

(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
2014年12月31日 (31.12.2014)



(10) 国际公布号
WO 2014/206186 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2014/079220
- (22) 国际申请日: 2014年6月5日 (05.06.2014)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201310268427.5 2013年6月28日 (28.06.2013) CN
- (71) 申请人: 百度在线网络技术(北京)有限公司
(BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) [CN/CN]; 中国北京市海淀区上地十街10号百度大厦, Beijing 100085 (CN)。
- (72) 发明人: 张伟 (ZHANG, Wei); 中国北京市海淀区上地十街10号百度大厦, Beijing 100085 (CN)。 李海波 (LI, Haibo); 中国北京市海淀区上地十街10号百度大厦, Beijing 100085 (CN)。 徐惠 (XU, Hui); 中国北京市海淀区上地十街10号百度大厦, Beijing 100085 (CN)。 卢佳 (LU, Jia); 中国北京市海淀区上地十街10号百度大厦, Beijing 100085 (CN)。
- (74) 代理人: 北京汉昊知识产权代理事务所(普通合伙) (HANHOW INTELLECTUAL PROPERTY); 中国北京市东长安街1号东方广场西1区1111室, Beijing 100738 (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, LT, 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, 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,

[见续页]

(54) Title: METHOD AND DEVICE FOR GENERATING ENTRY INFORMATION

(54) 发明名称: 一种用于生成词条信息的方法和装置

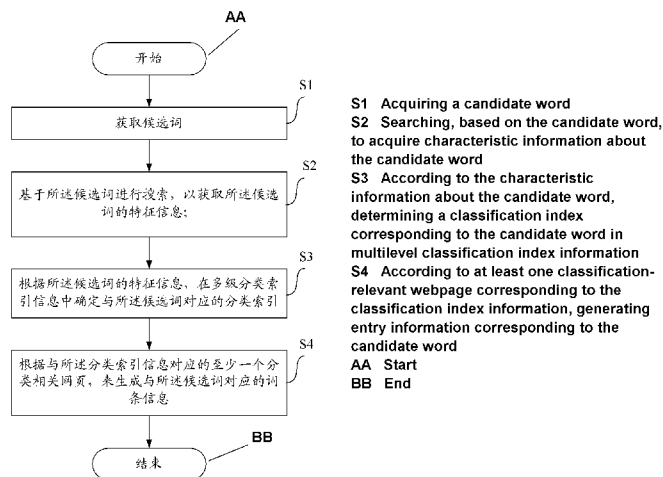


图 1 /FIG. 1

(57) Abstract: Provided are a method and device for generating entry information. The method comprises: acquiring a candidate word; searching, based on the candidate word, to acquire characteristic information about the candidate word; according to the characteristic information about the candidate word, determining a classification index corresponding to the candidate word in multilevel classification index information, wherein the classification index corresponds to at least one classification-relevant webpage; and according to at least one classification-relevant webpage corresponding to the classification index information, generating entry information corresponding to the candidate word. The present invention has the advantages that contents related to the entry can be fully excavated from a professional website related to the entry and the entry information can be automatically generated, so that the efficiency of generating the entry information is improved, and more comprehensive and complete entry information can be acquired.

(57) 摘要:

[见续页]



WO 2014/206186 A1

RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, **本国际公布:**
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, — 包括国际检索报告(条约第 21 条(3))。
TG)。

提供一种用于生成词条信息的方法和装置。该方法包括：获取候选词；基于所述候选词进行搜索，以获取所述候选词的特征信息；根据所述候选词的特征信息，在多级分类索引信息中确定与所述候选词对应的分类索引；其中，所述分类索引对应至少一个分类相关网页；根据与所述分类索引信息对应的至少一个分类相关网页，来生成与所述候选词对应的词条信息。优点在于，能与与词条相关的专业网站中，全面地挖掘与词条相关的内容并自动生成词条信息，从而提高了词条信息的生成效率，并且能够获得更加全面、完整的词条信息。

一种用于生成词条信息的方法和装置

技术领域

5 本发明涉及计算机技术领域，尤其涉及一种用于生成词条信息的方法和装置。

背景技术

在现有技术中，仅能依靠用户手动填写内容来生成百科词条的词条信息，然而，这种方式的效率较低，并且不能及时地对其进行更新；
10 此外，还有一种方式是依据搜索相关词条所得到的网页内容来自动生成词条信息，但是，这种方式所获得的网页类型较为繁杂，且其内容不成系统，所生成的词条信息不够完善，并且，往往无法有效地利用与词条相关的专业类网站中的网页内容。

发明内容

15 本发明的目的是提供一种用于生成词条信息的方法和装置。

根据本发明的一个方面，提供一种用于生成词条信息的方法，其中，所述方法包括以下步骤：

a 获取候选词；

20 b 基于所述候选词进行搜索，以获取所述候选词的特征信息；

c 根据所述候选词的特征信息，在多级分类索引信息中确定与所述候选词对应的分类索引；其中，所述分类索引对应至少一个分类相关网页；

25 d 根据与所述分类索引信息对应的至少一个分类相关网页，来生成与所述候选词对应的词条信息。

根据本发明的一个方面，提供一种用于生成词条信息的词条生成装置，其中，所述词条生成装置包括：

第一获取装置，用于获取候选词；

第二获取装置，用于基于所述候选词进行搜索，以获取所述候选词

的特征信息;

第一确定装置, 用于根据所述候选词的特征信息, 在多级分类索引信息中确定与所述候选词对应的分类索引; 其中, 所述分类索引对应至少一个分类相关网页;

5 第一生成装置, 用于根据与所述分类索引信息对应的至少一个分类相关网页, 来生成与所述候选词对应的词条信息。

本发明的优点在于, 能够从与词条相关的专业网站中, 挖掘与词条相关的内容并自动生成词条信息, 从而提高了词条信息的生成效率, 并且能够获得更加全面、完整的词条信息。

10

附图说明

通过阅读参照以下附图所作的对非限制性实施例所作的详细描述, 本发明的其它特征、目的和优点将会变得更明显:

15 图 1 为根据本发明的一个方面的一种用于生成词条信息的方法流程图;

图 2 为根据本发明的一个优选实施例的用于生成词条信息的方法流程图;

图 3 为根据本发明的又一个优选实施例的用于生成词条信息的方法流程图;

20 图 4 为根据本发明的又一个优选实施例的用于生成词条信息的方法流程图;

图 5 为根据本发明的一个方面用于生成词条信息的词条生成装置的结构示意图;

25 图 6 为根据本发明的一个优选实施例的用于生成词条信息的词条生成装置的结构示意图;

图 7 为根据本发明的又一个优选实施例的用于生成词条信息的词条生成装置的结构示意图;

图 8 为根据本发明的又一个优选实施例的用于生成词条信息的词条生成装置的结构示意图;

附图中相同或相似的附图标记代表相同或相似的部件。

具体实施方式

下面结合附图对本发明作进一步详细描述。

5 图 1 示意出了根据本发明的一个方面的一种用于生成词条信息的方法流程图。根据本发明的方法包括步骤 S1、步骤 S2、步骤 S3 和步骤 S4。

其中，根据本发明的方法通过计算机设备实现。所述计算机设备包括一种能够按照事先设定或存储的指令，自动进行数值计算和/或信息处理的电子设备，其硬件包括但不限于微处理器、专用集成电路 (ASIC)、可编程门阵列 (FPGA)、数字处理器 (DSP)、嵌入式设备等。所述计算机设备包括网络设备和/或用户设备。其中，所述用户设备包括但不限于任何一种可与用户通过键盘、鼠标、遥控器、触摸板、或声控设备等方式进行人机交互的电子产品，例如，个人计算机、平板电脑、智能手机、PDA、游戏机、或 IPTV 等。其中，所述用户设备所处的网络包括但不限于互联网、广域网、城域网、局域网、VPN 网络等。

需要说明的是，所述用户设备以及网络仅为举例，其他现有的或今后可能出现的用户设备以及网络如可适用于本发明，也应包含在本发明保护范围以内，并以引用方式包含于此。

参照图 1，在步骤 S1 中，计算机设备获取候选词。

具体地，所述获取候选词的方式包括但不限于以下任一种方式：

- 1) 实时获取用户输入的查询序列，并将其作为候选词；
- 2) 由预获取的多个查询序列中选择一个作为候选词。

25 接着，在步骤 S2 中，计算机设备基于所述候选词进行搜索，以获取所述候选词的特征信息。

其中，所述特征信息包括一项或多项文本信息。其中，所述文本信息包括但不限于以下任一项：

- a) 词语信息；

b) 段落语言信息。

优选地，所述特征信息包括一项或多项文本信息以及各项文本信息的权重信息。

具体地，所述计算机设备基于所述候选词进行搜索，以获取所述
5 候选词的特征信息的方式包括但不限于以下任一项：

1) 计算机设备在包含多个候选词及其对应的特征信息的词库中搜索，以获得与步骤 S1 中所获得的候选词对应的特征信息。

2) 计算机设备通过第一预定搜索引擎，基于所述候选词执行搜索，以获取与所述候选词对应的一个或多个搜索结果网页；接着，计
10 算机设备根据所述一个或多个搜索结果网页，来确定与所述候选词对应的特征信息。

其中，所述第一预定搜索引擎包括但不限于可基于候选词执行搜索并获取一个或多个搜索结果网页的搜索引擎。

其中，所述计算机设备根据所述一个或多个搜索结果网页，来确
15 定与所述候选词对应的特征信息的方式包括但不限于以下任一项：

a) 获取所述一个或多个搜索结果网页中所包含的至少一个关键词；获取所述至少一个关键词中的各个关键词的权重信息；基于所获得的各个关键词及其相应的权重信息，来确定与所述候选词对应的特征信息。

20 其中，所述权重信息根据以下至少任一项信息来确定：

i) 关键词在所述一个或多个搜索结果网页中的出现频率；

ii) 关键词在所述一个或多个搜索结果网页中的出现次数；

iii) 关键词在所述一个或多个搜索结果网页中出现的区域信息，例如，出现在网页标题部分，或者，出现在网页内容部分等。

25 优选地，所述权重信息基于各个关键词在所述一个或多个搜索结果网页中的词频反文档频率(TF-IDF, term frequency-inverse document frequency) 值来确定。

具体地，计算机设备对一个或多个搜索结果网页的网页内容进行切词处理，以获得至少一个关键词，并统计该至少一个关键

词中的各个关键词的权重信息，接着，根据所获得的各个关键词及其权重信息，由该至少一个关键词中选择一个或多个关键词作为与候选词对应的特征信息。

5 优选地，计算机设备由与所述候选词对应的所有搜索结果网页中选择一个或多个搜索结果网页，并基于该所选择的搜索结果网页来确定与该候选词对应的特征信息。

10 根据本发明的第一示例，计算机设备在步骤 S1 中获得的候选词包括“马尔代夫”，并且，计算机设备通过预定搜索引擎，如百度搜索引擎对“马尔代夫”进行搜索，并获得多个搜索结果网页。计算机设备选择在搜索结果中排名前十位的搜索结果网页 web1 至 web10 作为与该候选词对应的一个或多个搜索结果网页。接着，计算机设备对所选择的十个搜索结果网页的网页内容进行切词以获得多个关键词，并统计每个关键词相对于该十个搜索结果网页的 TF-IDF 值，且将所获得的 TF-IDF 值作为各个关键词的权重信息；并且，计算机设备根据 TF-IDF 值对各个关键词进行排序后选择排名前二十位的关键词，并将该排名前二十位的关键词及其各自对应的 TF-IDF 值作为候选词“马尔代夫”的特征信息。

15 b) 通过预定主题确定模型，根据所述一个或多个搜索结果网页中的各个网页的网页内容，来确定与所述一个或多个搜索结果网页对应的主题相关信息；接着，基于所确定的主题相关信息来确定与
20 所述候选词对应的特征信息。

25 其中，所述预定主题确定模型用于对给定的文本信息通过预定模型来执行数据挖掘等操作，以获得与所述文本信息对应的主题相关信息。例如，潜在狄利克雷分配模型（LDA, Latent Dirichlet Allocation）、概率潜在语义分析模型（PLSA, Probabilistic Latent Semantic Analysis）、带标签的潜在狄利克雷分配模型（Labeled LDA, Labeled Latent Dirichlet Allocation）模型等。

其中，所述主题相关信息包括用于表征所述文本信息的一个或多个主题的信息，例如，用于表征所述文本信息的主题的多个关键

词等。

优选地，所述主题相关信息还包括用于表征该一个或多个主题在所述文本中的权重的信息，例如，与用于表征所述文本信息的主题多个关键词相对应的关键词权重等。

5 其中，本领域技术人员可根据实际情况和需求确定所采用的主题模型，以及通过主题模型获得一个或多个主题相关信息的方式，在此不再赘述。

接着，在步骤 S3 中，计算机设备根据所述候选词的特征信息，在多级分类索引信息中确定与所述候选词对应的分类索引。

10 其中，所述多级分类索引信息包括多个基于预定拓扑结构相互关联的分类索引，其中，各个分类索引分别对应至少一个分类相关网页。

其中，确定多级分类索引信息的方式将在后续参照图 3 所示的实施例中予以详述，并以引用的方式包含于此，在此不再赘述。

15 具体地，计算机设备获取所述候选词的特征信息与多级分类索引信息中的各个分类索引的至少一个分类相关网页之间的相似度，并基于相似度来确定与所述候选词对应的分类索引。

接着，在步骤 S4 中，计算机设备根据与所述分类索引信息对应的至少一个分类相关网页，来确定与所述候选词对应的词条信息。

20 具体地，计算机设备由与所述分类索引相对应的至少一个分类相关网页中，获取与所述候选词相关的网页内容，以生成属于所述分类索引的、与所述候选词对应的词条信息。

其中，计算机设备由至少一个分类相关网页中获取与候选词相关的内容信息的方式包括：

25 计算机设备根据所述候选词及其特征信息，由所述至少一个分类相关网页中挖掘与所述候选词及其特征信息相对应的网页内容，作为与该候选词对应的词条信息的内容信息。

继续对前述第一示例进行说明，多级分类索引信息包括如下表 1 所示的基于预定的树状拓扑结构相关联的分类索引：

表 1

一级分类索引	二级分类索引	三级分类索引
旅游	路线	境内游
		境外游
	住宿	民宿
		酒店

并且，每个分类索引均对应多个分类相关网页，计算机设备在步骤 S3 中确定与候选词“马尔代夫”对应的分类索引为“境内游”，则计算机设备从与分类索引“境内游”对应的多个分类相关网页中获取与候选词“马尔代夫”及其特征信息相关的网页内容，并将其作为与“马尔代夫”这一候选词对应的词条信息的内容，以生成属于分类索引“出境游”的、与候选词“马尔代夫”对应的词条信息。

优选地，当已存在属于所述分类索引的、且与所述候选词对应的词条信息时，计算机设备由与所述分类索引相对应的至少一个分类相关网页中，获取与所述候选词相关的内容信息，以更新该候选词对应的词条信息。

根据本发明的方法，可自动由与候选词具有较高相似度的分类相关中获取词条信息的内容，从而极大的提高了词条信息的生成与更新的效。并且，能够更加充分地挖掘并利用分类相关网页的内容。

图 2 示意出了根据本发明的一个优选实施例的用于生成词条信息的方法流程图。根据本实施例的方法包括步骤 S1 至步骤 S4、步骤 S5、步骤 S6 以及步骤 S7。

其中，步骤 S1 至步骤 S4 已在参照图 1 所示的实施例中予以详述，并以引用的方式包含于此，不再赘述。

在步骤 S5 中，计算机设备获取与所述候选词对应的一项或多项网络发布信息。

其中，所述网络发布信息包括用于在互联网中发布的、具有一定的各类信息。优选地，所述网络发布信息包括广告。

其中，所述计算机设备获取与所述候选词对应的一项或多项网络

发布信息的方式包括但不限于以下任一项：

1) 计算机设备通过在第二预定搜索引擎中查询所述候选词，以获取与所述候选词对应的一项或多项网络发布信息。

其中，所述第二预定搜索引擎包括但不限于可基于候选词执行搜索并获取一个或多个网络发布信息的搜索引擎。

优选地，所述第二预定搜索引擎与前述参照图 1 的实施例中所述的第一预定搜索引擎为同一搜索引擎。

2) 计算机设备通过预定的各个候选词与网络发布信息的对应关系，来获取与该候选词对应的一项或多项网络发布信息。

接着，在步骤 S6 中，计算机设备根据所获得的一项或多项网络发布信息来确定所述候选词的重要度信息。

具体地，所述计算机设备根据所获得的一项或多项网络发布信息来确定所述候选词的重要度信息的方式包括但不限于以下任一项：

1) 计算机设备统计所述候选词相对于所述一项或多项网络发布信息的权重信息。

例如，计算机设备统计所述候选词相对于其所对应的多项广告中的 TF-IDF 值并将其作为候选词的重要度信息。

2) 计算机设备统计所述一项或多项网络发布信息数量，并将其作为所述候选词的重要度信息；

3) 计算机设备获取所述一项或多项网络发布信息的被使用信息，并根据所获得的被使用信息来确定所述候选词的重要度信息。其中，所述网络发布信息的被使用信息包括但不限于以下至少任一项：

a) 所述网络发布信息的被展现次数；

b) 所述网络发布信息的被点击次数等。

例如，计算机设备统计候选词所对应的所有广告的被点击次数，并将其作为候选词的重要度信息；又例如，计算机设备统计候选词所对应的广告的平均被点击次数，以将其作为候选词的重要度信息等。

接着，在步骤 S7 中，计算机设备判断所述候选词的重要度信息是否满足预定重要度条件。

其中，所述预定重要度条件包括预定重要度阈值；

具体地，计算机设备判断所述候选词的重要度信息是否满足预定阈值。

接着，根据本实施例的方法，在步骤 S2 中，当所述候选词的重要度信息满足预定重要度条件时，计算机设备获取所述候选词的特征信息。

作为本实施例的优选方案之一，所述步骤 S5 进一步包括步骤 S501（图未示）和步骤 S502（图未示），所述步骤 S6 进一步包括步骤 S60（图未示）1 和步骤 S602（图未示）。

10 在步骤 S501 中，计算机设备对所述候选词进行切词以获取多个子候选词。

在步骤 S502 中，计算机设备通过第二预定搜索引擎，基于各个子候选词执行搜索以获取与各个子候选词对应的网络发布信息。

其中，所述计算机设备通过第二预定搜索引擎，基于各个子候选词
15 执行搜索以获取与各个子候选词对应的网络发布信息的方式与前述计算机设备通过在第二预定搜索引擎中查询所述候选词，以获取与所述候选词对应的一项或多项网络发布信息的方式相同或相似，故不再赘述。

接着，在步骤 S601 中，计算机设备基于各个子候选词对应的网络
20 发布信息确定该子候选词的子重要度信息。

其中，计算机设备基于各个子候选词对应的网络发布信息确定该子候选词的子重要度信息的方式与前述计算机设备根据所获得的一项或多项网络发布信息来确定所述候选词的重要度信息的方式相同或相似，故不再赘述。

25 在步骤 S602 中，计算机设备基于各个子候选词的子重要度信息确定所述候选词的重要度信息。

具体地，所述计算机设备基于预定的统计规则，确定各个子候选词的子重要度信息。

优选地，计算机设备基于预定的统计规则，确定各个子候选词

的子重要度信息的方式包括但不限于以下任一种：

1) 计算机设备根据各个子候选词的子重要度信息，确定平均重要度信息，并将其作为候选词的重要度信息。

2) 计算机设备获取各个子候选词相对于其所属的候选词的权重值，并基于各个子候选词的子重要度信息以及各个子候选词的权重值，来确定候选词的重要度信息。

例如，基于各个子候选词在其所属的候选词中出现的次数来确定各个子候选词的权重值，并基于各个子候选词的子重要度信息以及各个子候选词的权重值，来确定候选词的重要度信息。

10 根据本实施例的方法，仅对满足预定重要度条件的候选词来生成词条，提高了词条生成效率。

图3示意出了根据本发明的又一个优选实施例的用于生成词条信息的方法流程图。根据本实施例的方法包括步骤S1至步骤S4、步骤S8以及步骤S9。

15 其中，步骤S1至步骤S4已在参照图1所示的实施例中予以详述，并以引用的方式包含于此，不再赘述。

在步骤S8中，计算机设备获取一个或多个网站的网页导航信息。

20 其中，所述一个或多个网站可以为人工指定的具有一定相似度的一个或多个网站，也可以为通过对大量网站的网页内容执行聚类分析后所确定的，具有一定相似度的一个或多个网站。

其中，所述网页导航信息包括但不限于基于网站中的网页栏目结构，为用户浏览网页提供提示的信息。

25 在步骤S9中，计算机设备根据所获得的一个或多个网页导航信息，来生成多级分类索引信息，其中，所述多级分类索引中的各个分类索引按照预定拓扑结构相互关联。

具体地，计算机设备根据所获得的一个或多个网页导航信息，来生成多级分类索引信息的方式包括但不限于以下任一项：

1) 直接将所获得的网页导航信息转换为多级分类索引。

例如，将网站的导航栏中的各个栏目作为分类索引，并依次保存

各个栏目之间的所属关系，以作为各个分类索引之间的所属关系，以生成多级分类索引。

2) 对多个网站的网页导航信息进行选择与合并，并基于选择合并后的结果来生成词条索引信息。

5 例如，将该多个网站的导航栏中共同包含的一个或多个栏目作为分类索引，并选择其中一个网站导航栏中的各个栏目之间的所属关系，作为所获得的各个分类索引之间的所属关系的参考，以生成多级分类索引。

10 作为本实施例的优选方案之一，根据本方案的方法还包括步骤 S10（图未示）、步骤 S11（图未示），所述步骤 S3 进一步包括步骤 S301（图未示）。

在步骤 S10 中，计算机设备基于与所述多级分类索引信息对应的所述一个或多个网站的网页导航信息，获取与该多级分类索引信息中的各个分类索引分别对应的分类相关网页。

15 具体地，计算机设备基于与所述多级分类索引信息对应的所述一个或多个网站的网页导航信息，确定分别与各个分类索引相对应的、所述一个或多个网站的网页导航信息中的部分导航信息，并获取所述一个或多个网站中与该部分导航信息对应的至少一个站点网页，作为与所述分类索引相对应的分类相关网页。

20 接着，在步骤 S11 中，计算机设备基于与所述各个分类索引相对应的分类相关网页来确定与该各个分类索引分别对应的分类特征信息。

其中，计算机设备基于与所述各个分类索引相对应的分类相关网页来确定与该各个分类索引分别对应的分类特征信息的方式与前述
25 参照图 1 所示实施例的步骤 S2 中，计算机设备根据所述一个或多个搜索结果网页，来确定与所述候选词对应的特征信息的方式相同或相似，此处不再赘述。

接着，在步骤 S301 中，计算机设备基于所述候选词的特征信息以及各个分类索引的分类特征信息，确定与所述候选词对应的分类索

引。

具体地，计算机设备将所述候选词的特征信息与各个分类索引的分类特征信息进行比较，并选择分类特征信息与候选词的特征信息的相似度满足预定相似度条件的分类索引，作为与所候选词对应的分类索引。

其中，所述预定相似度条件包括相似度满足预定相似度阈值。

作为本实施例的优选方案之一，所述预定拓扑结构包括多级的拓扑结构，其中相邻两级的分类索引之间为隶属关系，其中，所述步骤 S3 进一步包括步骤 S302（图未示）和步骤 S303（图未示）。

优选地，所述预定拓扑结构包括多级的树状结构，相邻的两级的分类索引之间为隶属关系。

在步骤 S302 中，计算机设备将所述候选词的特征信息与所述各个分类索引的分类特征信息相比较，以获取其分类特征信息与所述候选词的特征信息相似的分类索引。

具体地，计算机设备根据所述预定拓扑结构，按照预定遍历顺序，将所述候选词的特征信息逐个与所述各个分类索引的分类特征信息相比较，以获取其分类特征信息与所述候选词的特征信息相似的分类索引。

例如，当预定拓扑结构为树状结构，并且预定遍历顺序为随机遍历时，随机获取尚未被遍历的分类索引，并将该分类索引的分类特征信息与候选词的特征信息相比较，以获取其分类特征信息与所述候选词的特征信息相似的分类索引。

又例如，当预定拓扑结构为树状结构，并且预定遍历顺序为从叶结点逐层向上遍历时，先获取作为各个叶结点的分类索引，将该层的分类索引的分类特征信息与候选词的特征信息相比较，当未能在叶结点中获得与所述候选词的特征信息相似的分类索引时，再获取各个叶结点上一层的结点的分类索引，并将该层的分类索引的分类特征信息与候选词的特征信息相比较，依次逐层往上，直至获得与所述候选词的特征信息相似的分类索引。

在步骤 S303 中，当所获得的分类索引包含底层分类索引时，计算机设备将该底层分类索引作为所述候选词对应的分类索引。

具体地，计算机设备判断所获得的分类索引是否为底层分类索引，并当所获得的分类索引包含底层分类索引时，计算机设备将该底层分类索引作为所述候选词对应的分类索引。

优选地，根据本方案的方法，所述步骤 S3 还包括步骤 S304（图未示）和步骤 S305（图未示）。

在步骤 S304 中，当所获得的分类索引不包含底层索引节点时，计算机设备基于其中最低级别的分类索引所对应的一个或多个分类相关网页以及所述候选词，来生成位于该最低级别的分类索引的下级分类索引。

具体地，计算机设备基于其中最低级别的分类索引所对应的一个或多个分类相关网页以及所述候选词，来生成位于该最低级别的分类索引的下级分类索引的方式包括但不限于以下任一种：

1) 计算机设备基于候选词生成属于前述步骤 S302 中获得的分类索引的下级分类索引的名称，并基于候选词所对应的搜索结果页面以及所获得的分类索引所对应的分类相关页面，确定与该下级分类索引相对应的分类相关网页。

2) 计算机设备基于前述步骤 S302 中获得的分类索引对应的一个或多个站点网页，在该一个或多个站点网页中查询并获取与候选词相关的至少一个网页，并确定与所该网页对应的中心词，以将其作为前述步骤 S302 中获得的分类索引的下级分类索引的名称，并将该至少一个网页作为与该下级分类索引对应的分类相关网页。

接着，在步骤 S305 中，计算机设备将所生成的底层分类索引作为与所述候选词对应的分类索引。

根据本实施例的方法，通过获取一个或多个网站的网站导航信息来建立多级分类索引，从而使得词条的分类索引体系与实际使用中的体系相近，有利于更加全面的挖掘专业网站的内容信息，并且由于同时还可利用这些网站的网页内容作为分类索引的分类相关网页，故能够

为候选词生成能够有更加系统、完整的词条信息。

图4示意出了根据本发明的又一优选实施例的用于生成词条信息的方法流程图。根据本实施例的方法包括步骤S1至步骤S4、步骤S12、步骤S13、步骤S14以及步骤S15。

5 其中，步骤S1至步骤S4已在参照图1所示的实施例中予以详述，并以引用的方式包含于此，不再赘述。

在步骤S12中，计算机设备获取候选网站的一个或多个网页。

其中，计算机设备确定候选网站的方式包括但不限于以下任一种：

1) 获取人工置顶的网站作为候选网站；

10 2) 将抓取到的网页与多级分类索引信息中的各个分类索引所对应的网页进行比较，以获得站点网页与所述各个分类索引所对应的网页相似的网站。

接着，在步骤S13中，计算机设备根据所述候选网站的一个或多个网页，确定该候选网站的站点特征信息。

15 其中，计算机设备根据所述候选网站的一个或多个网页，确定该候选网站的站点特征信息的方式与前述参照图1所示实施例中的步骤S2中计算机设备根据所述一个或多个搜索结果网页，来确定与所述候选词对应的特征信息的方式相同或相似，在此不再赘述。

接着，在步骤S14中，计算机设备将所述候选网站的站点特征信息
20 与各个分类索引的分类特征信息进行比较，以确定与该候选网站对应的一个或多个分类索引。

其中，计算机设备将所述候选网站的站点特征信息与各个分类索引的分类特征信息进行比较，以确定与该候选网站对应的一个或多个分类索引的方式与前述参照图3所示实施例的步骤S302中，计算机设备
25 将所述候选词的特征信息与所述各个分类索引的分类特征信息相比较，以获取其分类特征信息与所述候选词的特征信息相似的分类索引的方式相同或相似，在此不再赘述。

接着，在步骤S15中，计算机设备向该候选网站对应的候选用户提供该一个或多个分类索引分别对应的一个或多个候选词。

作为本实施力的优选方案，根据本实施例的方法还包括步骤 S16（图未示）、步骤 S17（图未示）以及步骤 S18（图未示）。

在步骤 S16 中，计算机设备根据与所述候选网站对应的一个或多个分类索引，获取所述候选网站中与该一个或多个分类索引分别对应的一个或多个候选网页。

其中，所述计算机设备根据与所述候选网站对应的一个或多个分类索引，获取所述候选网站中与该一个或多个分类索引分别对应的一个或多个候选网页的方式包括但不限于以下任一种：

1) 计算机设备获取该一个或多个分类索引的分类相关网页，将所获得的分类相关网页与所述候选网站的站点网页进行比较，以获得与所述分类相关网页相似的一个或多个站点网页，并将其作为与该分类相关网页所对应的分类索引的候选网页。

2) 计算机设备根据该一个或多个分类索引的分类特征信息，由候选网站中获取分别与该一个或多个分类索引的分类特征信息相似的一个或多个候选网页。

接着，在步骤 S17 中，计算机设备基于与各个分类索引对应的、所述候选网站中的一个或多个候选网页，确定或更新与该各个分类索引对应的分类相关网页。

具体地，计算机设备将所确定的候选网页作为与分类索引对应的分类相关网页添加至与各个分类索引对应的分类相关网页库中。

在步骤 S18 中，计算机设备基于所述更新后的与各个分类索引对应的分类相关网页，更新各个分类索引所对应的候选词的词条信息。

具体地，对属于个各个分类索引的一个或多个候选词，分别采用更新后的该分类索引的分类相关网页来更新各个候选词的词条内容。其中，采用更新后的该分类索引的分类相关网页来更新各个候选词的词条内容的方式与前述参照图 1 所示实施例中的步骤 S4 中计算机设备根据与所述分类索引信息对应的至少一个分类相关网页，来确定与所述候选词对应的词条信息的方式相同或相似，此处不再赘述。

根据本实施例的方法，通过采用候选网站的内容来自动更新词条信息，使得词条内容能够尽快得到更新，并且提高了更新效率。

图 5 示意出了根据本发明的一个方面用于生成词条信息的词条生成装置的结构示意图。根据本发明的词条生成装置包括第一获取装置 1、第二获取装置 2、第一确定装置 3 和第一生成装置 4。

参照图 5，第一获取装置 1 获取候选词。

具体地，所述获取候选词的方式包括但不限于以下任一种方式：

- 1) 实时获取用户输入的查询序列，并将其作为候选词；
- 2) 由预获取的多个查询序列中选择一个作为候选词。

接着，第二获取装置 2 基于所述候选词进行搜索，以获取所述候选词的特征信息。

其中，所述特征信息包括一项或多项文本信息。其中，所述文本信息包括但不限于以下任一项：

- a) 词语信息；
- b) 段落语言信息。

优选地，所述特征信息包括一项或多项文本信息以及各项文本信息的权重信息。

具体地，所述第二获取装置 2 基于所述候选词进行搜索，以获取所述候选词的特征信息的方式包括但不限于以下任一项：

1) 第二获取装置 2 在包含多个候选词及其对应的特征信息的词库中搜索，以获得与步骤 S1 中所获得的候选词对应的特征信息。

2) 第二获取装置 2 中的第一搜索装置（图未示）的通过第一预定搜索引擎，基于所述候选词执行搜索，以获取与所述候选词对应的一个或多个搜索结果网页；接着，第二获取装置 2 中的第二确定装置（图未示）根据所述一个或多个搜索结果网页，来确定与所述候选词对应的特征信息。

其中，所述第一预定搜索引擎包括但不限于可基于候选词执行搜索并获取一个或多个搜索结果网页的搜索引擎。

其中，所述第二确定装置根据所述一个或多个搜索结果网页，来

确定与所述候选词对应的特征信息的方式包括但不限于以下任一项：

a) 第二确定装置中的关键词获取装置（图未示）获取所述一个或多个搜索结果网页中所包含的至少一个关键词；接着，第二确定装置中的权重获取装置（图未示）获取所述至少一个关键词中的各个关键词的权重信息；接着，第二确定装置中的第一子确定装置（图未示）基于所获得的各个关键词及其相应的权重信息，来确定与所述候选词对应的特征信息。

其中，所述权重信息根据以下至少任一项信息来确定：

i) 关键词在所述一个或多个搜索结果网页中的出现频率；

ii) 关键词在所述一个或多个搜索结果网页中的出现次数；

iii) 关键词在所述一个或多个搜索结果网页中出现的区域信息，例如，出现在网页标题部分，或者，出现在网页内容部分等。

优选地，所述权重信息基于各个关键词在所述一个或多个搜索结果网页中的词频反文档频率(TF-IDF, term frequency-inverse document frequency) 值来确定。

具体地，关键词获取装置对一个或多个搜索结果网页的网页内容进行切词处理，以获得至少一个关键词，接着，权重获取装置统计并确定该至少一个关键词中的各个关键词的权重信息，接着，第一子确定装置根据所获得的各个关键词及其权重信息，由该至少一个关键词中选择一个或多个关键词作为与候选词对应的特征信息。

优选地，计算机设备由与所述候选词对应的所有搜索结果网页中选择一个或多个搜索结果网页，并基于该所选择的搜索结果网页来确定与该候选词对应的特征信息。

根据本发明的第一示例，第一获取装置 1 获得的候选词包括“马尔代夫”，并且，第一搜索装置通过预定搜索引擎，如百度搜索引擎对“马尔代夫”进行搜索以获得多个搜索结果网页，并选择在搜索结果中排名前十位的搜索结果网页 web1 至 web10 作为与该候选词对应的一个或多个搜索结果网页。接着，关键词获取

装置权重获取装置对所选择的十个搜索结果网页的网页内容进行切词以获得多个关键词，由权重获取装置统计每个关键词相对于该十个搜索结果网页的 TF-IDF 值，将所获得的 TF-IDF 值作为各个关键词的权重信息；然后，第一子确定装置根据 TF-IDF 值对各个关键词进行排序后选择排名前二十位的关键词，并将该排名前二十位的关键词及其各自对应的 TF-IDF 值作为候选词“马尔代夫”的特征信息。

b) 第二确定装置中的模型确定装置（图未示）通过预定主题确定模型，根据所述一个或多个搜索结果网页中的各个网页的网页内容，来确定与所述一个或多个搜索结果网页对应的主题相关信息；接着，第二确定装置中的第二子确定装置（图未示）基于所确定的主题相关信息来确定与所述候选词对应的特征信息。

其中，所述预定主题确定模型用于对给定的文本信息通过预定模型来执行数据挖掘等操作，以获得与所述文本信息对应的主题相关信息。例如，潜在狄利克雷分配模型（LDA, Latent Dirichlet Allocation）、概率潜在语义分析模型（PLSA, Probabilistic Latent Semantic Analysis）、带标签的潜在狄利克雷分配模型（Labeled LDA, Labeled Latent Dirichlet Allocation）模型等。

其中，所述主题相关信息包括用于表征所述文本信息的一个或多个主题的信息，例如，用于表征所述文本信息的主题或多个关键词等。

优选地，所述主题相关信息还包括用于表征该一个或多个主题在所述文本中的权重的信息，例如，与用于表征所述文本信息的主题或多个关键词相对应的关键词权重等。

其中，本领域技术人员可根据实际情况和需求确定所采用的主题模型，以及通过主题模型获得一个或多个主题相关信息的方式，在此不再赘述。

接着，第一确定装置 3 根据所述候选词的特征信息，在多级分类索引信息中确定与所述候选词对应的分类索引。

其中，所述多级分类索引信息包括多个基于预定拓扑结构相互关联的分类索引，其中，各个分类索引分别对应至少一个分类相关网页。

其中，确定多级分类索引信息的方式将在后续参照图 3 所示的实施例中予以详述，并以引用的方式包含于此，在此不再赘述。

5 具体地，第一确定装置 3 获取所述候选词的特征信息与多级分类索引信息中的各个分类索引的至少一个分类相关网页之间的相似度，并基于相似度来确定与所述候选词对应的分类索引。

接着，第一生成装置 4 根据与所述分类索引信息对应的至少一个分类相关网页，来确定与所述候选词对应的词条信息。

10 具体地，第一生成装置 4 由与所述分类索引相对应的至少一个分类相关网页中，获取与所述候选词相关的网页内容，以生成属于所述分类索引的、与所述候选词对应的词条信息。

其中，第一生成装置 4 由至少一个分类相关网页中获取与候选词相关的内容信息的方式包括：

15 第一生成装置 4 根据所述候选词及其特征信息，由所述至少一个分类相关网页中挖掘与所述候选词及其特征信息相对应的网页内容，作为与该候选词对应的词条信息的内容信息。

继续对前述第一示例进行说明，多级分类索引信息包括如下表 2 所示的基于预定的树状拓扑结构相关联的分类索引：

20

表 2

一级分类索引	二级分类索引	三级分类索引
旅游	路线	境内游
		境外游
	住宿	民宿
		酒店

并且，每个分类索引均对应多个分类相关网页，第一确定装置 3 确定与候选词“马尔代夫”对应的分类索引为“境内游”，则第一生成装置 4 从与分类索引“境内游”对应的多个分类相关网页中获取与候选词“马尔代夫”及其特征信息相关的网页内容，并将其作为与“马

尔代夫”这一候选词对应的词条信息的内容,以生成属于分类索引“出境游”的、与候选词“马尔代夫”对应的词条信息。

优选地,当已存在属于所述分类索引的、且与所述候选词对应的词条信息时,计算机设备由与所述分类索引相对应的至少一个分类相关网页中,获取与所述候选词相关的内容信息,以更新该候选词对应的词条信息。

根据本发明的方案,可自动由与候选词具有较高相似度的分类相关中获取词条信息的内容,从而极大的提高了词条信息的生成与更新的效。并且,能够更加充分地挖掘并利用分类相关网页的内容。

10 图6示意出了根据本发明的一个优选实施例的用于生成词条信息的词条生成装置的结构示意图。根据本实施例的词条生成装置包括第一获取装置1、第二获取装置2、第一确定装置3、第一生成装置4、第三获取装置5、第三确定装置6以及判断装置7。

15 其中,第一获取装置1、第二获取装置2、第一确定装置3以及第一生成装置4已在参照图5所示的实施例中予以详述,并以引用的方式包含于此,不再赘述。

第三获取装置5获取与所述候选词对应的一项或多项网络发布信息。

20 其中,所述网络发布信息包括用于在互联网中发布的、具有一定的各类信息。优选地,所述网络发布信息包括广告。

其中,所述第三获取装置5获取与所述候选词对应的一项或多项网络发布信息的方式包括但不限于以下任一项:

1)第三获取装置5通过在第二预定搜索引擎中查询所述候选词,以获取与所述候选词对应的一项或多项网络发布信息。

25 其中,所述第二预定搜索引擎包括但不限于可基于候选词执行搜索并获取一个或多个网络发布信息的搜索引擎。

优选地,所述第二预定搜索引擎与前述参照图5的实施例中所述的第一预定搜索引擎为同一搜索引擎。

2)第三获取装置5通过预定的各个候选词与网络发布信息的对

应关系，来获取与该候选词对应的一项或多项网络发布信息。

接着，第三确定装置 6 根据所获得的一项或多项网络发布信息来确定所述候选词的重要度信息。

具体地，所述第三确定装置 6 根据所获得的一项或多项网络发布信息来确定所述候选词的重要度信息的方式包括但不限于以下任一项：

1) 第三确定装置 6 统计所述候选词相对于所述一项或多项网络发布信息的权重信息。

例如，第三确定装置 6 统计所述候选词相对于其所对应的多项广告中的 TF-IDF 值并将其作为候选词的重要度信息。

2) 第三确定装置 6 统计所述一项或多项网络发布信息数量，并将其作为所述候选词的重要度信息；

3) 第三确定装置 6 获取所述一项或多项网络发布信息的被使用信息，并根据所获得的被使用信息来确定所述候选词的重要度信息。其中，所述网络发布信息的被使用信息包括但不限于以下至少任一项：

a) 所述网络发布信息的被展现次数；

b) 所述网络发布信息的被点击次数等。

例如，第三确定装置 6 统计候选词所对应的所有广告的被点击次数，并将其作为候选词的重要度信息；又例如，第三确定装置 6 统计候选词所对应的广告的平均被点击次数，以将其作为候选词的重要度信息等。

接着，判断装置 7 判断所述候选词的重要度信息是否满足预定重要度条件。

其中，所述预定重要度条件包括预定重要度阈值；

具体地，判断装置 7 判断所述候选词的重要度信息是否满足预定阈值。

接着，根据本实施例的方案，当所述候选词的重要度信息满足预定重要度条件时，第二获取装置 2 获取所述候选词的特征信息。

作为本实施例的优选方案之一，所述第三获取装置 5 进一步包括第一子获取装置（图未示）和第二搜索装置（图未示），所述第三确定装

置进一步包括第三子确定装置(图未示)和第四子确定装置(图未示)。

第一子获取装置对所述候选词进行切词以获取多个子候选词。

第二搜索装置通过第二预定搜索引擎,基于各个子候选词执行搜索以获取与各个子候选词对应的网络发布信息。

5 其中,所述第二搜索装置通过第二预定搜索引擎,基于各个子候选词执行搜索以获取与各个子候选词对应的网络发布信息的方式与前述第三获取装置 5 通过在第二预定搜索引擎中查询所述候选词,以获取与所述候选词对应的一项或多项网络发布信息的方式相同或相似,故不再赘述。

10 接着,第三子确定装置基于各个子候选词对应的网络发布信息确定该子候选词的子重要度信息。

其中,第三子确定装置基于各个子候选词对应的网络发布信息确定该子候选词的子重要度信息的方式与前述计算机设备根据所获得的一项或多项网络发布信息来确定所述候选词的重要度信息的方式相同或
15 相似,故不再赘述。

第四子确定装置基于各个子候选词的子重要度信息确定所述候选词的重要度信息。

具体地,所述第四子确定装置基于预定的统计规则,确定各个子候选词的子重要度信息。

20 优选地,第四子确定装置基于预定的统计规则,确定各个子候选词的子重要度信息的方式包括但不限于以下任一种:

1)第四子确定装置根据各个子候选词的子重要度信息,确定平均重要度信息,并将其作为候选词的重要度信息。

2)第四子确定装置获取各个子候选词相对于其所属的候选词的权
25 重值,并基于各个子候选词的子重要度信息以及各个子候选词的权重值,来确定候选词的重要度信息。

例如,第四子确定装置基于各个子候选词在其所属的候选词中出现的次数来确定各个子候选词的权重值,并基于各个子候选词的子重要度信息以及各个子候选词的权重值,来确定候选词的重要度信

息。

根据本实施例的方案，仅对满足预定重要度条件的候选词来生成词条，提高了词条生成效率。

图7示意出了根据本发明的又一个优选实施例的用于生成词条信息的词条生成装置的结构示意图。根据本实施例的词条生成装置包括
5 第一获取装置1、第二获取装置2、第一确定装置3、第一生成装置4、导航获取装置8以及第二生成装置9。

其中，第一获取装置1、第二获取装置2、第一确定装置3以及第一生成装置4已在参照图5所示的实施例中予以详述，并以引用的
10 方式包含于此，不再赘述。

导航获取装置8获取一个或多个网站的网页导航信息。

其中，所述一个或多个网站可以为人工指定的具有一定相似度的一个或多个网站，也可以为通过对大量网站的网页内容执行聚类分析后所确定的，具有一定相似度的一个或多个网站。

其中，所述网页导航信息包括但不限于基于网站中的网页栏目结构，为用户浏览网页提供提示的信息。

第二生成装置9根据所获得的一个或多个网页导航信息，来生成多级分类索引信息，其中，所述多级分类索引中的各个分类索引按照
20 预定拓扑结构相互关联。

具体地，第二生成装置9根据所获得的一个或多个网页导航信息，来生成多级分类索引信息的方式包括但不限于以下任一项：

1) 第二生成装置9直接将所获得的网页导航信息转换为多级分类索引。

例如，第二生成装置9将网站的导航栏中的各个栏目作为分类索引，并依次保存各个栏目之间的所属关系，以作为各个分类索引之间的
25 所属关系，以生成多级分类索引。

2) 第二生成装置9对多个网站的网页导航信息进行选择与合并，并基于选择合并后的结果来生成词条索引信息。

例如，第二生成装置9将该多个网站的导航栏中共同包含的一个

或多个栏目作为分类索引，并选择其中一个网站导航栏中的各个栏目之间的所属关系，作为所获得的各个分类索引之间的所属关系的参考，以生成多级分类索引。

5 作为本实施例的优选方案之一，根据本方案的词条生成装置还包括第四获取装置（图未示）、第一特征确定装置（图未示）。

第四获取装置基于与所述多级分类索引信息对应的所述一个或多个网站的网页导航信息，获取与该多级分类索引信息中的各个分类索引分别对应的分类相关网页。

10 具体地，第四获取装置基于与所述多级分类索引信息对应的所述一个或多个网站的网页导航信息，确定分别与各个分类索引相对应的、所述一个或多个网站的网页导航信息中的部分导航信息，并获取所述一个或多个网站中与该部分导航信息对应的至少一个站点网页，作为与所述分类索引相对应的分类相关网页。

15 接着，第一特征确定装置基于与所述各个分类索引相对应的分类相关网页来确定与该各个分类索引分别对应的分类特征信息。

20 其中，第一特征确定装置基于与所述各个分类索引相对应的分类相关网页来确定与该各个分类索引分别对应的分类特征信息的方式与前述参照图5所示实施例中第二确定装置根据所述一个或多个搜索结果网页，来确定与所述候选词对应的特征信息的方式相同或相似，此处不再赘述。

接着，根据本实施例的第一确定装置3基于所述候选词的特征信息以及各个分类索引的分类特征信息，确定与所述候选词对应的分类索引。

25 具体地，第一确定装置3将所述候选词的特征信息与各个分类索引的分类特征信息进行比较，并选择分类特征信息与候选词的特征信息的相似度满足预定相似度条件的分类索引，作为与所候选词对应的分类索引。

其中，所述预定相似度条件包括相似度满足预定相似度阈值。

作为本实施例的优选方案之一，所述预定拓扑结构包括多级的

拓扑结构，其中相邻两级的分类索引之间为隶属关系，其中，所述第一确定装置 3 进一步包括比较获取装置（图未示）和第一分类确定装置（图未示）。

5 优选地，所述预定拓扑结构包括多级的树状结构，相邻的两级的分类索引之间为隶属关系。

比较获取装置将所述候选词的特征信息与所述各个分类索引的分类特征信息相比较，以获取其分类特征信息与所述候选词的特征信息相似的分类索引。

10 具体地，比较获取装置根据所述预定拓扑结构，按照预定遍历顺序，将所述候选词的特征信息逐个与所述各个分类索引的分类特征信息相比较，以获取其分类特征信息与所述候选词的特征信息相似的分类索引。

15 例如，当预定拓扑结构为树状结构，并且预定遍历顺序为随机遍历时，随机获取尚未被遍历的分类索引，并将该分类索引的分类特征信息与候选词的特征信息相比较，以获取其分类特征信息与所述候选词的特征信息相似的分类索引。

20 又例如，当预定拓扑结构为树状结构，并且预定遍历顺序为从叶结点逐层向上遍历时，先获取作为各个叶结点的分类索引，将该层的分类索引的分类特征信息与候选词的特征信息相比较，当未能在叶结点中获得与所述候选词的特征信息相似的分类索引时，再获取各个叶结点上一层的结点的分类索引，并将该层的分类索引的分类特征信息与候选词的特征信息相比较，依次逐层往上，直至获得与所述候选词的特征信息相似的分类索引。

25 当所获得的分类索引包含底层分类索引时，第一分类确定装置将该底层分类索引作为所述候选词对应的分类索引。

具体地，第一分类确定装置判断所获得的分类索引是否为底层分类索引，并当所获得的分类索引包含底层分类索引时，第一分类确定装置将该底层分类索引作为所述候选词对应的分类索引。

优选地，根据本方案的词条生成装置中，所述第一确定装置 3 还

包括第三生成装置（图未示）和第二分类确定装置（图未示）。

当所获得的分类索引不包含底层索引节点时，第三生成装置基于其中最低级别的分类索引所对应的一个或多个分类相关网页以及所述候选词，来生成位于该最低级别的分类索引的下级分类索引。

5 具体地，第三生成装置基于其中最低级别的分类索引所对应的一个或多个分类相关网页以及所述候选词，来生成位于该最低级别的分类索引的下级分类索引的方式包括但不限于以下任一种：

1) 第三生成装置基于候选词生成属于由前述第一分类确定装置所获得的分类索引的下级分类索引的名称，并基于候选词所对应的搜索结果页面以及所获得的分类索引所对应的分类相关页面，确定与该下级分类索引相对应的分类相关网页。

15 2) 第三生成装置基于前述第一分类确定装置所获得的分类索引对应的一个或多个站点网页，在该一个或多个站点网页中查询并获取与候选词相关的至少一个网页，并确定与所该网页对应的中心词，将其作为前述第一分类确定装置获得的分类索引的下级分类索引的名称，并将该至少一个网页作为与该下级分类索引对应的分类相关网页。

接着，第二分类确定装置将所生成的底层分类索引作为与所述候选词对应的分类索引。

20 根据本实施例的方案，通过获取一个或多个网站的网站导航信息来建立多级分类索引，从而使得词条的分类索引体系与实际使用中的体系相近，有利于更加全面的挖掘专业网站的内容信息，并且由于同时还可利用这些网站的网页内容作为分类索引的分类相关网页，故能够为候选词生成能够有更加系统、完整的词条信息。

25 图8示意出了根据本发明的又一优选实施例的用于生成词条信息的词条生成装置的结构示意图。根据本实施例的词条生成装置包括第一获取装置1、第二获取装置2、第一确定装置3、第一生成装置4、第一网页获取装置10、第二特征确定装置11、第三分类确定装置12以及提供装置13。

其中，第一获取装置 1、第二获取装置 2、第一确定装置 3 以及第一生成装置 4 已在参照图 5 所示的实施例中予以详述，并以引用的方式包含于此，不再赘述。

第一网页获取装置 10 获取候选网站的一个或多个网页。

5 其中，第一网页获取装置 10 确定候选网站的方式包括但不限于以下任一种：

1) 第一网页获取装置 10 获取人工指定的网站作为候选网站；

2) 第一网页获取装置 10 将抓取到的网页页面与多级分类索引信息中的各个分类索引所对应的网页进行比较，以获得站点网页与所述各
10 个分类索引所对应的网页相似的网站。

接着，第二特征确定装置 11 根据所述候选网站的一个或多个网页，确定该候选网站的站点特征信息。

其中，第二特征确定装置 11 根据所述候选网站的一个或多个网页，确定该候选网站的站点特征信息的方式与前述参照图 5 所示实施例中第二确定装置根据所述一个或多个搜索结果网页，来确定与所述候
15 选词对应的特征信息的方式相同或相似，在此不再赘述。

接着，第三分类确定装置 12 将所述候选网站的站点特征信息与各个分类索引的分类特征信息进行比较，以确定与该候选网站对应的一
个或多个分类索引。

20 其中，第三分类确定装置 12 将所述候选网站的站点特征信息与各个分类索引的分类特征信息进行比较，以确定与该候选网站对应的一
个或多个分类索引的方式与前述参照图 7 所示实施例中比较确定装置将
所述候选词的特征信息与所述各个分类索引的分类特征信息相比较，
以获取其分类特征信息与所述候选词的特征信息相似的分类索引的方
25 式相同或相似，在此不再赘述。

接着，提供装置 13 向该候选网站对应的候选用户提供该一个或多个分类索引分别对应的一个或多个候选词。

作为本实施力的优选方案，根据本实施例的词条生成装置还包括第二网页获取装置（图未示）、第一更新装置（图未示）以及第一更新装

置（图未示）。

第二网页获取装置根据与所述候选网站对应的一个或多个分类索引，获取所述候选网站中与该一个或多个分类索引分别对应的一个或多个候选网页。

5 其中，所述第二网页获取装置根据与所述候选网站对应的一个或多个分类索引，获取所述候选网站中与该一个或多个分类索引分别对应的一个或多个候选网页的方式包括但不限于以下任一种：

1) 第二网页获取装置获取该一个或多个分类索引的分类相关网页，将所获得的分类相关网页与所述候选网站的站点网页进行比较，以获
10 得与所述分类相关网页相似的一个或多个站点网页，并将其作为与该分类相关网页所对应的分类索引的候选网页。

2) 第二网页获取装置根据该一个或多个分类索引的分类特征信息，由候选网站中获取分别与该一个或多个分类索引的分类特征信息相似的一个或多个候选网页。

15 接着，第一更新装置基于与各个分类索引对应的、所述候选网站中的一个或多个候选网页，确定或更新与该各个分类索引对应的分类相关网页。

具体地，第一更新装置将所确定的候选网页作为与分类索引对应的分类相关网页添加至与各个分类索引对应的分类相关网页库中。

20 第一更新装置基于所述更新后的与各个分类索引对应的分类相关网页，更新各个分类索引所对应的候选词的词条信息。

具体地，第一更新装置对属于个各个分类索引的一个或多个候选词，分别采用更新后的该分类索引的分类相关网页来更新各个候选词的词条内容。

25 其中，第一更新装置采用更新后的该分类索引的分类相关网页来更新各个候选词的词条内容的方式与前述参照图5所示实施例中第一生成装置根据与所述分类索引信息对应的至少一个分类相关网页，来确定与所述候选词对应的词条信息的方式相同或相似，此处不再赘述。

根据本实施例的方案，通过采用候选网站的内容来自动更新词条信息，使得词条内容能够尽快得到更新，并且提高了更新效率。

本发明的软件程序可以通过处理器执行以实现上文所述步骤或功能。同样地，本发明的软件程序（包括相关的数据结构）可以被存
5 储到计算机可读记录介质中，例如，RAM 存储器，磁或光驱动器或软磁盘及类似设备。另外，本发明的一些步骤或功能可采用硬件来实现，例如，作为与处理器配合从而执行各个功能或步骤的电路。

另外，本发明的一部分可被应用为计算机程序产品，例如计算机程序指令，当其被计算机执行时，通过该计算机的操作，可以调用或
10 提供根据本发明的方法和/或技术方案。而调用本发明的方法的程序指令，可能被存储在固定的或可移动的记录介质中，和/或通过广播或其他信号承载媒体中的数据流而被传输，和/或被存储在根据所述程序指令运行的计算机设备的工作存储器中。在此，根据本发明的一个实施例包括一个装置，该装置包括用于存储计算机程序指令的存储器和用
15 于执行程序指令的处理器，其中，当该计算机程序指令被该处理器执行时，触发该装置运行基于前述根据本发明的多个实施例的方法和/或技术方案。

对于本领域技术人员而言，显然本发明不限于上述示范性实施例的细节，而且在不背离本发明的精神或基本特征的情况下，能够以其
20 他的具体形式实现本发明。因此，无论从哪一点来看，均应将实施例看作是示范性的，而且是非限制性的，本发明的范围由所附权利要求而不是上述说明限定，因此旨在将落在权利要求的等同要件的含义和范围内的所有变化涵括在本发明内。不应将权利要求中的任何附图标
25 记视为限制所涉及的权利要求。此外，显然“包括”一词不排除其他单元或步骤，单数不排除复数。系统权利要求中陈述的多个单元或装置也可以由一个单元或装置通过软件或者硬件来实现。第一，第二等词语用来表示名称，而并不表示任何特定的顺序。

权 利 要 求 书

1. 一种用于生成词条信息的方法，其中，所述方法包括以下步骤：

- 5 a 获取候选词；
- b 基于所述候选词进行搜索，以获取所述候选词的特征信息；
- c 根据所述候选词的特征信息，在多级分类索引信息中确定与所述候选词对应的分类索引；其中，所述分类索引对应至少一个分类相关网页；
- 10 d 根据与所述分类索引信息对应的至少一个分类相关网页，来确定与所述候选词对应的词条信息。

2. 根据权利要求 1 所述的方法，其中，所述步骤 b 包括以下步骤：

- b1 通过第一预定搜索引擎，基于所述候选词执行搜索，以获取与所述候选词对应的一个或多个搜索结果网页；
- 15 b2 根据所述一个或多个搜索结果网页，来确定与所述候选词对应的特征信息。

3. 根据权利要求 2 所述的方法，其中，所述步骤 b2 包括以下步骤：

- 获取所述一个或多个搜索结果网页中所包含的至少一个关键词；
- 20 - 获取所述至少一个关键词中的各个关键词的权重信息；
- 基于所获得的各个关键词及其相应的权重信息，来确定与所述候选词对应的特征信息。

4. 根据权利要求 2 所述的方法，其中，所述步骤 b2 包括以下步骤：

- 25 - 通过预定主题确定模型，根据所述一个或多个搜索结果网页中的各个网页的网页内容，来确定与所述一个或多个搜索结果网页对应的主题相关信息；

 - 基于所确定主题相关信息来确定与所述候选词对应的特征信息。

5. 根据权利要求 1 至 3 中任一项所述的方法，其中，所述方法还包

括以下步骤:

x 获取与所述候选词对应的一项或多项网络发布信息;

y 根据所获得的一项或多项网络发布信息来确定所述候选词的重要度信息;

5 其中, 所述方法还包括以下步骤:

- 判断所述候选词的重要度信息是否满足预定重要度条件;

其中, 所述步骤 b 包括以下步骤:

- 当所述候选词的重要度信息满足预定重要度条件时, 获取所述候选词的特征信息。

10 6. 根据权利要求 5 所述的方法, 其中, 所述步骤 x 包括以下步骤:

- 对所述候选词进行切词以获取多个子候选词;

- 通过第二预定搜索引擎, 基于各个子候选词执行搜索以获取与各个子候选词对应的网络发布信息;

其中, 所述步骤 y 包括以下步骤:

15 - 基于各个子候选词对应的网络发布信息确定该子候选词的子重要度信息;

- 基于各个子候选词的子重要度信息确定所述候选词的重要度信息。

20 7. 根据权利要求 1 至 6 中任一项所述的方法, 其中, 所述方法还包括以下步骤:

- 获取一个或多个网站的网页导航信息;

- 根据所获得的一个或多个网页导航信息, 来生成多级分类索引信息, 其中, 所述多级分类索引中的各个分类索引按照预定拓扑结构相互关联。

25 8. 根据权利要求 7 所述的方法, 其中, 所述方法包括以下步骤:

- 基于与所述多级分类索引信息对应的所述一个或多个网站的网页导航信息, 获取与该多级分类索引信息中的各个分类索引分别对应的网页;

- 基于与所述各个分类索引相对应的网页来确定与该各个分类索引

分别对应的分类特征信息;

其中, 所述步骤 c 包括以下步骤:

- 基于所述候选词的特征信息以及各个分类索引的分类特征信息, 确定与所述候选词对应的分类索引。

5 9. 根据权利要求 8 所述的方法, 其中, 所述预定拓扑结构包括多级的拓扑结构, 其中相邻两级的分类索引之间为隶属关系, 其中, 所述步骤 c 包括以下步骤:

- 将所述候选词的特征信息与所述各个分类索引的分类特征信息相比较, 以获取其分类特征信息与所述候选词的特征信息相似的分类索引;

- 当所获得的分类索引包含底层分类索引时, 将该底层分类索引作为所述候选词对应的分类索引。

10 10. 根据权利要求 9 所述的方法, 其中, 所述步骤 c 还包括以下步骤:

15 - 当所获得的分类索引不包含底层索引节点时, 基于其中最低级别的分类索引所对应的一个或多个分类相关网页以及所述候选词, 来生成位于该最低级别的分类索引的下级分类索引;

- 将所生成的底层分类索引作为与所述候选词对应的分类索引。

20 11. 根据权利要求 1 至 9 中任一项所述的方法, 其中, 所述方法还包括以下步骤:

- 获取候选网站的一个或多个网页;

- 根据所述候选网站的一个或多个网页, 确定该候选网站的站点特征信息;

25 - 将所述候选网站的站点特征信息与各个分类索引的分类特征信息进行比较, 以确定与该候选网站对应的一个或多个分类索引;

- 向该候选网站对应的候选用户提供该一个或多个分类索引分别对应的一个或多个候选词。

12. 根据权利要求 11 所述的方法, 其中, 所述方法还包括以下步骤:

- 根据与所述候选网站对应的一个或多个分类索引，获取所述候选网站中与该一个或多个分类索引分别对应的一个或多个候选网页；

- 基于与各个分类索引对应的、所述候选网站中的一个或多个候选网页，确定或更新与该各个分类索引对应的分类相关网页；

5 - 基于所述更新后的与各个分类索引对应的分类相关网页，更新各个分类索引所对应的候选词的词条信息。

13. 一种用于生成词条信息的词条生成装置，其中，所述词条生成装置包括：

第一获取装置，用于获取候选词；

10 第二获取装置，用于基于所述候选词进行搜索，以获取所述候选词的特征信息；

第一确定装置，用于根据所述候选词的特征信息，在多级分类索引信息中确定与所述候选词对应的分类索引；其中，所述分类索引对应至少一个分类相关网页；

15 第一生成装置，用于根据与所述分类索引信息对应的至少一个分类相关网页，来确定与所述候选词对应的词条信息。

14. 根据权利要求 13 所述的词条生成装置，其中，所述第二获取装置包括：

20 第一搜索装置，用于通过第一预定搜索引擎，基于所述候选词执行搜索，以获取与所述候选词对应的一个或多个搜索结果网页；

第二确定装置，用于根据所述一个或多个搜索结果网页，来确定与所述候选词对应的特征信息。

15. 根据权利要求 14 所述的词条生成装置，其中，所述第二确定装置包括：

25 关键词获取装置，用于获取所述一个或多个搜索结果网页中所包含的至少一个关键词；

权重获取装置，用于获取所述至少一个关键词中的各个关键词的权重信息；

第一子确定装置，用于基于所获得的各个关键词及其相应的权重信

息，来确定与所述候选词对应的特征信息。

16. 根据权利要求 14 所述的词条生成装置，其中，所述第二确定装置包括以下步骤：

模型确定装置，用于通过预定主题确定模型，根据所述一个或多个搜索结果网页中的各个网页的网页内容，来确定与所述一个或多个搜索结果网页对应的主题相关信息；

第二子确定装置，用于基于所确定的主题相关信息来确定与所述候选词对应的特征信息。

17. 根据权利要求 13 至 16 中任一项所述的词条生成装置，其中，所述词条生成装置还包括：

第三获取装置，用于获取与所述候选词对应的一项或多项网络发布信息；

第三确定装置，用于根据所获得的一项或多项网络发布信息来确定所述候选词的重要度信息；

15 判断装置，用于判断所述候选词的重要度信息是否满足预定重要度条件；

其中，所述第二获取装置用于：

- 当所述候选词的重要度信息满足预定重要度条件时，获取所述候选词的特征信息。

20 18. 根据权利要求 17 所述的词条生成装置，其中，所述第三获取装置包括：

第一子获取装置，用于对所述候选词进行切词以获取多个子候选词；

25 第二搜索装置，用于通过第二预定搜索引擎，基于各个子候选词执行搜索以获取与各个子候选词对应的网络发布信息；

其中，所述第三确定装置包括：

第三子确定装置，用于基于各个子候选词对应的网络发布信息确定该子候选词的子重要度信息；

第四子确定装置，用于基于各个子候选词的子重要度信息确定所述

候选词的重要度信息。

19. 根据权利要求 13 至 18 中任一项所述的词条生成装置，其中，所述词条生成装置还包括：

导航获取装置，用于获取一个或多个网站的网页导航信息；

5 第二生成装置，用于根据所获得的一个或多个网页导航信息，来生成多级分类索引信息，其中，所述多级分类索引中的各个分类索引按照预定拓扑结构相互关联。

20. 根据权利要求 19 所述的词条生成装置，其中，所述词条生成装置包括以下步骤：

10 第四获取装置，用于基于与所述多级分类索引信息对应的所述一个或多个网站的网页导航信息，获取与该多级分类索引信息中的各个分类索引分别对应的网页；

第一特征确定装置，用于基于与所述各个分类索引相对应的网页来确定与该各个分类索引分别对应的分类特征信息；

15 其中，所述第一确定装置用于：

- 基于所述候选词的特征信息以及各个分类索引的分类特征信息，确定与所述候选词对应的分类索引。

21. 根据权利要求 20 所述的词条生成装置，其中，所述预定拓扑结构包括多级的拓扑结构，其中相邻两级的分类索引之间为隶属关系，其中，所述第一确定装置包括：

比较获取装置，用于将所述候选词的特征信息与所述各个分类索引的分类特征信息相比较，以获取其分类特征信息与所述候选词的特征信息相似的分类索引；

25 第一分类确定装置，用于当所获得的分类索引包含底层分类索引时，将该底层分类索引作为所述候选词对应的分类索引。

22. 根据权利要求 21 所述的词条生成装置，其中，所述第一确定装置还包括：

第三生成装置，用于当所获得的分类索引不包含底层索引节点时，基于其中最低级别的分类索引所对应的一个或多个分类相关网页以及

所述候选词，来生成位于该最低级别的分类索引的下级分类索引；

第二分类确定装置，用于将所生成的底层分类索引作为与所述候选词对应的分类索引。

23. 根据权利要求 13 至权利要求 21 所述的词条生成装置，其中，
5 所述词条生成装置还包括：

第一网页获取装置，用于获取候选网站的一个或多个网页；

第二特征确定装置，用于根据所述候选网站的一个或多个网页，确定该候选网站的站点特征信息；

10 第三分类确定装置，用于将所述候选网站的站点特征信息与各个分类索引的分类特征信息进行比较，以确定与该候选网站对应的一个或多个分类索引；

提供装置，用于向该候选网站对应的候选用户提供该一个或多个分类索引分别对应的一个或多个候选词。

24. 根据权利要求 23 所述的词条生成装置，其中，所述词条生成装
15 置还包括：

第二网页获取装置，用于根据与所述候选网站对应的一个或多个分类索引，获取所述候选网站中与该一个或多个分类索引分别对应的一个或多个候选网页；

20 第一更新装置，用于基于与各个分类索引对应的、所述候选网站中的一个或多个候选网页，确定与该各个分类索引对应的分类相关网页；

第一更新装置，用于基于所述更新后的与各个分类索引对应的分类相关网页，更新各个分类索引所对应的候选词的词条信息。

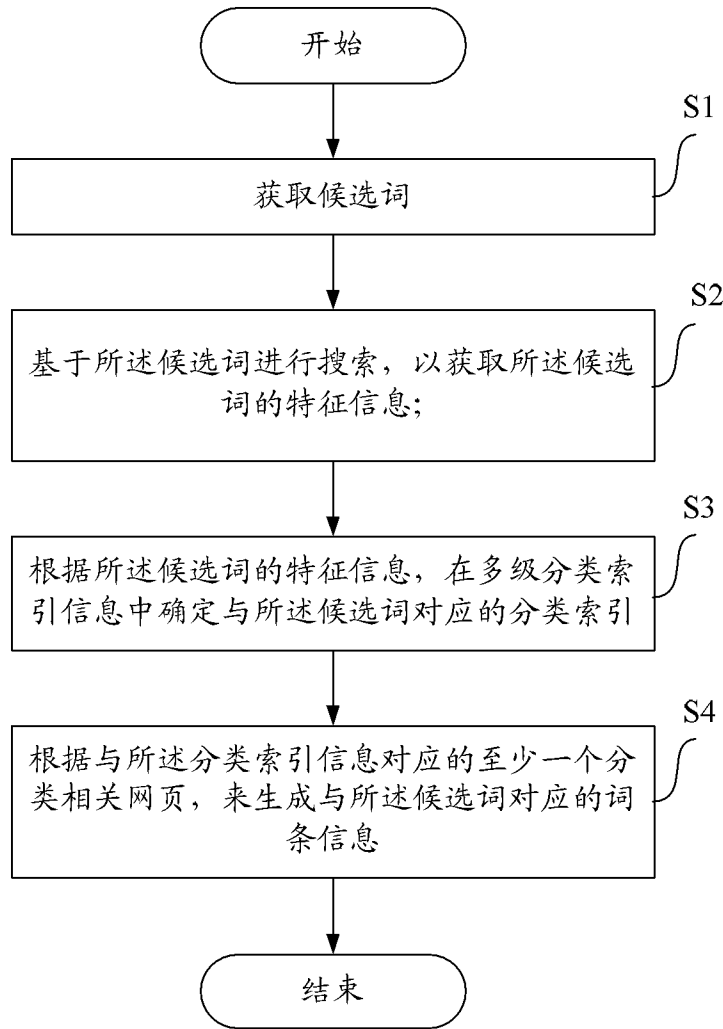


图 1

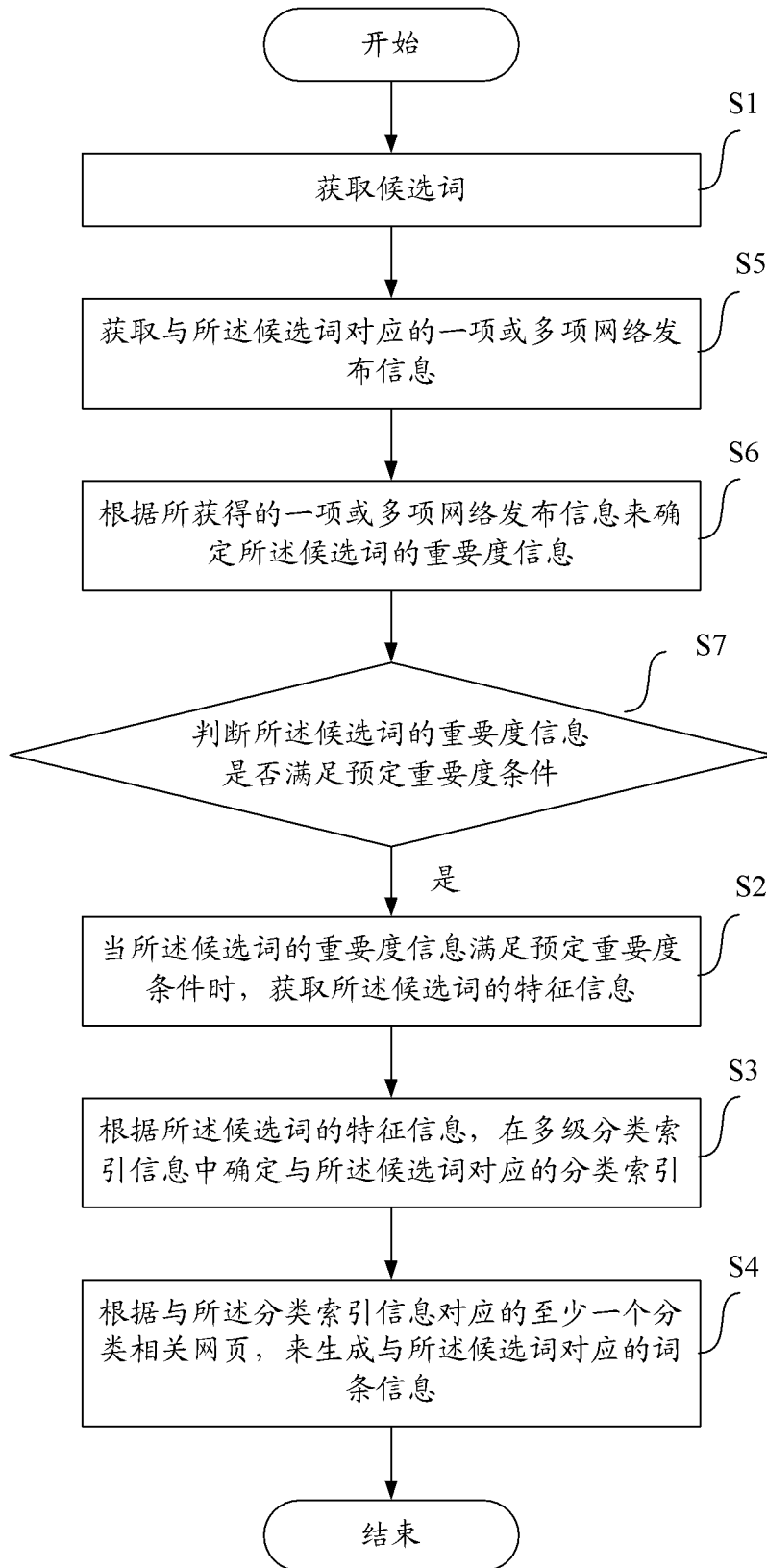


图 2

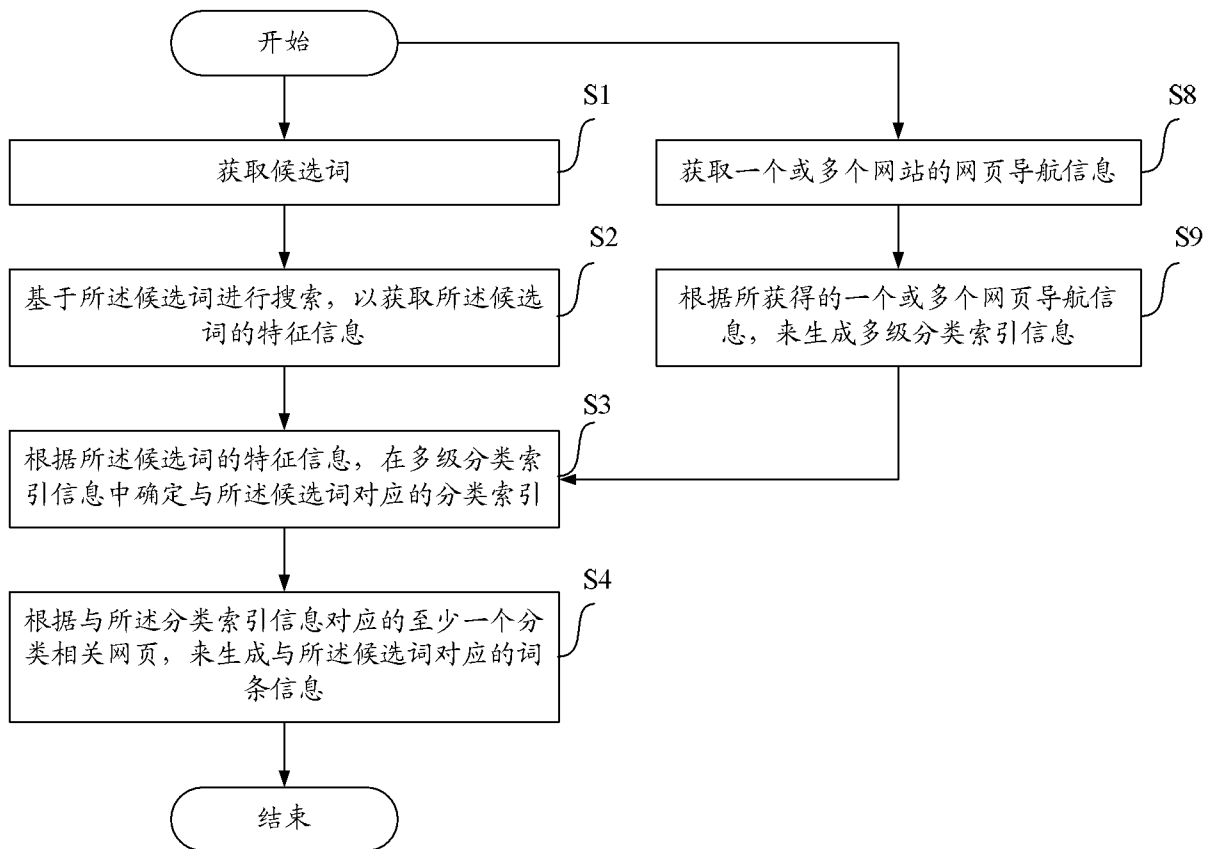


图 3

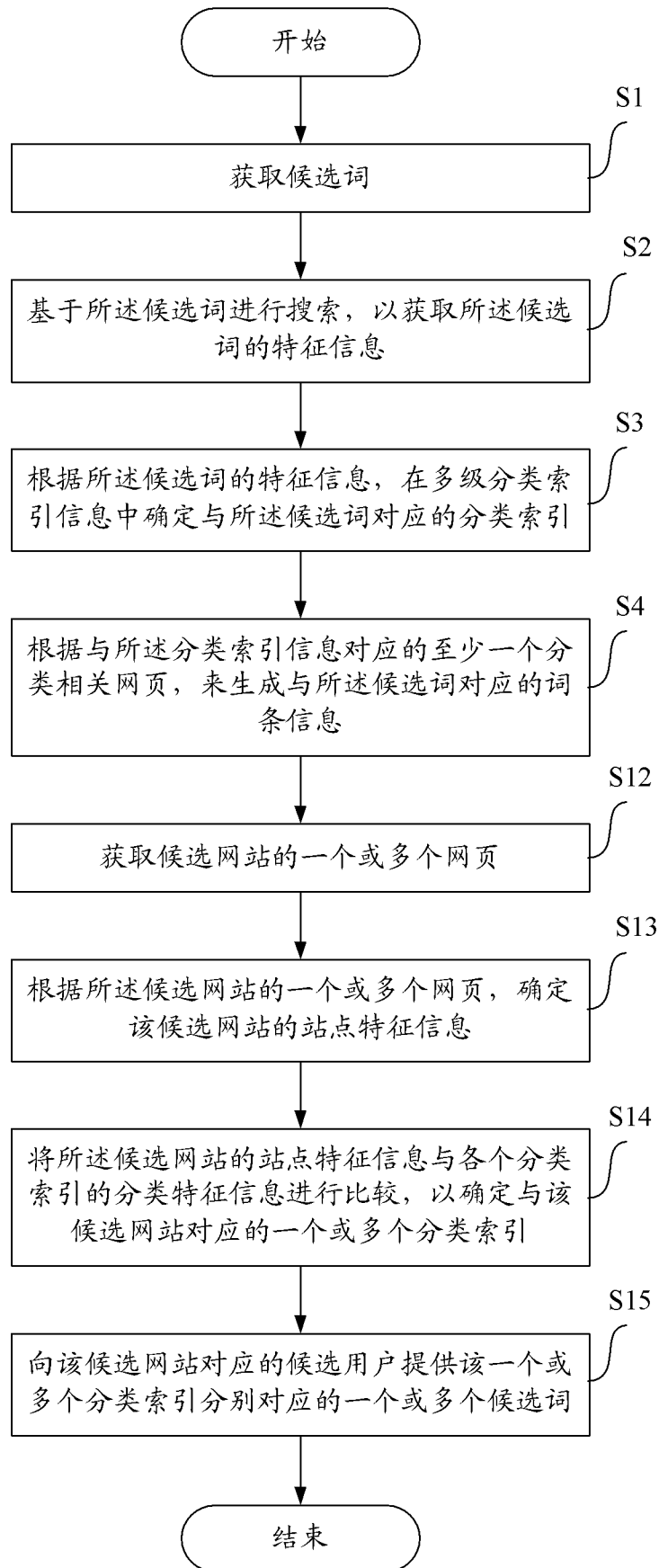


图 4

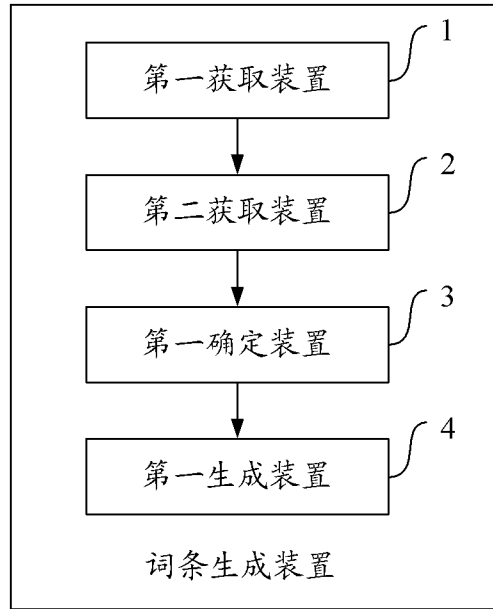


图 5

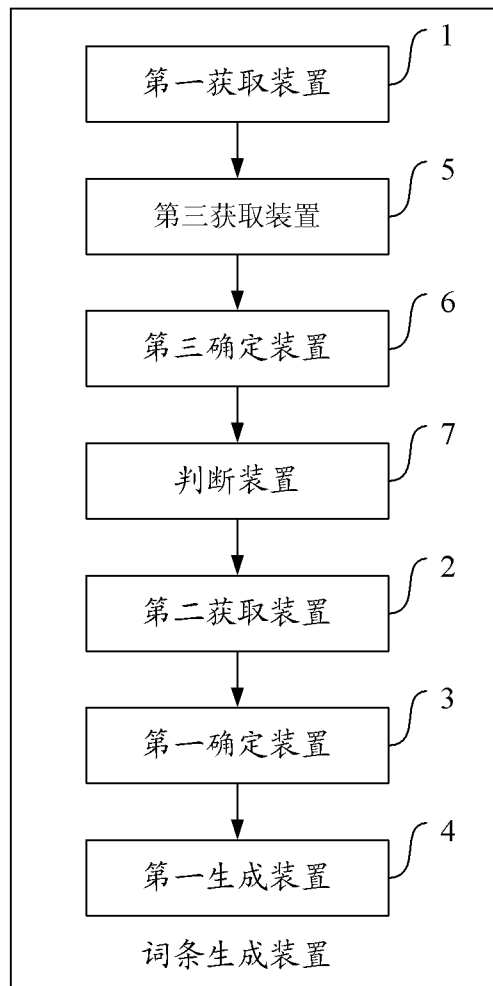


图 6

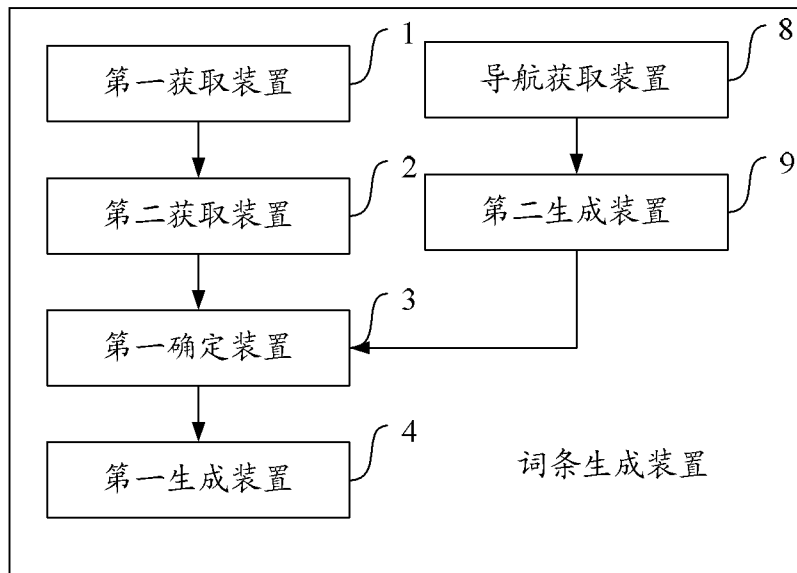


图 7

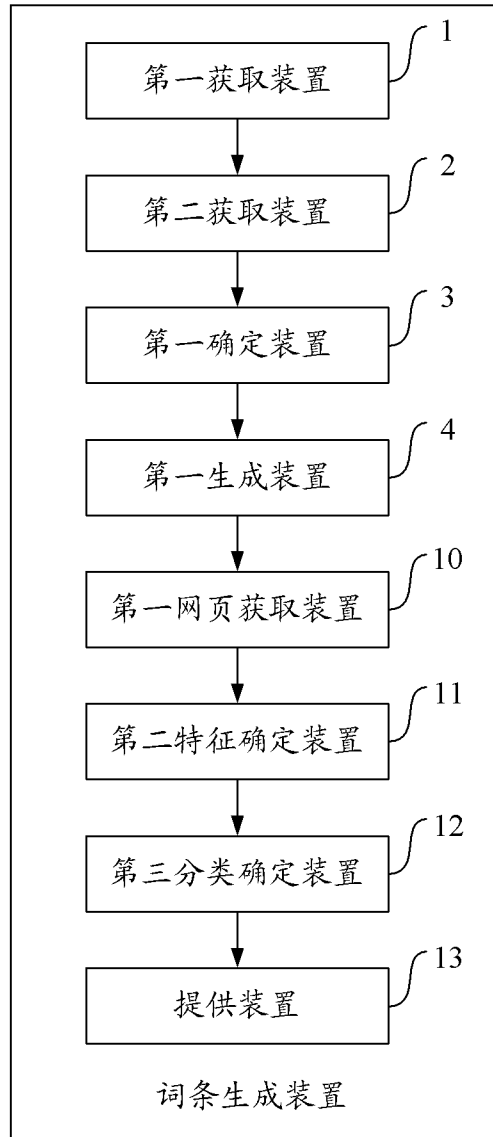


图 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2014/079220

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)

EPODOC, WPI, CNPAT, CNKI, IEEE: entry, generate, add, new, word, candidate, key, feature, query, catalog, weight, classify, search, index, divide

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 101986310 A (INVENTEC BESTA XI AN CO., LTD.) 16 March 2011 (16.03.2011) description, paragraphs [0032] to [0056]	1-24
Y	CN 101957831 A (LIU, Erzong) 26 January 2011 (26.01.2011) description, paragraphs [0084] to [0177], and figure 5	1-24
A	CN 101566995 A (BEIJING SOGOU TECH DEV CO.) 28 October 2009 (28.10.2009) the whole document	1-24
A	CN 101251854 A (SHENZHEN INST OF ADV TECH CAS) 27 August 2008 (27.08.2008) the whole document	1-24
A	US 2009094020 A1 (FUJITSU LIMITED) 09 April 2009 (09.04.2009) the whole document	1-24

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>
---	---

<p>Date of the actual completion of the international search</p> <p style="text-align: center;">15 August 2014</p>	<p>Date of mailing of the international search report</p> <p style="text-align: center;">01 September 2014</p>
<p>Name and mailing address of the ISA</p> <p>State Intellectual Property Office of the P. R. China</p> <p>No. 6, Xitucheng Road, Jimenqiao</p> <p>Haidian District, Beijing 100088, China</p> <p>Facsimile No. (86-10) 62019451</p>	<p>Authorized officer</p> <p style="text-align: center;">JIANG, Xiaoqing</p> <p>Telephone No. (86-10) 61648093</p>

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2014/079220

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2012000335 A1 (BAIDU ONLINE NETWORK TECHNOLOGY BEIJING CO., LTD. et al.) 05 January 2012 (05.01.2012) the whole document	1-24

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2014/079220

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 101986310 A	16 March 2011	None	
CN 101957831 A	26 January 2011	WO 2011006412 A1	20 January 2011
		US 2012271810 A1	25 October 2012
		CN 101694666 A	14 April 2010
CN 101566995 A	28 October 2009	None	
CN 101251854 A	27 August 2008	None	
US 2009094020 A1	09 April 2009	JP 2009093649 A	30 April 2009
		EP 2045740 A1	08 April 2009
		CN 101408887 A	15 April 2009
WO 2012000335 A1	05 January 2012	CN 102314439 A	11 January 2012

<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>EPODOC, WPI, CNPAT, CNKI, IEEE: 词条, 生成, 增加, 新词, 候选词, 关键词, 特征, 查询, 索引, 目录, 分类, 权重, 切词, entry, generate, add, new, word, candidate, key, feature, query, search, index, catalog, classify, weight, divide</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>Y</td> <td>CN 101986310 A (无敌科技西安有限公司) 2011年 3月 16日 (2011 - 03 - 16) 说明书第0032-0056段</td> <td>1-24</td> </tr> <tr> <td>Y</td> <td>CN 101957831 A (刘二中) 2011年 1月 26日 (2011 - 01 - 26) 说明书0084-0177段及附图5</td> <td>1-24</td> </tr> <tr> <td>A</td> <td>CN 101566995 A (北京搜狗科技发展有限公司) 2009年 10月 28日 (2009 - 10 - 28) 全文</td> <td>1-24</td> </tr> <tr> <td>A</td> <td>CN 101251854 A (深圳先进技术研究院) 2008年 8月 27日 (2008 - 08 - 27) 全文</td> <td>1-24</td> </tr> <tr> <td>A</td> <td>US 2009094020 A1 (FUJITSU LIMITED) 2009年 4月 09日 (2009 - 04 - 09) 全文</td> <td>1-24</td> </tr> <tr> <td>A</td> <td>WO 2012000335 A1 (百度在线网络技术北京有限公司等) 2012年 1月 05日 (2012 - 01 - 05) 全文</td> <td>1-24</td> </tr> </tbody> </table> <p><input type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。</p> <p>* 引用文件的具体类型:</p> <table border="0"> <tr> <td>“A” 认为不特别相关的表示了现有技术一般状态的文件</td> <td>“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</td> </tr> <tr> <td>“E” 在国际申请日的当天或之后公布的在先申请或专利</td> <td>“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</td> </tr> <tr> <td>“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)</td> <td>“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</td> </tr> <tr> <td>“O” 涉及口头公开、使用、展览或其他方式公开的文件</td> <td>“&” 同族专利的文件</td> </tr> <tr> <td>“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</td> <td></td> </tr> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	Y	CN 101986310 A (无敌科技西安有限公司) 2011年 3月 16日 (2011 - 03 - 16) 说明书第0032-0056段	1-24	Y	CN 101957831 A (刘二中) 2011年 1月 26日 (2011 - 01 - 26) 说明书0084-0177段及附图5	1-24	A	CN 101566995 A (北京搜狗科技发展有限公司) 2009年 10月 28日 (2009 - 10 - 28) 全文	1-24	A	CN 101251854 A (深圳先进技术研究院) 2008年 8月 27日 (2008 - 08 - 27) 全文	1-24	A	US 2009094020 A1 (FUJITSU LIMITED) 2009年 4月 09日 (2009 - 04 - 09) 全文	1-24	A	WO 2012000335 A1 (百度在线网络技术北京有限公司等) 2012年 1月 05日 (2012 - 01 - 05) 全文	1-24	“A” 认为不特别相关的表示了现有技术一般状态的文件	“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	“E” 在国际申请日的当天或之后公布的在先申请或专利	“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	“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)	“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	“O” 涉及口头公开、使用、展览或其他方式公开的文件	“&” 同族专利的文件	“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	
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																															
Y	CN 101986310 A (无敌科技西安有限公司) 2011年 3月 16日 (2011 - 03 - 16) 说明书第0032-0056段	1-24																															
Y	CN 101957831 A (刘二中) 2011年 1月 26日 (2011 - 01 - 26) 说明书0084-0177段及附图5	1-24																															
A	CN 101566995 A (北京搜狗科技发展有限公司) 2009年 10月 28日 (2009 - 10 - 28) 全文	1-24																															
A	CN 101251854 A (深圳先进技术研究院) 2008年 8月 27日 (2008 - 08 - 27) 全文	1-24																															
A	US 2009094020 A1 (FUJITSU LIMITED) 2009年 4月 09日 (2009 - 04 - 09) 全文	1-24																															
A	WO 2012000335 A1 (百度在线网络技术北京有限公司等) 2012年 1月 05日 (2012 - 01 - 05) 全文	1-24																															
“A” 认为不特别相关的表示了现有技术一般状态的文件	“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																																
“E” 在国际申请日的当天或之后公布的在先申请或专利	“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																																
“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)	“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																																
“O” 涉及口头公开、使用、展览或其他方式公开的文件	“&” 同族专利的文件																																
“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																																	
<p>国际检索实际完成的日期</p> <p>2014年 8月 15日</p>	<p>国际检索报告邮寄日期</p> <p>2014年 9月 01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) 北京市海淀区蓟门桥西土城路6号 100088 中国</p> <p>传真号 (86-10)62019451</p>	<p>授权官员</p> <p>姜晓庆</p> <p>电话号码 (86-10)61648093</p>																																

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2014/079220

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	101986310	A	2011年 3月 16日	无			
CN	101957831	A	2011年 1月 26日	WO	2011006412	A1	2011年 1月 20日
				US	2012271810	A1	2012年 10月 25日
				CN	101694666	A	2010年 4月 14日
CN	101566995	A	2009年 10月 28日	无			
CN	101251854	A	2008年 8月 27日	无			
US	2009094020	A1	2009年 4月 09日	JP	2009093649	A	2009年 4月 30日
				EP	2045740	A1	2009年 4月 08日
				CN	101408887	A	2009年 4月 15日
WO	2012000335	A1	2012年 1月 05日	CN	102314439	A	2012年 1月 11日