N 个子集，然后取 N 个子集的并集；共得到 M 个并集，取在 M 个并集中出现总次数超过设定阈值的各名称，确定为标准名称集合。又比如，得到 N 个子集，然后取 N 个子集的交集确定为标准名称集合。

上面对“初筛”过程进行了说明，为了便于理解，本申请实施例还提供了一种实际应用场景下，名称匹配方法中初筛的一种具体实施方案的流程简图，如图 2 所示。

在该实际应用场景下，假定第一名称集合为某英文人名名单，预先设定了该名单中各名称的索引（通过索引还可以索引得到对应名称以外的相关信息），索引具体为其对应的名称中的词，各索引包含在以索引对应的词为主键建立的索引表中，待匹配名称为英文人名“Kit Wai Jackson Wong”，以空格作为分隔符对该人名进行分词，分词结果如下表 2 所示。

表 2

待匹配人名	Kit Wai Jackson Wong			
分词后	词 1: Kit	词 2: Wai	词 3: Jackson	词 4: Wong

在图 2 中，M=4，分词结果具体为 {Kit, Wai, Jackson, Wong}，采用前缀树匹配算法、字典树匹配算法、字符串相似度匹配算法（比如 Simstring 算法等）、发音相似匹配算法（比如 metaphone 算法等）分别对分词得到的每个词与各索引分别进行匹配，输出单个词通过这四种匹配算法得到的索引对应的名称集合，匹配结果 1、2、3、4；

将每个词的匹配结果取并集，得到每个词的综合匹配结果；

取至少包含在两个综合匹配结果中的名称构成集合，作为生成的初筛结果。

在本申请实施例中，确定标准名称集合后，可以对待匹配名称进行检测。不过，在检测之前，还可以进行一项或多项预处理，预处理有利于提高后续检测结果的可靠性。所述预处理可以包括对齐处理、大小写统一处理、简繁体处理等。

以对齐处理为例。对于步骤 S103，所述对所述待匹配名称进行检测前，还可以执行：根据所述标准名称集合中名称包含的各词与所述待匹配名称包含的各词的相似度，对所述标准名称集合中名称包含的各词与所述待匹配名称包含的各词进行对齐处理。

在实际应用中，基于相似度进行对齐处理的具体原理可以是：基于相似度最大化原则，即将相似度最大的词所在的名称位置对齐。

例如，假定待匹配名称为“Kate Lee Smith”，标准名称集合中的某名称为“Smith Catherine Lee”对齐结果如下表 3 所示。

表 3

待匹配名称	Kate	Lee	Smith
标准名称集合中对齐处理后的某名称	Catherine	Lee	Smith

进一步地，在进行了对齐处理的情况下，对于步骤 S103，所述对所述待匹配名称进行检测，具体可以包括：根据对齐后的所述标准名称集合，对所述待匹配名称进行检测，以判定所述待匹配名称是否与所述标准名称集合中至少一个名称同义而字符不全相同。

在本申请实施例中，步骤 S103 中的检测是提高人名模糊匹配的匹配准确度的关键部分。具体地，对于步骤 S103，所述对所述待匹配名称进行检测，具体包括：

对所述待匹配名称进行缩写检测、称谓词检测、多语言检测、别称检测中的至少一种，以判定所述待匹配名称是否与所述标准名称集合中至少一个名称同义而字符不全相同。下面分别对这几种检测进行说明。

在本申请实施例中，缩写检测中最常见的一种是首字母缩写检测，除此之外还有部分词省略缩写检测等。在一种具体实施方式下，所述对所述待匹配名称进行缩写检测，具体可以包括：获得预定的缩写对照组合数据，每个所述缩写对照组合反映至少一个词与其缩写词的缩写对应关系；根据所述缩写对照组合数据，检测所述待匹配名称包含的词与所述标准名称集合中的名称包含的词是否具有所述缩写对应关系；根据检测结果，判定所述待匹配名称是否与所述标准名称集合中至少一个名称同义而字符不全相同。

例如，若预先确定了英文人名“Ben Williams”可缩写为“B. Williams”，则可以将“Ben Williams”与“B. Williams”作为一个缩写对照组合，若检测出待匹配名称与标准名称集合中的某个名称即为该缩写对照组合，则可以确定存在同义而字符不全相同的情况。

在本申请实施例中，所述对所述待匹配名称进行称谓词检测，具体可以包括：获得预定的称谓词数据；根据所述称谓词数据，检测所述待匹配名称是否包含所述称谓词，所述称谓词被认为不影响包含该称谓词的名称之义；根据检测结果，以判定所述待匹配名称是否与所述标准名称集合中至少一个名称同义而字符不全相同。

就人名而言，称谓词一般是通用尊称（比如，Mr. 或 Miss. 等）、头衔（Dr.、Prof. 等）等附加于至少部分原始人名的词。在人名匹配的环境下，称谓词并不会影响其所对应的所述至少部分原始人名之义。因此，若确定待匹配名称中包含有称谓词，且除了称谓词的其他部分与标准名称集合中的某个名称可以匹配，则可以判定待匹配名称与标准名称集合中至少一个名称同义而字符不全相同。

在本申请实施例中，所述对所述待匹配名称进行别称检测，具体可以包括：获得预定的别称对照组合数据，每个所述缩写对照组合反映至少一个词与其别称的别称对应关系；根据所述别称对照组合数据，检测所述待匹配名称包含的词与所述标准名称集合中的名称包含的词是否具有所述别称对应关系；根据检测结果，判定所述待匹配名称是否与所述标准名称集合中至少一个名称同义而字符不全相同。

在实际应用中，别称可以包括其对应的名称的昵称（比如，表 1 中的 Mick

是 Mikey 的昵称等)、或其对应的名称在不同领域下的同义名称。对于后者,不同领域具体可以是不同地域(比如,不同国家、不同省份等)、不同语言(比如,不同国家的语言、不同民族的语言等)、不同的行业等。

相应地,对所述待匹配名称进行昵称检测,和/或,对所述待匹配名称进行不同领域下的同义名称检测。

别称也不会影响其对应的名称之义,因此,若确定待匹配名称为标准名称集合中的某个名称的别称,则可以判定待匹配名称与标准名称集合中至少一个名称同义而字符不全相同。

需要说明的是,本申请对上述的对照组合数据、称谓词数据的具体存储形式并不做限定,常用的一种方式是在对应的数据表中,需要使用时从数据库中读取。

在本申请实施例中,所述对所述待匹配名称进行多语言检测,具体可以包括:确定所述待匹配名称对应的语言;获得所述语言和/或其他语言的拼写变形同义规则和/或拼写变形异义规则;根据所述拼写变形同义规则和/或拼写变形异义规则,对所述待匹配名称进行检测,以判定所述待匹配名称是否与标准名称集合中至少一个名称同义而字符不全相同。

多语言检测主要是针对诸如以下这种情况:比如,英文“Pooh”在德文中拼写为“puh”,两者在人名中出现是属于同义。

需要说明的是,在前面进行的各项检测中,可能会使用到用于计算字符串匹配度的算法。

在本申请实施例中,所述根据检测结果,确定所述待匹配名称的匹配结果,具体可以包括:若判定所述待匹配名称与标准名称集合中至少一个名称同义而字符不全相同,将所述至少一个名称确定为所述待匹配名称的匹配结果;否则(也即,通过上述检测,未能匹配成功时),通过采用一种或多种相似度算法对所述待匹配名称与标准名称集合中的名称进行匹配,确定所述待匹配名称的匹配结果。

在本申请实施例中,所述多种相似度算法可以是基于不同维度的,从而有利于提高匹配结果的可靠性。根据这样的思路,相似度算法具体可以是用于计算文本匹配度的算法(比如,n-gram算法等)、用于计算语音匹配度的算法(比

如, Phonex 算法等)、用于计算字符串匹配度的算法(比如, Jaro-Winkler 算法等)等。

在采用了多种相似度算法的情况下,可以通过对各相似度算法对应的匹配结果的综合衡量,得到综合匹配结果,本申请对具体衡量方式并不做限定,常用的一种方式,是加权相加。

例如,采用 n-gram 算法,算法输入为待匹配名称包含的各词与其对齐位置的各词,算法输出为对齐的每一对词的文本匹配度,记作 F1;

采用 Jaro-Winkler 算法,算法输入为待匹配名称包含的各词与其对齐位置的各词,算法输出为对齐的每一对词的字符串匹配度,记作 F2;

10 采用 Phonex 算法,算法输入为待匹配名称包含的各词与其对齐位置的各词,算法输出为对齐的每一对词的语音匹配度,记作 F3;

通过对文本匹配度、字符串匹配度和语音匹配度加权相加得到对齐的每一对词的综合匹配度 F,如下公式所示:

$$F=w1*F1+ w2*F2+ w3*F3, 其中, w1+w2+w3=1;$$

15 根据各对词的综合匹配度 F,通过取平均值得到待匹配名称与标准名称集中的名称之间的匹配结果,比如,对于表 3 中的名称对,得到的匹配结果如下表 4 所示。

表 4

入检名字	Kate	Lee	Smith
名单名字	Catherine	Lee	Smith
综合匹配度	0.872	1	1
匹配结果	$(0.872+1+1)/3=0.957$		

20 上面对初筛后的检测以及匹配过程进行了说明,该过程中可能使用到多种匹配相关的算法,本申请的方案在具体实施时,可以对这些可能使用的算法进行集成,则该过程即是一种集成算法模糊匹配的过程。

在本申请实施例中,在进行集成算法模糊匹配后,还可以进一步地进行一些后置规则过滤,比如,将匹配结果中的匹配度映射为文字描述信息,或者根据具体场景对匹配度进行适当补偿或者消减等。

根据上面的说明,更直观地,本申请实施例还提供了一种实际应用场景下,名称匹配方法中集成算法模糊匹配的一种具体实施方案的流程简图,如图3所示。在图3中,各项检测和各种方式的匹配度计算的执行顺序只是示例,并非对本申请的限定。

5 在图3中,若通过上述任一项检测对待匹配名称匹配成功,则可以直接输出匹配结果,反之,则可以通过一种或多种方式的匹配度计算(比如,文本匹配度计算、语音匹配度计算、字符串匹配度计算等),确定并输出待匹配名称的匹配结果。

当然,也可以完整地执行各项检测和各种方式的匹配度计算,再综合考虑
10 各项检测结果和各种方式的匹配度计算结果,确定待匹配字符串的匹配结果。

进一步地,本申请实施例还提供了一种实际应用场景下,名称匹配方法的一种具体实施方案的流程简图,如图4所示。在图4中,前置规则判决可以包括:判决待匹配名称是否未包含在无需匹配名单集合中等,扫描名单索引即为上述的标准名称集合中名称的索引。

15 图3、图4中的步骤在上面都进行了详细说明,在此不赘述。

上面对本申请实施例提供的名称匹配方法进行了说明,基于同样的发明思路,本申请实施例还提供了对应的装置,如图5所示。

图5为本申请实施例提供的对应于图1的一种名称匹配装置的结构示意图,虚线表示可选的模块,所述装置包括:

20 获得模块501,获得待匹配名称;

确定模块502,确定用于与所述待匹配名称匹配的标准名称集合;

检测模块503,对所述待匹配名称进行检测,以判定所述待匹配名称是否与
所述标准名称集合中至少一个名称同义而字符不全相同;

匹配模块504,根据检测结果,确定所述待匹配名称的匹配结果。

25 可选地,所述确定模块502确定用于与所述待匹配名称匹配的标准名称集合前,获得预定的无需匹配名称集合,确定所述待匹配名称未包含在所述无需匹配名单集合中。

可选地,所述确定模块502确定用于与所述待匹配名称匹配的标准名称集合,具体包括:

所述确定模块 502 获得可用于与所述待匹配名称匹配的第一名称集合，通过将所述待匹配名称包含的各词与所述第一名称集合中名称包含的各词进行相似性匹配，确定用于与所述待匹配名称匹配的标准名称集合。

5 可选地，所述确定模块 502 通过将所述待匹配名称包含的各词与所述第一名称集合中名称包含的各词进行相似性匹配，确定用于与所述待匹配名称匹配的标准名称集合，具体包括：

所述确定模块 502 获得所述第一名称集合包含的各名称的索引，所述名称的索引为该名称包含的任一词，以及对所述待匹配名称进行分词，获得所述待匹配名称包含的各词，分别将对所述待匹配名称包含的各词，与各所述索引进行相似性匹配，获得由匹配成功的各索引所索引的名称构成的所述第一名称集合的子集，根据各所述子集，确定用于与所述待匹配名称匹配的标准名称集合。

10 可选地，所述确定模块 502 分别将对所述待匹配名称包含的各词，与各所述索引进行相似性匹配，具体包括：

所述确定模块 502 采用字符串匹配算法，分别将对所述待匹配名称包含的各词，与各所述索引进行相似性匹配，所述字符串匹配算法包括以下至少一种：前缀树匹配算法、字典树匹配算法、字符串相似度匹配算法、发音相似匹配算法。

可选地，所述装置还包括：

20 对齐模块 505，在所述检测模块 503 对所述待匹配名称进行检测前，根据所述标准名称集合中名称包含的各词与所述待匹配名称包含的各词的相似度，对所述标准名称集合中名称包含的各词与所述待匹配名称包含的各词进行对齐处理；

所述检测模块 503 对所述待匹配名称进行检测，具体包括：

25 所述检测模块 503 根据对齐后的所述标准名称集合，对所述待匹配名称进行检测，以判定所述待匹配名称是否与所述标准名称集合中至少一个名称同义而字符不全相同。

可选地，所述检测模块 503 对所述待匹配名称进行检测，具体包括：

所述检测模块 503 对所述待匹配名称进行缩写检测、称谓词检测、多语言检测、别称检测中的至少一种，以判定所述待匹配名称是否与所述标准名称集

合中至少一个名称同义而字符不全相同。

可选地，所述检测模块 503 对所述待匹配名称进行缩写检测，具体包括：

所述检测模块获得预定的缩写对照组合数据，每个所述缩写对照组合反映至少一个词与其缩写词的缩写对应关系，根据所述缩写对照组合数据，检测所述待匹配名称包含的词与所述标准名称集中的名称包含的词是否具有所述缩写对应关系，根据检测结果，判定所述待匹配名称是否与所述标准名称集中至少一个名称同义而字符不全相同。

可选地，所述检测模块 503 对所述待匹配名称进行称谓词检测，具体包括：

所述检测模块 503 获得预定的称谓词数据，根据所述称谓词数据，检测所述待匹配名称是否包含所述称谓词，所述称谓词被认为不影响包含该称谓词的名称之义，根据检测结果，判定所述待匹配名称是否与所述标准名称集中至少一个名称同义而字符不全相同。

可选地，所述别称包括其对应的名称的昵称、或其对应的名称在不同领域下的同义名称；

所述检测模块 503 对所述待匹配名称进行别称检测，具体包括：

所述检测模块 503 对所述待匹配名称进行昵称检测，和/或，对所述待匹配名称进行不同领域下的同义名称检测。

可选地，所述检测模块 503 对所述待匹配名称进行多语言检测，具体包括：

所述检测模块 503 确定所述待匹配名称对应的语言，获得所述语言的拼写变形同义规则和/或拼写变形异义规则，根据所述拼写变形同义规则和/或拼写变形异义规则，对所述待匹配名称进行检测，以判定所述待匹配名称是否与所述标准名称集中至少一个名称同义而字符不全相同。

可选地，所述检测模块 503 对所述待匹配名称进行别称检测，具体包括：

所述检测模块 503 获得预定的别称对照组合数据，每个所述缩写对照组合反映至少一个词与其别称的别称对应关系，根据所述别称对照组合数据，检测所述待匹配名称包含的词与所述标准名称集中的名称包含的词是否具有所述别称对应关系，根据检测结果，判定所述待匹配名称是否与所述标准名称集中至少一个名称同义而字符不全相同。

可选地，所述匹配模块 504 根据检测结果，确定所述待匹配名称的匹配结

果，具体包括：

所述匹配模块 504 若所述检测模块 503 判定所述待匹配名称与所述标准名称集合中至少一个名称同义而字符不全相同，将所述至少一个名称确定为所述待匹配名称的匹配结果，否则，通过采用一种或多种相似度算法对所述待匹配名称与所述标准名称集合中的名称进行匹配，确定所述待匹配名称的匹配结果。

可选地，所述相似度算法包括以下至少一种：用于计算文本匹配度的算法、用于计算语音匹配度的算法、用于计算字符串匹配度的算法。

可选地，所述名称为人名。

本申请实施例提供的装置与方法是一一对应的，因此，装置也具有与其对应的方法类似的有益技术效果，由于上面已经对方法的有益技术效果进行了详细说明，因此，这里不再赘述对应装置的有益技术效果。

另外，在本申请对上述装置与方法的具体应用场景并不做限定，除了背景技术中提到的风险控制领域（比如，反洗钱、用户认证等具体领域）以外，其他任何可能需要用到名称匹配技术的领域，本申请的方案均可能适用。

在 20 世纪 90 年代，对于一个技术的改进可以很明显地区分是硬件上的改进（例如，对二极管、晶体管、开关等电路结构的改进）还是软件上的改进（对于方法流程的改进）。然而，随着技术的发展，当今的很多方法流程的改进已经可以视为硬件电路结构的直接改进。设计人员几乎都通过将改进的方法流程编程到硬件电路中来得到相应的硬件电路结构。因此，不能说一个方法流程的改进就不能用硬件实体模块来实现。例如，可编程逻辑器件（Programmable Logic Device, PLD）（例如现场可编程门阵列（Field Programmable Gate Array, FPGA））就是这样一种集成电路，其逻辑功能由用户对器件编程来确定。由设计人员自行编程来把一个数字系统“集成”在一片 PLD 上，而不需要请芯片制造厂商来设计和制作专用的集成电路芯片。而且，如今，取代手工地制作集成电路芯片，这种编程也多半改用“逻辑编译器（logic compiler）”软件来实现，它与程序开发撰写时所用的软件编译器相类似，而要编译之前的原始代码也得用特定的编程语言来撰写，此称之为硬件描述语言（Hardware Description Language, HDL），而 HDL 也并非仅有一种，而是有许多种，如 ABEL（Advanced

Boolean Expression Language)、AHDL (Altera Hardware Description Language)、Confluence、CUPL (Cornell University Programming Language)、HDCal、JHDL (Java Hardware Description Language)、Lava、Lola、MyHDL、PALASM、RHDL (Ruby Hardware Description Language) 等, 目前最普遍使用
5 的是 VHDL (Very-High-Speed Integrated Circuit Hardware Description Language) 与 Verilog。本领域技术人员也应该清楚, 只需要将方法流程用上述几种硬件描述语言稍作逻辑编程并编程到集成电路中, 就可以很容易得到实现该逻辑方法流程的硬件电路。

控制器可以按任何适当的方式实现, 例如, 控制器可以采取例如微处理器
10 或处理器以及存储可由该(微)处理器执行的计算机可读程序代码(例如软件或固件)的计算机可读介质、逻辑门、开关、专用集成电路(Application Specific Integrated Circuit, ASIC)、可编程逻辑控制器和嵌入微控制器的形式, 控制器的例子包括但不限于以下微控制器: ARC 625D、Atmel AT91SAM、Microchip PIC18F26K20 以及 Silicone Labs C8051F320, 存储器控制器还可
15 以被实现为存储器的控制逻辑的一部分。本领域技术人员也知道, 除了以纯计算机可读程序代码方式实现控制器以外, 完全可以通过将方法步骤进行逻辑编程来使得控制器以逻辑门、开关、专用集成电路、可编程逻辑控制器和嵌入微控制器等的形式来实现相同功能。因此这种控制器可以被认为是一种硬件部
20 件, 而对其内包括的用于实现各种功能的装置也可以视为硬件部件内的结构。或者甚至, 可以将用于实现各种功能的装置视为既可以是实现方法的软件模块又可以是硬件部件内的结构。

上述实施例阐明的系统、装置、模块或单元, 具体可以由计算机芯片或实体实现, 或者由具有某种功能的产品来实现。一种典型的实现设备为计算机。具体的, 计算机例如可以为个人计算机、膝上型计算机、蜂窝电话、相机电话、
25 智能电话、个人数字助理、媒体播放器、导航设备、电子邮件设备、游戏控制台、平板计算机、可穿戴设备或者这些设备中的任何设备的组合。

为了描述的方便, 描述以上装置时以功能分为各种单元分别描述。当然, 在实施本申请时可以把各单元的功能在同一个或多个软件和/或硬件中实现。

本领域内的技术人员应明白, 本发明的实施例可提供为方法、系统、或计

计算机程序产品。因此，本发明可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本发明可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质（包括但不限于磁盘存储器、CD-ROM、光学存储器等）上实施的计算机程序产品的形式。

5 本发明是参照根据本发明实施例的方法、设备（系统）、和计算机程序产品的流程图和 / 或方框图来描述的。应理解可由计算机程序指令实现流程图和 / 或方框图中的每一流程和 / 或方框、以及流程图和 / 或方框图中的流程和 / 或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机
10 或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的装置。

这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制品，该指令装置实现在流程图一个流程或多个
15 流程和 / 或方框图一个方框或多个方框中指定的功能。

这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个
20 流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的步骤。

在一个典型的配置中，计算设备包括一个或多个处理器（CPU）、输入/输出接口、网络接口和内存。

内存可能包括计算机可读介质中的非永久性存储器，随机存取存储器（RAM）和/或非易失性内存等形式，如只读存储器（ROM）或闪存（flash RAM）。内存是计算机可读介质的示例。

25 计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括，但不限于相变内存（PRAM）、静态随机存取存储器（SRAM）、动态随机存取存储器（DRAM）、其他类型的随机存取存储器（RAM）、只读存储器（ROM）、电可擦除可编程只读存储器（EEPROM）、快

闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带, 磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质, 可用于存储可以被计算设备访问的信息。按照本文中的界定, 计算机可读介质不包括暂存电脑可读媒体(transitory media), 如调制
5 的数据信号和载波。

还需要说明的是, 术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含, 从而使得包括一系列要素的过程、方法、商品或者设备不仅包括那些要素, 而且还包括没有明确列出的其他要素, 或者是还包括为这种过程、方法、商品或者设备所固有的要素。在没有更多限制的情况下, 由语句“包
10 括一个……”限定的要素, 并不排除在包括所述要素的过程、方法、商品或者设备中还存在另外的相同要素。

本领域技术人员应明白, 本申请的实施例可提供为方法、系统或计算机程序产品。因此, 本申请可采用完全硬件实施例、完全软件实施例或结合软件和硬件方面的实施例的形式。而且, 本申请可采用在一个或多个其中包含有计算机
15 可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

本申请可以在由计算机执行的计算机可执行指令的一般上下文中描述, 例如程序模块。一般地, 程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构等等。也可以在分布式计算环境中实践本
20 申请, 在这些分布式计算环境中, 由通过通信网络而被连接的远程处理设备来执行任务。在分布式计算环境中, 程序模块可以位于包括存储设备在内的本地和远程计算机存储介质中。

本说明书中的各个实施例均采用递进的方式描述, 各个实施例之间相同相似的部分互相参见即可, 每个实施例重点说明的都是与其他实施例的不同之处。尤其, 对于系统实施例而言, 由于其基本相似于方法实施例, 所以描述的比较
25 简单, 相关之处参见方法实施例的部分说明即可。

以上所述仅为本申请的实施例而已, 并不用于限制本申请。对于本领域技术人员来说, 本申请可以有各种更改和变化。凡在本申请的精神和原理之内所作的任何修改、等同替换、改进等, 均应包含在本申请的权利要求范围之内。

权利要求书

1、一种名称匹配方法，其特征在于，包括：

获得待匹配名称；

确定用于与所述待匹配名称匹配的标准名称集合；

5 对所述待匹配名称进行检测，以判定所述待匹配名称是否与所述标准名称集合中至少一个名称同义而字符不全相同；

根据检测结果，确定所述待匹配名称的匹配结果。

2、如权利要求 1 所述的方法，其特征在于，所述确定用于与所述待匹配名称匹配的标准名称集合前，所述方法还包括：

10 获得预定的无需匹配名称集合；

确定所述待匹配名称未包含在所述无需匹配名称集合中。

3、如权利要求 1 所述的方法，其特征在于，所述确定用于与所述待匹配名称匹配的标准名称集合，具体包括：

确定可用于与所述待匹配名称匹配的第一名称集合；

15 通过将所述待匹配名称包含的各词与所述第一名称集合中名称包含的各词进行相似性匹配，确定用于与所述待匹配名称匹配的标准名称集合。

4、如权利要求 3 所述的方法，其特征在于，所述通过将所述待匹配名称包含的各词与所述第一名称集合中名称包含的各词进行相似性匹配，确定用于与所述待匹配名称匹配的标准名称集合，具体包括：

20 获得所述第一名称集合包含的各名称的索引，所述名称的索引为该名称包含的任一词；以及

对所述待匹配名称进行分词，获得所述待匹配名称包含的各词；

分别将对所述待匹配名称包含的各词，与各所述索引进行相似性匹配，获得由匹配成功的各索引所索引的名称构成的所述第一名称集合的子集；

25 根据各所述子集，确定用于与所述待匹配名称匹配的标准名称集合。

5、如权利要求 4 所述的方法，其特征在于，所述分别将对所述待匹配名

称包含的各词，与各所述索引进行相似性匹配，具体包括：

采用字符串匹配算法，分别将对所述待匹配名称包含的各词，与各所述索引进行相似性匹配；

所述字符串匹配算法包括以下至少一种：前缀树匹配算法、字典树匹配算法、字符串相似度匹配算法、发音相似匹配算法。

6、如权利要求 1 所述的方法，其特征在于，所述对所述待匹配名称进行检测前，所述方法还包括：

根据所述标准名称集合中名称包含的各词与所述待匹配名称包含的各词的相似度，对所述标准名称集合中名称包含的各词与所述待匹配名称包含的各词进行对齐处理；

所述对所述待匹配名称进行检测，具体包括：

根据对齐后的所述标准名称集合，对所述待匹配名称进行检测，以判定所述待匹配名称是否与所述标准名称集合中至少一个名称同义而字符不全相同。

7、如权利要求 1 所述的方法，其特征在于，所述对所述待匹配名称进行检测，具体包括：

对所述待匹配名称进行缩写检测、称谓词检测、多语言检测、别称检测中的至少一种，以判定所述待匹配名称是否与所述标准名称集合中至少一个名称同义而字符不全相同。

8、如权利要求 7 所述的方法，其特征在于，所述对所述待匹配名称进行缩写检测，具体包括：

获得预定的缩写对照组合数据，每个所述缩写对照组合反映至少一个词与其缩写词的缩写对应关系；

根据所述缩写对照组合数据，检测所述待匹配名称包含的词与所述标准名称集合中的名称包含的词是否具有所述缩写对应关系；

根据检测结果，判定所述待匹配名称是否与所述标准名称集合中至少一个名称同义而字符不全相同。

9、如权利要求 7 所述的方法，其特征在于，所述对所述待匹配名称进行称谓词检测，具体包括：

获得预定的称谓词数据；

根据所述称谓词数据，检测所述待匹配名称是否包含所述称谓词，所述称谓词被认为不影响包含该称谓词的名称之义；

根据检测结果，判定所述待匹配名称是否与所述标准名称集合中至少一个名称同义而字符不全相同。

10、如权利要求 7 所述的方法，其特征在于，所述别称包括其对应的名称的昵称、或其对应的名称在不同领域下的同义名称；

10 所述对所述待匹配名称进行别称检测，具体包括：

对所述待匹配名称进行昵称检测，和/或，对所述待匹配名称进行不同领域下的同义名称检测。

11、如权利要求 7 所述的方法，其特征在于，所述对所述待匹配名称进行多语言检测，具体包括：

15 确定所述待匹配名称对应的语言；

获得所述语言和/或其他语言的拼写变形同义规则和/或拼写变形异义规则；

根据所述拼写变形同义规则和/或拼写变形异义规则，对所述待匹配名称进行检测，以判定所述待匹配名称是否与所述标准名称集合中至少一个名称同义而字符不全相同。

12、如权利要求 7 所述的方法，其特征在于，所述对所述待匹配名称进行别称检测，具体包括：

获得预定的别称对照组合数据，每个所述缩写对照组合反映至少一个词与其别称的别称对应关系；

25 根据所述别称对照组合数据，检测所述待匹配名称包含的词与所述标准名称集合中的名称包含的词是否具有所述别称对应关系；

根据检测结果,判定所述待匹配名称是否与所述标准名称集合中至少一个名称同义而字符不全相同。

13、如权利要求 1 所述的方法,其特征在于,所述根据检测结果,确定所述待匹配名称的匹配结果,具体包括:

5 若判定所述待匹配名称与所述标准名称集合中至少一个名称同义而字符不全相同,将所述至少一个名称确定为所述待匹配名称的匹配结果;

否则,通过采用一种或多种相似度算法对所述待匹配名称与所述标准名称集合中的名称进行匹配,确定所述待匹配名称的匹配结果。

14、如权利要求 13 所述的方法,其特征在于,所述相似度算法包括以下
10 至少一种:用于计算文本匹配度的算法、用于计算语音匹配度的算法、用于计算字符串匹配度的算法。

15、如权利要求 1~14 任一项所述的方法,其特征在于,所述名称为人名。

16、一种名称匹配装置,其特征在于,包括:

获得模块,获得待匹配名称;

15 确定模块,确定用于与所述待匹配名称匹配的标准名称集合;

检测模块,对所述待匹配名称进行检测,以判定所述待匹配名称是否与所述标准名称集合中至少一个名称同义而字符不全相同;

匹配模块,根据检测结果,确定所述待匹配名称的匹配结果。

17、如权利要求 16 所述的装置,其特征在于,所述确定模块确定用于与
20 所述待匹配名称匹配的标准名称集合前,获得预定的无需匹配名称集合,确定所述待匹配名称未包含在所述无需匹配名单集合中。

18、如权利要求 16 所述的装置,其特征在于,所述确定模块确定用于与
所述待匹配名称匹配的标准名称集合,具体包括:

所述确定模块获得可用于与所述待匹配名称匹配的第一名称集合,通过将
25 所述待匹配名称包含的各词与所述第一名称集合中名称包含的各词进行相似性匹配,确定用于与所述待匹配名称匹配的标准名称集合。

19、如权利要求 18 所述的装置，其特征在于，所述确定模块通过将所述待匹配名称包含的各词与所述第一名称集合中名称包含的各词进行相似性匹配，确定用于与所述待匹配名称匹配的标准名称集合，具体包括：

所述确定模块获得所述第一名称集合包含的各名称的索引，所述名称的索引为该名称包含的任一词，以及对所述待匹配名称进行分词，获得所述待匹配名称包含的各词，分别将对所述待匹配名称包含的各词，与各所述索引进行相似性匹配，获得由匹配成功的各索引所索引的名称构成的所述第一名称集合的子集，根据各所述子集，确定用于与所述待匹配名称匹配的标准名称集合。

20、如权利要求 19 所述的装置，其特征在于，所述确定模块分别将对所述待匹配名称包含的各词，与各所述索引进行相似性匹配，具体包括：

所述确定模块采用字符串匹配算法，分别将对所述待匹配名称包含的各词，与各所述索引进行相似性匹配，所述字符串匹配算法包括以下至少一种：前缀树匹配算法、字典树匹配算法、字符串相似度匹配算法、发音相似匹配算法。

21、如权利要求 16 所述的装置，其特征在于，所述装置还包括：

对齐模块，在所述检测模块对所述待匹配名称进行检测前，根据所述标准名称集合中名称包含的各词与所述待匹配名称包含的各词的相似度，对所述标准名称集合中名称包含的各词与所述待匹配名称包含的各词进行对齐处理；

所述检测模块对所述待匹配名称进行检测，具体包括：

所述检测模块根据对齐后的所述标准名称集合，对所述待匹配名称进行检测，以判定所述待匹配名称是否与所述标准名称集合中至少一个名称同义而字符不全相同。

22、如权利要求 16 所述的装置，其特征在于，所述检测模块对所述待匹配名称进行检测，具体包括：

所述检测模块对所述待匹配名称进行缩写检测、称谓词检测、多语言检测、别称检测中的至少一种，以判定所述待匹配名称是否与所述标准名称集合中至

少一个名称同义而字符不全相同。

23、如权利要求 22 所述的装置，其特征在于，所述检测模块对所述待匹配名称进行缩写检测，具体包括：

所述检测模块获得预定的缩写对照组合数据，每个所述缩写对照组合反映至少一个词与其缩写词的缩写对应关系，根据所述缩写对照组合数据，检测所述待匹配名称包含的词与所述标准名称集中的名称包含的词是否具有所述缩写对应关系，根据检测结果，判定所述待匹配名称是否与所述标准名称集中至少一个名称同义而字符不全相同。

24、如权利要求 22 所述的装置，其特征在于，所述检测模块对所述待匹配名称进行称谓词检测，具体包括：

所述检测模块获得预定的称谓词数据，根据所述称谓词数据，检测所述待匹配名称是否包含所述称谓词，所述称谓词被认为不影响包含该称谓词的名称之义，根据检测结果，判定所述待匹配名称是否与所述标准名称集中至少一个名称同义而字符不全相同。

25、如权利要求 22 所述的装置，其特征在于，所述别称包括其对应的名称的昵称、或其对应的名称在不同领域下的同义名称；

所述检测模块对所述待匹配名称进行别称检测，具体包括：

所述检测模块对所述待匹配名称进行昵称检测，和/或，对所述待匹配名称进行不同领域下的同义名称检测。

26、如权利要求 22 所述的装置，其特征在于，所述检测模块对所述待匹配名称进行多语言检测，具体包括：

所述检测模块确定所述待匹配名称对应的语言，获得所述语言和/或其他语言的拼写变形同义规则和/或拼写变形异义规则，根据所述拼写变形同义规则和/或拼写变形异义规则，对所述待匹配名称进行检测，以判定所述待匹配名称是否与所述标准名称集中至少一个名称同义而字符不全相同。

27、如权利要求 22 所述的装置，其特征在于，所述检测模块对所述待匹

配名称进行别称检测，具体包括：

所述检测模块获得预定的别称对照组合数据，每个所述缩写对照组合反映至少一个词与其别称的别称对应关系，根据所述别称对照组合数据，检测所述待匹配名称包含的词与所述标准名称集中的名称包含的词是否具有所述别
5 称对应关系，根据检测结果，判定所述待匹配名称是否与所述标准名称集中至少一个名称同义而字符不全相同。

28、如权利要求 16 所述的装置，其特征在于，所述匹配模块根据检测结果，确定所述待匹配名称的匹配结果，具体包括：

所述匹配模块若所述检测模块判定所述待匹配名称与所述标准名称集合
10 中至少一个名称同义而字符不全相同，将所述至少一个名称确定为所述待匹配名称的匹配结果，否则，通过采用一种或多种相似度算法对所述待匹配名称与
所述标准名称集中的名称进行匹配，确定所述待匹配名称的匹配结果。

29、如权利要求 28 所述的装置，其特征在于，所述相似度算法包括以下
至少一种：用于计算文本匹配度的算法、用于计算语音匹配度的算法、用于计
15 算字符串匹配度的算法。

30、如权利要求 16~29 任一项所述的装置，其特征在于，所述名称为人名。

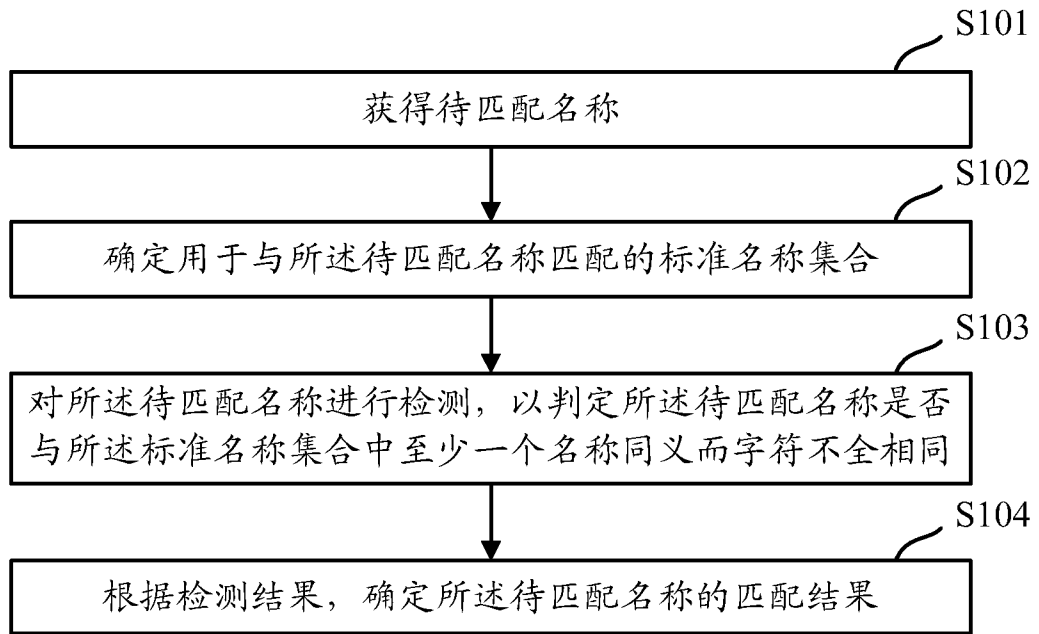


图 1

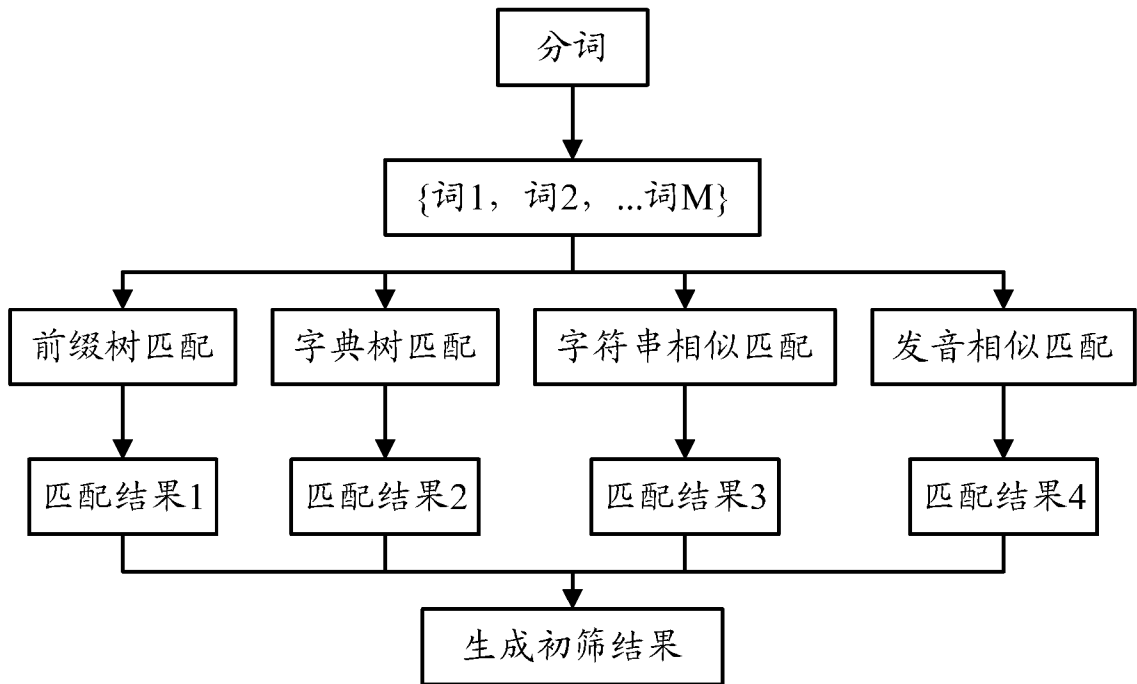


图 2

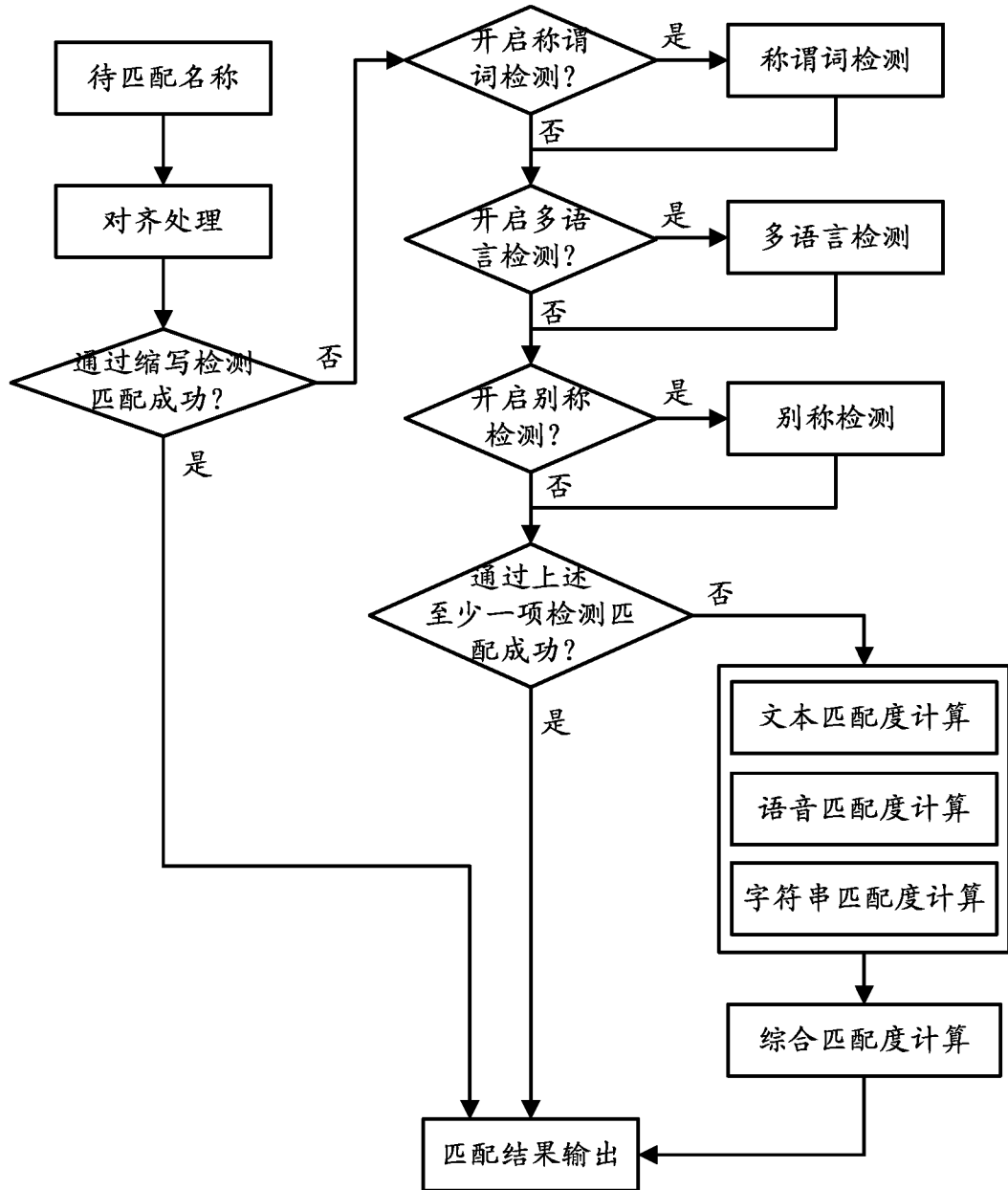


图 3

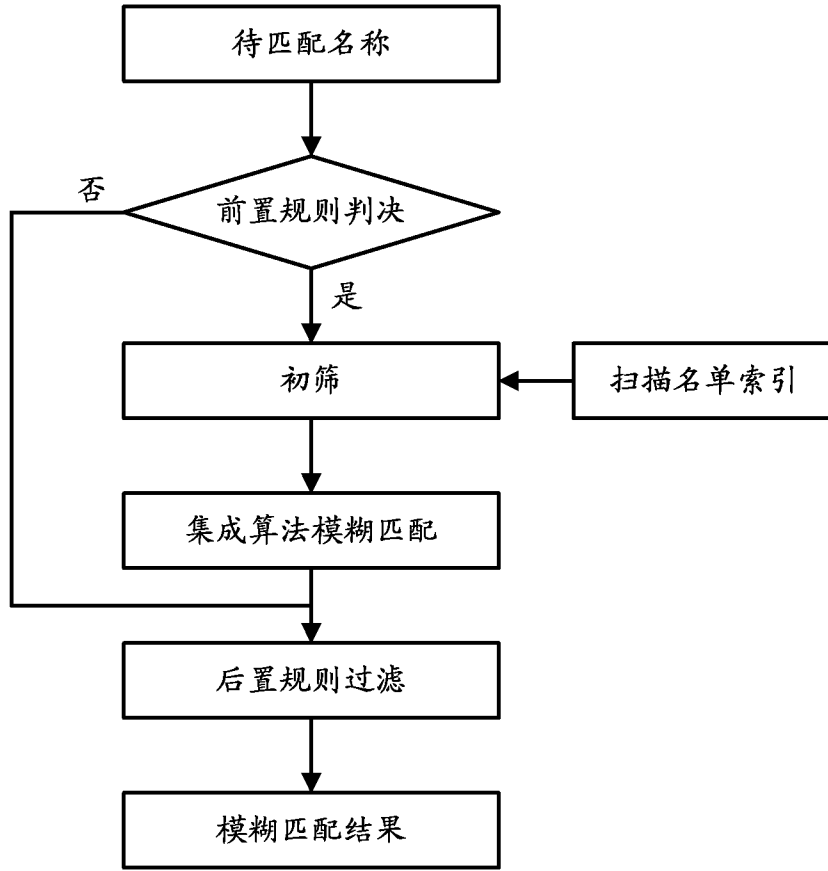


图 4

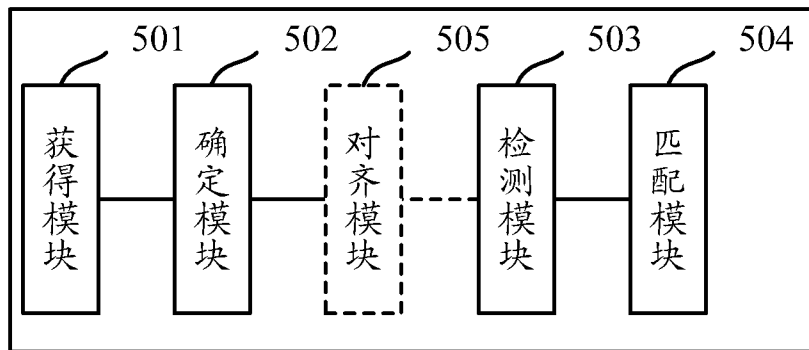


图 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/111604

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)

WPI, EPODOC, CNPAT, CNKI: 匹配, 名称, 人名, 域名, 文本, 字符, 字符串, 相似性, 同义, 索引, 分词, 对齐, 语言, 缩写, 称谓, 别称, 昵称, match+, name, title, domain, text, character, string, similar+, synonymous, synonym, index, participle, language, abbreviat+, appellation, alternative name, nickname

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 103167056 A (COMPUTER NETWORK INFORMATION CENTER, CHINESE ACADEMY OF SCIENCES) 19 June 2013 (19.06.2013), claims 1-6, and description, paragraphs [0006]-[0018]	1, 2, 6-12, 15-17, 21-27, 30
A	CN 105843950 A (LETV HOLDING (BEIJING) CO., LTD. et al.) 10 August 2016 (10.08.2016), entire document	1-30
A	CN 104331475 A (ZHENGZHOU XIZHI INFORMATION TECHNOLOGY CO., LTD.) 04 February 2015 (04.02.2015), entire document	1-30
A	CN 104765858 A (SHANGHAI BRANCH OF BEIJING AEROSPACE CHANGFENG SCIENCE TECHNOLOGY INDUSTRY GROUP CORP. et al.) 08 July 2015 (08.07.2015), entire document	1-30

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>“&” document member of the same patent family</p>
---	---

Date of the actual completion of the international search 06 February 2018	Date of mailing of the international search report 14 February 2018
---	--

<p>Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451</p>	<p>Authorized officer HU, Ni Telephone No. (86-10) 53961429</p>
--	---

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/111604

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 2860645 A1 (G-CLOUD TECHNOLOGY LTD.) 15 April 2015 (15.04.2015), entire document	1-30

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/111604

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 103167056 A	19 June 2013	None	
CN 105843950 A	10 August 2016	None	
CN 104331475 A	04 February 2015	None	
CN 104765858 A	08 July 2015	None	
EP 2860645 A1	15 April 2015	CN 103425739 A	04 December 2013
		WO 2015003421 A1	15 January 2015
		US 2016224552 A1	04 August 2016

国际检索报告

国际申请号

PCT/CN2017/111604

<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>WPI, EPODOC, CNPAT, CNKI: 匹配, 名称, 人名, 域名, 文本, 字符, 字符串, 相似性, 同义, 索引, 分词, 对齐, 语言, 缩写, 称谓, 别称, 昵称, match+, name, title, domain, text, character, string, similar+, synonymous, synonym, index, participle, language, abbreviat+, appellation, alternative name, nickname</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>X</td> <td>CN 103167056 A (中国科学院计算机网络信息中心) 2013年 6月 19日 (2013 - 06 - 19) 权利要求1-6, 说明书第[0006]-[0018]段</td> <td>1-2, 6-12, 15-17, 21-27, 30</td> </tr> <tr> <td>A</td> <td>CN 105843950 A (乐视控股北京有限公司 等) 2016年 8月 10日 (2016 - 08 - 10) 全文</td> <td>1-30</td> </tr> <tr> <td>A</td> <td>CN 104331475 A (郑州悉知信息技术有限公司) 2015年 2月 4日 (2015 - 02 - 04) 全文</td> <td>1-30</td> </tr> <tr> <td>A</td> <td>CN 104765858 A (北京航天长峰科技工业集团有限公司上海分公司 等) 2015年 7月 8日 (2015 - 07 - 08) 全文</td> <td>1-30</td> </tr> <tr> <td>A</td> <td>EP 2860645 A1 (G-CLOUD TECHNOLOGY LTD.) 2015年 4月 15日 (2015 - 04 - 15) 全文</td> <td>1-30</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	X	CN 103167056 A (中国科学院计算机网络信息中心) 2013年 6月 19日 (2013 - 06 - 19) 权利要求1-6, 说明书第[0006]-[0018]段	1-2, 6-12, 15-17, 21-27, 30	A	CN 105843950 A (乐视控股北京有限公司 等) 2016年 8月 10日 (2016 - 08 - 10) 全文	1-30	A	CN 104331475 A (郑州悉知信息技术有限公司) 2015年 2月 4日 (2015 - 02 - 04) 全文	1-30	A	CN 104765858 A (北京航天长峰科技工业集团有限公司上海分公司 等) 2015年 7月 8日 (2015 - 07 - 08) 全文	1-30	A	EP 2860645 A1 (G-CLOUD TECHNOLOGY LTD.) 2015年 4月 15日 (2015 - 04 - 15) 全文	1-30
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
X	CN 103167056 A (中国科学院计算机网络信息中心) 2013年 6月 19日 (2013 - 06 - 19) 权利要求1-6, 说明书第[0006]-[0018]段	1-2, 6-12, 15-17, 21-27, 30																		
A	CN 105843950 A (乐视控股北京有限公司 等) 2016年 8月 10日 (2016 - 08 - 10) 全文	1-30																		
A	CN 104331475 A (郑州悉知信息技术有限公司) 2015年 2月 4日 (2015 - 02 - 04) 全文	1-30																		
A	CN 104765858 A (北京航天长峰科技工业集团有限公司上海分公司 等) 2015年 7月 8日 (2015 - 07 - 08) 全文	1-30																		
A	EP 2860645 A1 (G-CLOUD TECHNOLOGY LTD.) 2015年 4月 15日 (2015 - 04 - 15) 全文	1-30																		
<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>“&” 同族专利的文件</p>																				
国际检索实际完成的日期	国际检索报告邮寄日期																			
2018年 2月 6日	2018年 2月 14日																			
ISA/CN的名称和邮寄地址	受权官员																			
中华人民共和国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088	胡妮																			
传真号 (86-10)62019451	电话号码 (86-10)53961429																			

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/111604

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	103167056	A	2013年 6月 19日	无			
CN	105843950	A	2016年 8月 10日	无			
CN	104331475	A	2015年 2月 4日	无			
CN	104765858	A	2015年 7月 8日	无			
EP	2860645	A1	2015年 4月 15日	CN	103425739	A	2013年 12月 4日
				WO	2015003421	A1	2015年 1月 15日
				US	2016224552	A1	2016年 8月 4日

表 PCT/ISA/210 (同族专利附件) (2009年7月)