

(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 9 月 28 日 (28.09.2017)



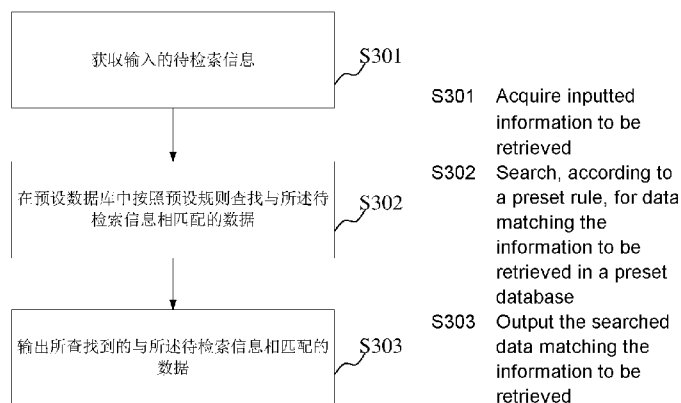
(10) 国际公布号
WO 2017/161749 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2016/088684
- (22) 国际申请日: 2016 年 7 月 5 日 (05.07.2016)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201610161559.1 2016 年 3 月 21 日 (21.03.2016) CN
- (71) 申请人: 乐视控股(北京)有限公司 (LE HOLDINGS (BEIJING) CO., LTD.) [CN/CN]; 中国北京市朝阳区姚家园路 105 号 3 号楼 10 层 1102, Beijing 100025 (CN)。 乐视网信息技术(北京)股份有限公司 (LE SHI INTERNET INFORMATION & TECHNOLOGY CORP., BEIJING) [CN/CN]; 中国北京市朝阳区姚家园路 105 号乐视大厦 10 层, Beijing 100025 (CN)。
- (72) 发明人: 王乙丁 (WANG, Yiding); 中国北京市朝阳区姚家园路 105 号乐视大厦 10 层, Beijing 100025 (CN)。
- (74) 代理人: 北京弘权知识产权代理事务所(普通合伙) (CHINABLE IP); 中国北京市朝阳区安定路 35 号六层 35-10-2 内 620 室, 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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, 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)。

[见续页]

(54) Title: METHOD AND DEVICE FOR INFORMATION MATCHING

(54) 发明名称: 一种信息匹配方法及装置



(57) Abstract: A method and a device for information matching, the method comprising: acquiring inputted information to be retrieved (S301); searching in a preset database, according to a preset rule, for data matching the information to be retrieved (S302); outputting the searched data which matches the information to be retrieved (S303); the data in the preset database being stored in a Trie tree, the Trie tree comprising fuzzy nodes in addition to ordinary nodes, the fuzzy node being used for crossing over respective sister ordinary nodes in the same level to directly reach sub nodes of the respective sister ordinary nodes in the same level when a fuzzy search is needed. The method improves the structure of the Trie tree by adding fuzzy nodes on the basis of ordinary nodes of the Trie tree so as to make it a more suitable data structure for a fuzzy match, in practice reducing the search and comparison times, thereby improving the speed of the fuzzy search.

(57) 摘要:

[见续页]



本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

一种信息匹配方法及装置，其中方法包括：获取输入的待检索信息(S301)；在预设数据库中按照预设规则查找与所述待检索信息相匹配的数据(S302)；输出所查找到的与所述待检索信息相匹配的数据(S303)；其中，所述预设数据库中的数据按照 Trie 树方式存储，并且，所述 Trie 树中除了普通节点外还包括模糊节点，所述模糊节点用于在需要进行模糊查找时越过同一级中的各兄弟普通节点而直接达到所述同一级中的各兄弟普通节点的子节点。该方法对 Trie 树的结构进行了改进，在 Trie 树的普通节点的基础上添加了模糊节点，使之变为更适合模糊匹配的数据结构，使用时可大大减少查找和比较的次数，进而提高了模糊查找的速度。

一种信息匹配方法及装置

本申请要求于 2016 年 3 月 21 日提交中国专利局、申请号为 2016101615591、发明名称为“一种信息匹配方法及装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

5 技术领域

本发明涉及计算机技术，尤其涉及一种信息匹配方法及装置。

背景技术

当前很多网页或 APP 等都提供检索功能，用户在使用时可能会输入错误的查询词，或者是输入模糊的查询词，此时就需要使用模糊匹配技术，为用户提供尽可能正确的检索结果。在现有技术中，可以通过在 Trie 树的结构上使用深度优先搜索算法，来实现对检索词的模糊匹配。

Trie 树又可称前缀树或字典树，是一种匹配查找的常用方法，可以用来排序和保存大量的字符串等信息，所以经常被搜索引擎系统用于文本词频统计。它的优点是：利用字符串的公共前缀来减少查询时间，最大限度地减少无谓的字符串比较，查询效率比哈希树高。深度优先搜索则是一种常用的枚举搜索法，其规则简单说来就是从根节点出发，对每一个可能的分支路径深入到不能再深入为止，且每个节点只能访问一次。

发明人在实现本发明的过程中发现，现有的这种检索方式在使用时查找和比较的次数太多，导致运行速度很慢，效率很低，用户体验不好。

20 发明内容

为克服现有技术中存在的问题，本发明实施例提供一种信息匹配方法及装置，提高模糊检索时的效率。

根据本发明实施例的第一方面，提供一种信息匹配方法，所述方法包括：
获取输入的待检索信息；
在预设数据库中按照预设规则查找与所述待检索信息相匹配的数据；
输出所查找到的与所述待检索信息相匹配的数据；
其中，所述预设数据库中的数据按照 Trie 树方式存储，并且，所述 Trie

树中除了普通节点外还包括模糊节点,所述模糊节点用于在需要进行模糊查找时越过同一级中的各兄弟普通节点而直接达到所述同一级中的各兄弟普通节点的子节点;

所述预设规则包括:

- 5 按照深度优先算法在所述预设数据库查找与所述待检索信息相匹配的数据,当需要选取下一个节点时,如果此时为模糊查找,则只选取下一个节点中的模糊节点。

可选的:

- 10 所述模糊节点分为一级模糊节点和二级模糊节点,所述 Trie 树的每个普通节点下带有一个一级模糊节点,每个一级模糊节点下带有一个二级模糊节点,所述一级模糊节点的父节点的子子节点中的所有普通节点均链接到当前一级模糊节点下,所述二级模糊节点的祖父节点的子子子节点中所有普通节点均链接到当前二级模糊节点下,作为叶子节点的一级模糊节点储存有同父的所有兄弟普通节点中所储存的数据,作为叶子节点的二级模糊节点储存有同祖父的所有兄弟普通节点中所储存的数据。

可选的,在获取输入的待检索信息之前,所述方法还包括:

获取用于生成所述预设数据库的数据;

将所述用于生成所述预设数据库的数据按照 Trie 树方式存储;

- 20 在每个 Trie 树的普通节点下添加一个一级模糊节点,在每个一级模糊节点下添加一个二级模糊节点,将一级模糊节点的父节点的子子节点中的所有普通节点均链接到当前一级模糊节点下,将二级模糊节点的祖父节点的子子子节点中的所有普通节点均链接到当前二级模糊节点下,并令作为叶子节点的一级模糊节点储存有同父的所有兄弟普通节点所储存的数据,作为叶子节点的二级模糊节点储存有同祖父的所有兄弟普通节点所储存的数据,以生成所述预设数据库。

可选的,通过如下方式判断是否为模糊查找:

如果所述待检索信息中包括模糊通配符,则当所述查找进行到所述模糊通配符的位置处时,判断为此时进行模糊查找。

可选的,所述输出所查找到的与所述待检索信息相匹配的数据,包括:

- 30 输出所查找到的叶子节点中所存储的数据。

根据本发明实施例的第二方面,提供一种信息匹配装置,所述装置包括:

获取单元,用于获取输入的待检索信息;

匹配单元,用于在预设数据库中按照预设规则查找与所述待检索信息相匹

配的数据；

输出单元，用于输出所查找到的与所述待检索信息相匹配的数据；

其中，所述预设数据库中的数据按照 Trie 树方式存储，并且，所述 Trie 树中除了普通节点外还包括模糊节点，所述模糊节点用于在需要进行模糊查找时越过同一级中的各兄弟普通节点而直接达到所述同一级中的各兄弟普通节点的子节点；

所述预设规则包括：

按照深度优先算法在所述预设数据库查找与所述待检索信息相匹配的数据，当需要选取下一个节点时，如果此时为模糊查找，则只选取下一个节点中的模糊节点。

可选的：

所述模糊节点分为一级模糊节点和二级模糊节点，所述 Trie 树的每个普通节点下带有一个一级模糊节点，每个一级模糊节点下带有一个二级模糊节点，所述一级模糊节点的父节点的子节点中的所有普通节点均链接到当前一级模糊节点下，所述二级模糊节点的祖父节点的子节点中所有普通节点均链接到当前二级模糊节点下，作为叶子节点的一级模糊节点储存有同父的所有兄弟普通节点中所储存的数据，作为叶子节点的二级模糊节点储存有同祖父的所有兄弟普通节点中所储存的数据。

可选的，所述装置还包括：

预设数据库生成单元，用于：

获取用于生成所述预设数据库的数据；将所述用于生成所述预设数据库的数据按照 Trie 树方式存储；在每个 Trie 树的普通节点下添加一个一级模糊节点，在每个一级模糊节点下添加一个二级模糊节点，将一级模糊节点的父节点的子节点中的所有普通节点均链接到当前一级模糊节点下，将二级模糊节点的祖父节点的子节点中的所有普通节点均链接到当前二级模糊节点下，并令作为叶子节点的一级模糊节点储存有同父的所有兄弟普通节点所储存的数据，作为叶子节点的二级模糊节点储存有同祖父的所有兄弟普通节点所储存的数据，以生成所述预设数据库。

可选的，所述匹配单元通过如下方式判断是否为模糊查找：

如果所述待检索信息中包括模糊通配符，则当所述查找进行到所述模糊通配符的位置处时，判断为此时进行模糊查找。

可选的，所述输出单元用于：

输出所查找到的叶子节点中所存储的数据。

根据本发明实施例的第三方面，还提供一种计算机存储介质，其中，该计算机存储介质可存储有程序，该程序执行时可实现本发明实施例第一方面提供的一种信息匹配方法的各实现方式中的部分或全部步骤。

5 本发明的实施例提供的技术方案可以包括以下有益效果：

本发明实施例对 Trie 树的结构进行了改进，在 Trie 树的普通节点的基础上添加了模糊节点，使之变为更适合模糊匹配的数据结构，使用时可大大减少查找和比较的次数，进而提高了模糊查找的速度。

应当理解的是，以上的一般描述和后文的细节描述仅是示例性和解释性的，并不能限制本发明。

附图说明

此处的附图被并入说明书中并构成本说明书的一部分，示出了符合本发明的实施例，并与说明书一起用于解释本发明的原理。

图 1 是现有技术中 Trie 树的示意图；

15 图 2 是现有技术中在 Trie 树进行查找时的查找顺序示意图；

图 3 是根据本发明一示例性实施例示出的一种信息匹配方法的流程图；

图 4 是根据本发明一示例性实施例示出的带模糊节点的 Trie 树示意图；

图 5 是根据本发明一示例性实施例示出的模糊节点链接示意图；

图 6 是现有技术中进行模糊查找的示意图；

20 图 7 是根据本发明一示例性实施例示出的进行模糊查找的示意图；

图 8 是根据本发明一示例性实施例示出的一种信息匹配方法的流程图；

图 9 是根据本发明一示例性实施例示出的一种信息匹配装置的示意图；

图 10 是根据本发明一示例性实施例示出的一种信息匹配装置的示意图。

具体实施方式

25 这里将详细地对示例性实施例进行说明，其示例表示在附图中。下面的描述涉及附图时，除非另有表示，不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本发明相一致的所有实施方式。相反，它们仅是与如所附权利要求书中所详述的、本发明的一些方面相一致的装置和方法的例子。

30

首先对 Trie 树进行一下简单介绍。假设 Trie 树中存储的数据是“勇者大

冒险”、“勇者大冲关”、“勇者传说”、“圣域传说”、“神话传说”这些词条，参见图 1 所示。再假设某时刻用户输入的待查询关键词是“勇者传说”，则检索系统会将“勇者传说”分为 4 个单字‘勇’、‘者’、‘传’、‘说’，然后从根节点出发，检索‘勇’字，进入该节点，从‘勇’字节点出发，检索‘者’字，进入该节点，如此下去，直至发现全部单字，命中“勇者传说”，搜索结束。该过程可参见图 2 所示，在图 2 中，节点旁边的数字代表搜索节点时的顺序，下同。

图 3 是根据一示例性实施例示出的一种信息匹配方法的流程图。该方法可用于手机、平板电脑、台式机等终端设备，也可用于服务器等。

10 参见图 3 所示，该方法可以包括：

步骤 S301，获取输入的待检索信息。

例如当用户需要检索某信息时，可在手机的搜索框中输入待检索信息，然后点击检索按钮。

15 步骤 S302，在预设数据库中按照预设规则查找与所述待检索信息相匹配的数据。

其中，所述预设数据库中的数据按照 Trie 树方式存储，并且，所述 Trie 树中除了普通节点外还包括模糊节点，所述模糊节点用于在需要进行模糊查找时越过同一级中的各兄弟普通节点而直接达到所述同一级中的各兄弟普通节点的子节点；

20 所述预设规则包括：

按照深度优先算法在所述预设数据库查找与所述待检索信息相匹配的数据，当需要选取下一个节点时，如果此时为模糊查找，则只选取下一个节点中的模糊节点。

25 同一级的各兄弟普通节点可以是同父的子节点，也可以是同祖父的子子节点。

具体的，在本实施例或本发明其他某些实施例中：

30 所述模糊节点分为一级模糊节点和二级模糊节点，所述 Trie 树的每个普通节点下带有一个一级模糊节点，每个一级模糊节点下带有一个二级模糊节点，所述一级模糊节点的父节点的子子节点中的所有普通节点均链接到当前一级模糊节点下，所述二级模糊节点的祖父节点的子子子节点中的所有普通节点均链接到当前二级模糊节点下，作为叶子节点的一级模糊节点储存有同父的所有兄弟普通节点中所储存的数据，作为叶子节点的二级模糊节点储存有同祖父的所有兄弟普通节点中所储存的数据。

作为示例可参见图 4~图 5 所示,在图 4 中,F1 节点为一级模糊节点、F2 节点为二级模糊节点。在图 5 中,F1 节点的父节点的子节点中的所有普通节点均链接到当前 F1 节点下,成为 F1 的子节点,F2 节点的祖父节点的子节点中的所有普通节点均链接到当前 F2 节点下,成为 F2 的子节点。

- 5 对于图 4 和图 5 需要说明的是,理论上每个普通节点下都应该增加一个一级模糊节点,然而实际中有时增加了一级模糊节点后,该模糊节点并无意义,所以在图 4 和图 5 中有些普通节点下并未绘出模糊节点。以‘冲’字节点为例,该节点下并无子节点,所以即使为该节点添加了 F1 节点,该 F1 节点下也没有可以链接的普通节点,所以该 F1 节点也就无存在意义了,故图 4 和图 5 中
- 10 ‘冲’字节点下未绘出一级模糊节点。同理,有些一级模糊节点下也并未绘出二级模糊节点。

步骤 S303,输出所查找到的与所述待检索信息相匹配的数据。

作为示例,所述输出所查找到的与所述待检索信息相匹配的数据,可以包括:

- 15 输出所查找到的叶子节点中所存储的数据。

下面结合示例对 Trie 树所存储的数据进行说明。Trie 树中所述存储可以为字符串,例如由自然语言构成的词组。数据通常主要存储在叶子节点。例如对于“勇者传说”这一词组,对应四个普通节点,各节点分别用于索引‘勇’‘者’‘传’‘说’,而实际的数据即“勇者传说”这一词条,则存储在叶子节点即‘说’字节点中。另外,因为“勇者”一词也可以单独做一个词条,所以

20 ‘者’字节点也可以存储“勇者”一词。

此外,‘说’字节点所存储的数据除了包括“勇者传说”这一词条外,还可以进一步包括其他相关内容,如地址、图片、简介、链接等等,这样用户检索到“勇者传说”时就可以进一步看到该检索结果所关联的其他相关内容了。

- 25 在本发明或本发明其他某些实施例中,可以通过如下方式判断是否为模糊查找:

如果所述待检索信息中包括模糊通配符,则当所述查找进行到所述模糊通配符的位置处时,判断为此时进行模糊查找。

例如模糊通配符可以为“*”、“%”等。

- 30 下面结合具体示例对使用深度优先算法的模糊检索过程进行进一步说明:
假设要检索的内容是“**传说”,其中‘*’代表任意字。如果是在普通 Trie 树上进行模糊检索,参见图 6 所示,其过程可以如下:

1) 将“**传说”分为 4 个单字,即‘*’、‘*’、‘传’、‘说’。

2) 从根节点出发, 遍历其所有的子节点, 首先进入节点‘勇’。

3) 由于第二个字也是‘*’, 从‘勇’节点出发, 遍历其所有节点, 首先进入节点‘者’。

4) 从‘者’子节点出发, 检索‘传’字, 进入该节点。

5) 从‘传’子节点出发, 检索‘说’字, 进入该节点, 发现符合条件的词条“勇者传说”, 于是记录下来。

6) 回到第3)步, 继续‘勇’节点的子节点的遍历, 重复3)~5)。在这里‘勇’节点下面只有一个子节点, 所以遍历结束。

7) 回到第2)步, 继续根节点的子节点的遍历, 重复2)~6)。在这里根节点下面有三个子节点, 要一一遍历。

8) 统计所有记录下的符合条件的结果(“勇者传说”、“圣域传说”、“神话传说”), 搜索结束。

而如果使用本发明中添加了模糊节点的 Trie 树, 参见图 7 所示, 其过程可以如下:

a) 将“**传说”分为 4 个单字, 即‘*’、‘*’、‘传’、‘说’。

b) 从根节点出发, 只进入到它的子节点‘F1’。

c) 由于第二个字也是‘*’, 从‘F1’节点出发, 进入到它的子节点‘F2’。

d) 从‘F2’节点出发, 检索‘传’字, 发现多个节点, 遍历这些节点。

e) 从第一个‘传’子节点出发, 检索‘说’字, 进入该节点, 发现符合条件的词条“勇者传说”, 记录下来。

f) 重复第 e 步, 完成遍历。

g) 统计所有找到的符合条件的结果(“勇者传说”、“圣域传说”、“神话传说”), 搜索结束。

通过对比图 6 和图 7, 可以明确看出本发明检索速度的优势, 因为需要比较的次数大大减少, 在需要进行模糊查找时会越过同一级中的各兄弟普通节点而直接达到所述同一级中的各兄弟普通节点的子节点。

在现实搜索中, 服务器上会储存大量的数据, 一般每个节点下面将会由 3000 左右的子节点。本发明避免两重遍历, 优化了约 3000x3000 倍的速度。这样使得模糊匹配的速度有了显著的提升, 使得算法变得实际可用。当有 80 万数据量的时候, 优化前的 Trie 树模糊匹配, 一次搜索大约用时 5~6 分钟, 优化后的模糊匹配算法搜索用时可忽略不计(小于 1 毫秒)。

此外参见图 8 所示, 在获取输入的待检索信息之前, 所述方法还包括:

步骤 S801, 获取用于生成所述预设数据库的数据。

步骤 S802, 将所述用于生成所述预设数据库的数据按照 Trie 树方式存储。

步骤 S803, 在每个 Trie 树的普通节点下添加一个一级模糊节点, 在每个一级模糊节点下添加一个二级模糊节点, 将一级模糊节点的父节点的子节点中的所有普通节点均链接到当前一级模糊节点下, 将二级模糊节点的祖父节点的子节点中所有普通节点均链接到当前二级模糊节点下, 并令作为叶子节点的一级模糊节点储存有同父的所有兄弟普通节点所储存的数据, 作为叶子节点的二级模糊节点储存有同祖父的所有兄弟普通节点所储存的数据, 以生成所述预设数据库。

本发明实施例对 Trie 树的结构进行了改进, 为 Trie 树的普通节点添加了模糊节点, 使之变为更适合模糊匹配的数据结构, 使用时可大大减少查找和比较的次数, 从而提高了查找速度。

下述为本发明装置实施例, 可以用于执行本发明方法实施例。对于本发明装置实施例中未披露的细节, 请参照本发明方法实施例。

图 9 是根据一示例性实施例示出的一种信息匹配装置的示意图。该装置可用于手机、平板电脑、台式机等终端设备, 也可用于服务器等。

参见图 9 所示, 该装置可以包括:

获取单元 901, 用于获取输入的待检索信息。

例如当用户需要检索某信息时, 可在手机的搜索框中输入待检索信息, 然后点击检索按钮。

匹配单元 902, 用于在预设数据库中按照预设规则查找与所述待检索信息相匹配的数据。

其中, 所述预设数据库中的数据按照 Trie 树方式存储, 并且, 所述 Trie 树中除了普通节点外还包括模糊节点, 所述模糊节点用于在需要进行模糊查找时越过同一级中的各兄弟普通节点而直接达到所述同一级中的各兄弟普通节点的子节点;

所述预设规则包括:

按照深度优先算法在所述预设数据库查找与所述待检索信息相匹配的数据, 当需要选取下一个节点时, 如果此时为模糊查找, 则只选取下一个节点中的模糊节点。

具体的, 在本实施例或本发明其他某些实施例中:

所述模糊节点分为一级模糊节点和二级模糊节点, 所述 Trie 树的每个普通节点下带有一个一级模糊节点, 每个一级模糊节点下带有一个二级模糊节点

点,所述一级模糊节点的父节点的子子节点中的所有普通节点均链接到当前一级模糊节点下,所述二级模糊节点的祖父节点的子子子节点中所有普通节点均链接到当前二级模糊节点下,作为叶子节点的一级模糊节点储存有同父的所有兄弟普通节点中所储存的数据,作为叶子节点的二级模糊节点储存有同祖父的所有兄弟普通节点中所储存的数据。

输出单元 903,用于输出所查找到的与所述待检索信息相匹配的数据。

作为示例,所述输出单元 903 具体可以用于:

输出所查找到的叶子节点中所存储的数据。

在本发明或本发明其他某些实施例中,所述匹配单元 902 可以通过如下方式判断是否为模糊查找:

如果所述待检索信息中包括模糊通配符,则当所述查找进行到所述模糊通配符的位置处时,判断为此时进行模糊查找。

此外,参见图 10 所示,所述装置还可以包括:

预设数据库生成单元 904,用于:

获取用于生成所述预设数据库的数据;将所述用于生成所述预设数据库的数据按照 Trie 树方式存储;在每个 Trie 树的普通节点下添加一个一级模糊节点,在每个一级模糊节点下添加一个二级模糊节点,将一级模糊节点的父节点的子子节点中的所有普通节点均链接到当前一级模糊节点下,将二级模糊节点的祖父节点的子子子节点中的所有普通节点均链接到当前二级模糊节点下,并令作为叶子节点的一级模糊节点储存有同父的所有兄弟普通节点所储存的数据,作为叶子节点的二级模糊节点储存有同祖父的所有兄弟普通节点所储存的数据,以生成所述预设数据库。

本发明实施例对 Trie 树的结构进行了改进,为 Trie 树的普通节点添加了模糊节点,使之变为更适合模糊匹配的数据结构,使用时可大大减少查找和比较的次数,从而提高了查找速度。

关于上述实施例中的装置,其中各个单元\模块执行操作的具体方式已经在有关该方法的实施例中进行了详细描述,此处将不做详细阐述说明。

本发明实施例还提供一种计算机存储介质,其中,该计算机存储介质可存储有程序,该程序执行时可实现图 1~图 8 所涉及实施例提供的一种信息匹配方法的各实现方式中的部分或全部步骤。

本领域技术人员在考虑说明书及实践这里公开的发明后,将容易想到本发

明的其它实施方案。本申请旨在涵盖本发明的任何变型、用途或者适应性变化，这些变型、用途或者适应性变化遵循本发明的一般性原理并包括本发明未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的，本发明的真正范围和精神由所附的权利要求指出。

- 5 应当理解的是，本发明并不局限于上面已经描述并在附图中示出的精确结构，并且可以在不脱离其范围进行各种修改和改变。本发明的范围仅由所附的权利要求来限制。

权 利 要 求

1、一种信息匹配方法，其特征在于，所述方法包括：

获取输入的待检索信息；

在预设数据库中按照预设规则查找与所述待检索信息相匹配的数据；

5 输出所查找到的与所述待检索信息相匹配的数据；

其中，所述预设数据库中的数据按照 Trie 树方式存储，并且，所述 Trie 树中除了普通节点外还包括模糊节点，所述模糊节点用于在需要进行模糊查找时越过同一级中的各兄弟普通节点而直接达到所述同一级中的各兄弟普通节点的子节点；

10 所述预设规则包括：

按照深度优先算法在所述预设数据库查找与所述待检索信息相匹配的数据，当需要选取下一个节点时，如果此时为模糊查找，则只选取下一个节点中的模糊节点。

2、根据权利要求 1 所述的方法，其特征在于：

15 所述模糊节点分为一级模糊节点和二级模糊节点，所述 Trie 树的每个普通节点下带有一个一级模糊节点，每个一级模糊节点下带有一个二级模糊节点，所述一级模糊节点的父节点的子节点中的所有普通节点均链接到当前一级模糊节点下，所述二级模糊节点的祖父节点的子节点中所有普通节点均链接到当前二级模糊节点下，作为叶子节点的一级模糊节点储存有同父的所有兄弟普通节点中所储存的数据，作为叶子节点的二级模糊节点储存有同祖父的所有兄弟普通节点中所储存的数据。

3、根据权利要求 2 所述的方法，其特征在于，在获取输入的待检索信息之前，所述方法还包括：

获取用于生成所述预设数据库的数据；

25 将所述用于生成所述预设数据库的数据按照 Trie 树方式存储；

在每个 Trie 树的普通节点下添加一个一级模糊节点，在每个一级模糊节点下添加一个二级模糊节点，将一级模糊节点的父节点的子节点中的所有普通节点均链接到当前一级模糊节点下，将二级模糊节点的祖父节点的子节点中的所有普通节点均链接到当前二级模糊节点下，并令作为叶子节点的一级模糊节点储存有同父的所有兄弟普通节点所储存的数据，作为叶子节点的二级模糊节点储存有同祖父的所有兄弟普通节点所储存的数据，以生成所述预设数据库。

30

4、根据权利要求 1 所述的方法，其特征在于，通过如下方式判断是否为模糊查找：

如果所述待检索信息中包括模糊通配符，则当所述查找进行到所述模糊通配符的位置处时，判断为此时进行模糊查找。

5 5、根据权利要求 1 所述的方法，其特征在于，所述输出所查找到的与所述待检索信息相匹配的数据，包括：

输出所查找到的叶子节点中所存储的数据。

6、一种信息匹配装置，其特征在于，所述装置包括：

10 获取单元，用于获取输入的待检索信息；

匹配单元，用于在预设数据库中按照预设规则查找与所述待检索信息相匹配的数据；

输出单元，用于输出所查找到的与所述待检索信息相匹配的数据；

15 其中，所述预设数据库中的数据按照 Trie 树方式存储，并且，所述 Trie 树中除了普通节点外还包括模糊节点，所述模糊节点用于在需要进行模糊查找时越过同一级中的各兄弟普通节点而直接达到所述同一级中的各兄弟普通节点的子节点；

所述预设规则包括：

20 按照深度优先算法在所述预设数据库查找与所述待检索信息相匹配的数据，当需要选取下一个节点时，如果此时为模糊查找，则只选取下一个节点中的模糊节点。

7、根据权利要求 6 所述的装置，其特征在于：

25 所述模糊节点分为一级模糊节点和二级模糊节点，所述 Trie 树的每个普通节点下带有一个一级模糊节点，每个一级模糊节点下带有一个二级模糊节点，所述一级模糊节点的父节点的子节点中的所有普通节点均链接到当前一级模糊节点下，所述二级模糊节点的祖父节点的子节点中所有普通节点均链接到当前二级模糊节点下，作为叶子节点的一级模糊节点储存有同父的所有兄弟普通节点中所储存的数据，作为叶子节点的二级模糊节点储存有同祖父的所有兄弟普通节点中所储存的数据。

30 8、根据权利要求 7 所述的装置，其特征在于，所述装置还包括：

预设数据库生成单元，用于：

获取用于生成所述预设数据库的数据；将所述用于生成所述预设数据库的数据按照 Trie 树方式存储；在每个 Trie 树的普通节点下添加一个一级模糊节

点，在每个一级模糊节点下添加一个二级模糊节点，将一级模糊节点的父节点的子节点中的所有普通节点均链接到当前一级模糊节点下，将二级模糊节点的祖父节点的子节点中的所有普通节点均链接到当前二级模糊节点下，并令作为叶子节点的一级模糊节点储存有同父的所有兄弟普通节点所储存的数据，作为叶子节点的二级模糊节点储存有同祖父的所有兄弟普通节点所储存的数据，以生成所述预设数据库。

9、根据权利要求 6 所述的装置，其特征在于，所述匹配单元通过如下方式判断是否为模糊查找：

如果所述待检索信息中包括模糊通配符，则当所述查找进行到所述模糊通配符的位置处时，判断为此时进行模糊查找。

10、根据权利要求 6 所述的装置，其特征在于，所述输出单元用于：
输出所查找到的叶子节点中所存储的数据。

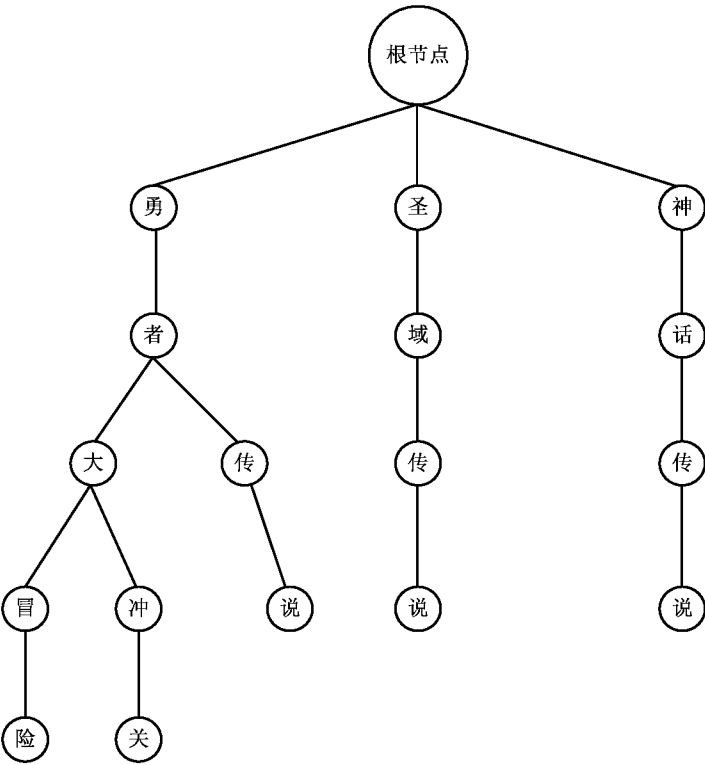


图 1

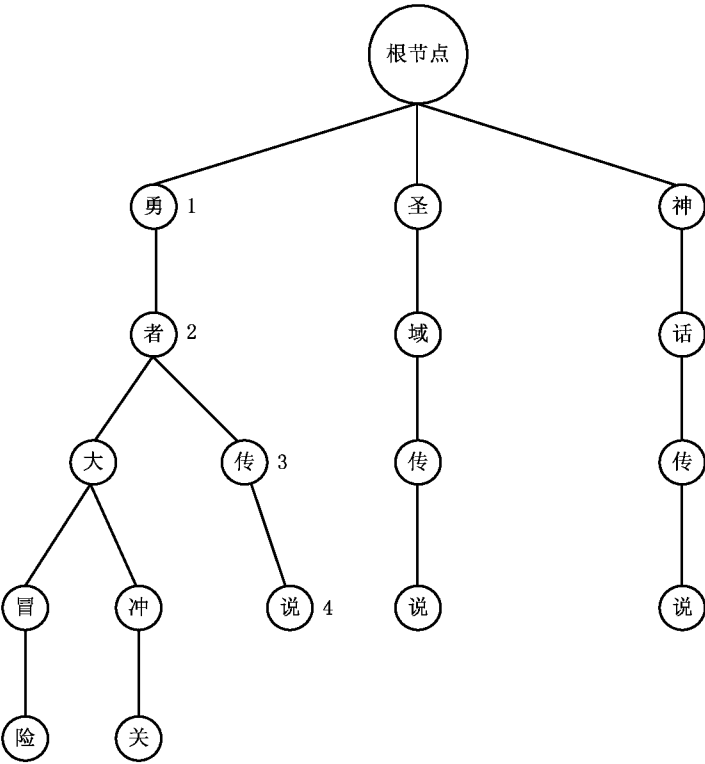


图 2

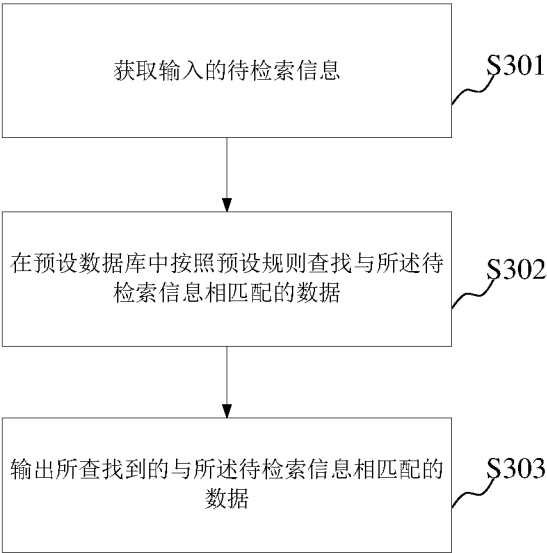


图 3

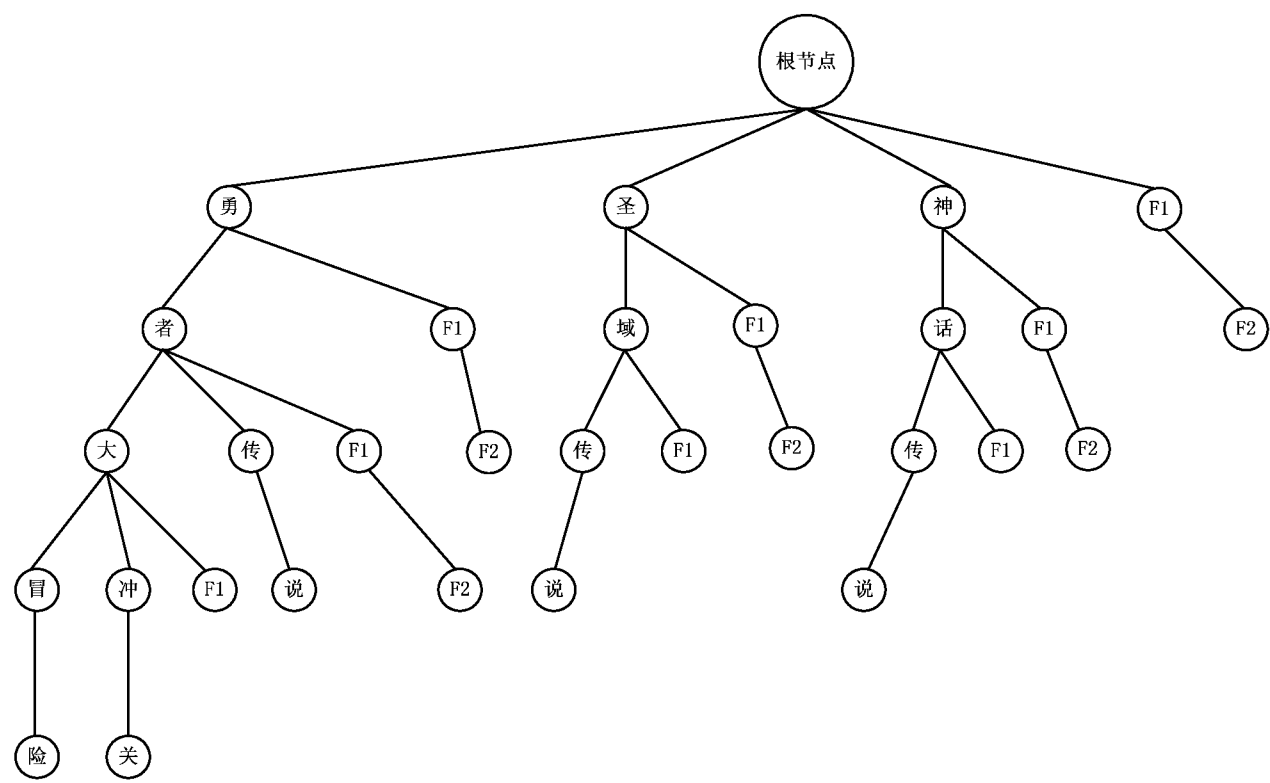


图 4

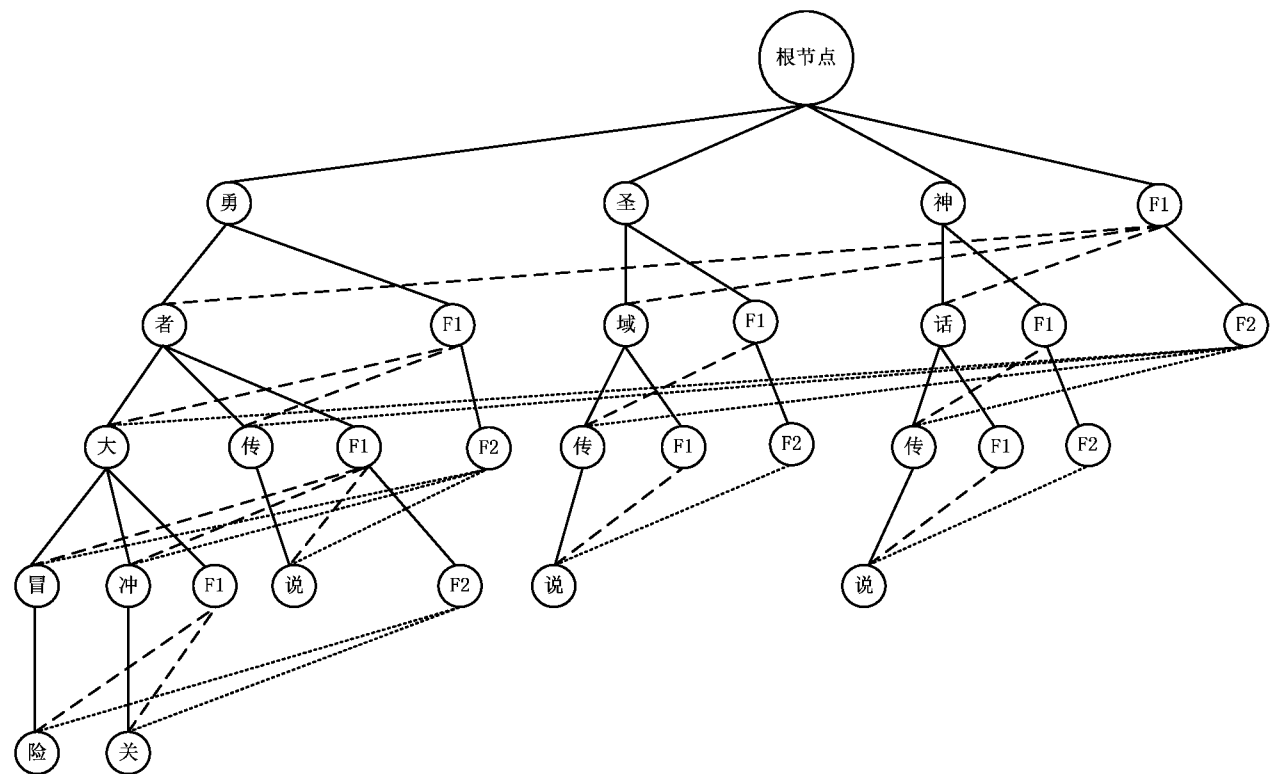


图 5

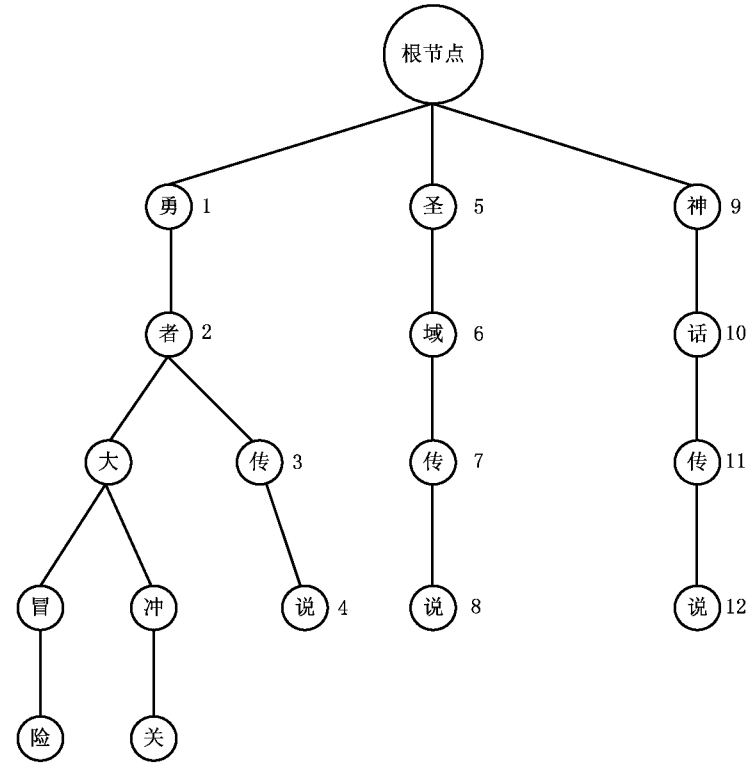


图 6

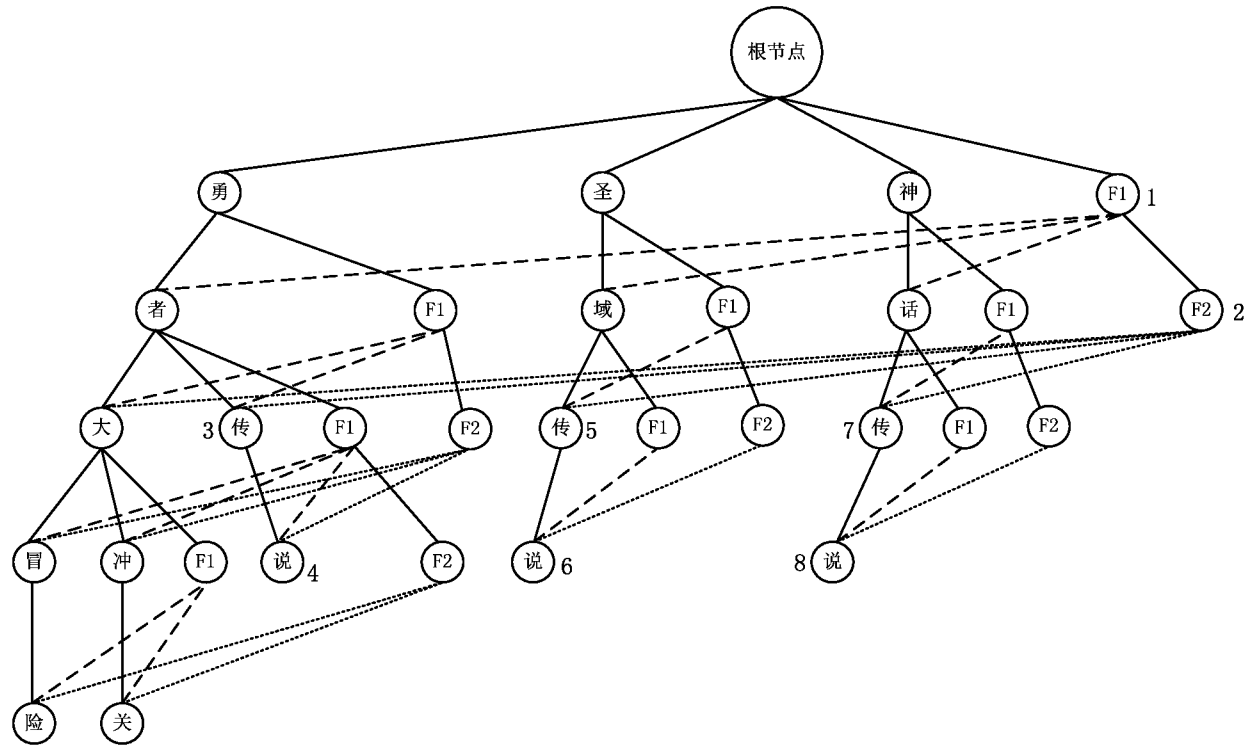


图 7

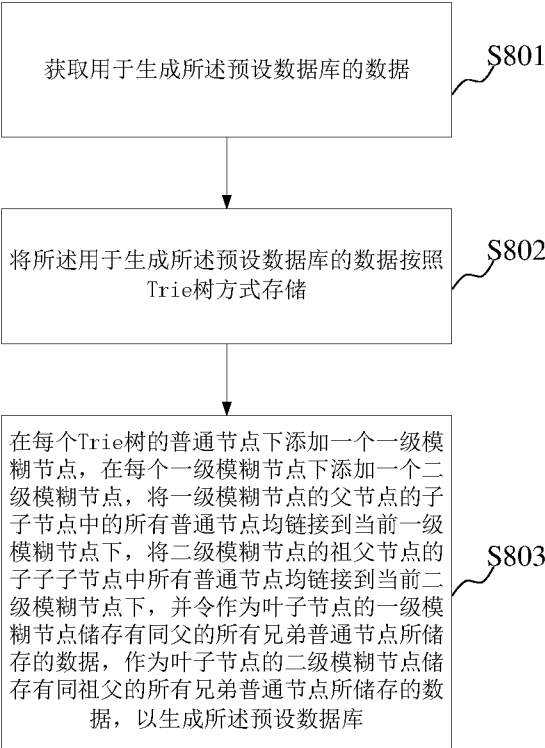


图 8

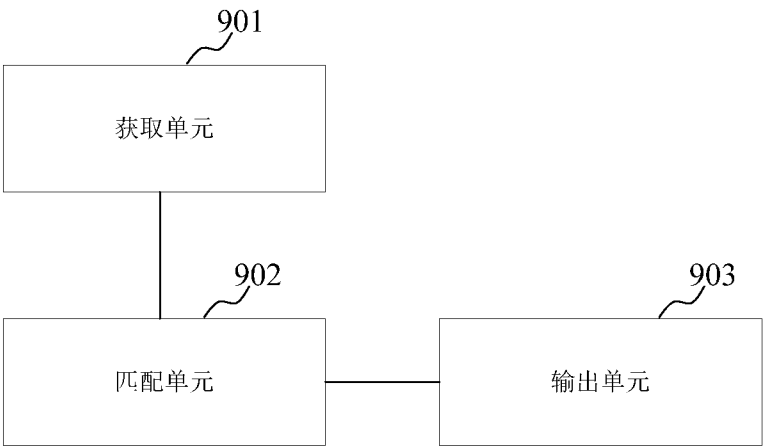


图 9

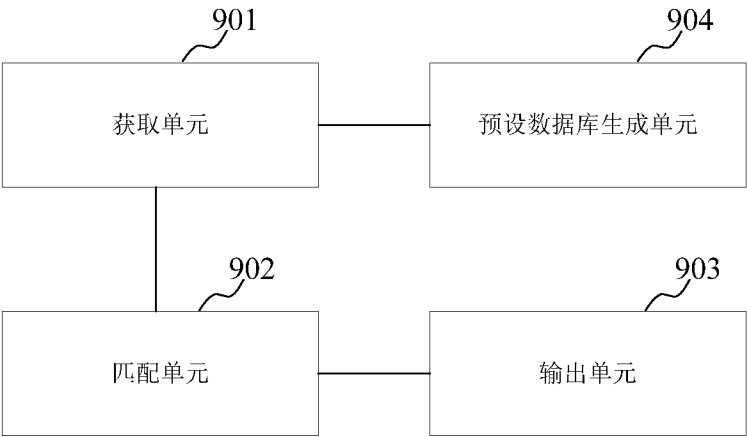


图 10

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/088684

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, IEEE, GOOGLE, CNKI: LETV, WANG, Yiding; TRIE, TRIE tree, vague, prefix tree, lexicographic tree, query, “?”, “*”, level, depth, virtual, add, grandchild node, grandparent node, grandchild 1w node, retriev+, match+, search+, fuzzy, wildcard+, tree, path+, dummy, index, node?, prefix, auxiliary, additional

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 103020052 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.), 03 April 2013 (03.04.2013), description, paragraphs [0239]-[0262] and [0296]-[0300], and figure 7	1-10
A	CN 102646115 A (BEIJING STAR-NET RUIJIE NETWORKS TECHNOLOGY CO., LTD.), 22 August 2012 (22.08.2012), the whole document	1-10
A	CN 101286988 A (BEIJING VENUSTECH INC.), 15 October 2008 (15.10.2008), the whole document	1-10
A	CN 101083608 A (HUAWEI TECHNOLOGIES CO., LTD.), 05 December 2007 (05.12.2007), the whole document	1-10
A	CN 103778124 A (PEKING UNIVERSITY FOUNDER GROUP CORP. et al.), 07 May 2014 (07.05.2014), the whole document	1-10
A	US 6594655 B2 (EZCHIP TECHNOLOGIES LTD.), 15 July 2003 (15.07.2003), the whole document	1-10

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“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
“A” document defining the general state of the art which is not considered to be of particular relevance	“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
“E” earlier application or patent but published on or after the international filing date	“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
“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)	“&” document member of the same patent family
“O” document referring to an oral disclosure, use, exhibition or other means	
“P” document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 15 December 2016 (15.12.2016)	Date of mailing of the international search report 26 December 2016 (26.12.2016)
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 LU, Ran Telephone No.: (86-10) 62414059

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/088684

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 102360385 A (SHENGLE INFORMATION TECHNOLOGY (SHANGHAI) CO., LTD.), 22 February 2012 (22.02.2012), the whole document	1-10
A	CN 101807184 A (ALCATEL-LUCENT), 18 August 2010 (18.08.2010), the whole document	1-10

International application No.
PCT/CN2016/088684

Form PCT/ISA/210 (patent family annex) (July 2009)

A. 主题的分类 G06F 17/30 (2006.01) i 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		
B. 检索领域 检索的最低限度文献 (标明分类系统和分类号) G06F 包含在检索领域中的除最低限度文献以外的检索文献 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)) CNPAT, WPI, EPODOC, IEEE, GOOGLE, CNKI: 乐视, 王乙丁, 树, TRIE, 特里树, 模糊, 含糊, 字典树, 词典树, 辞典树, 前缀树, 搜索, 检索, 匹配, 查找, 查询, 索引, 前缀, 通配, "?", "*", 路径, 层次, 深度, 虚拟, 虚设, 辅助, 增设, 增加, 附加, 节点, 子节点, 祖父节点, 孙 1w 节点, retriev+, match+, search+, fuzzy, wildcard+, tree, path+, dummy, index, node?, prefix, auxiliary, additional		
C. 相关文件		
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	CN 103020052 A (北京百度网讯科技有限公司) 2013年 4月 3日 (2013 - 04 - 03) 说明书第[0239]-[0262], [0296]-[0300]]段、附图7	1-10
A	CN 102646115 A (北京星网锐捷网络技术有限公司) 2012年 8月 22日 (2012 - 08 - 22) 全文	1-10
A	CN 101286988 A (北京启明星辰信息技术股份有限公司) 2008年 10月 15日 (2008 - 10 - 15) 全文	1-10
A	CN 101083608 A (华为技术有限公司) 2007年 12月 5日 (2007 - 12 - 05) 全文	1-10
A	CN 103778124 A (北大方正集团有限公司 等) 2014年 5月 7日 (2014 - 05 - 07) 全文	1-10
A	US 6594655 B2 (EZCHIP TECHNOLOGIES LTD.) 2003年 7月 15日 (2003 - 07 - 15) 全文	1-10
☒ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。		
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “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) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “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 “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 “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 “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 “&” 同族专利的文件		
国际检索实际完成的日期 2016年 12月 15日		国际检索报告邮寄日期 2016年 12月 26日
ISA/CN的名称和邮寄地址 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 传真号 (86-10) 62019451		受权官员 陆然 电话号码 (86-10) 62414059

C. 相关文件		
类 型*	引用文件，必要时，指明相关段落	相关的权利要求
A	CN 102360385 A (盛乐信息技术上海有限公司) 2012年 2月 22日 (2012 - 02 - 22) 全文	1-10
A	CN 101807184 A (阿尔卡特朗讯) 2010年 8月 18日 (2010 - 08 - 18) 全文	1-10

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2016/088684

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	103020052	A	2013年 4月 3日	无	
CN	102646115	A	2012年 8月 22日	无	
CN	101286988	A	2008年 10月 15日	CN 101286988	B 2011年 5月 4日
CN	101083608	A	2007年 12月 5日	无	
CN	103778124	A	2014年 5月 7日	无	
US	6594655	B2	2003年 7月 15日	US 2002143747	A1 2002年 10月 3日
CN	102360385	A	2012年 2月 22日	无	
CN	101807184	A	2010年 8月 18日	CN 101807184	B 2013年 5月 1日