

(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 1 月 5 日 (05.01.2017)



(10) 国际公布号
WO 2017/000402 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2015/090703
- (22) 国际申请日: 2015 年 9 月 25 日 (25.09.2015)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201510385768.X 2015 年 6 月 30 日 (30.06.2015) 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 号百度大厦三层, Beijing 100085 (CN)。
- (72) 发明人: 裘皓萍 (QIU, Haoping); 中国北京市海淀区
上地十街 10 号百度大厦三层, Beijing 100085 (CN)。
徐云峰 (XU, Yunfeng); 中国北京市海淀区上地十
街 10 号百度大厦三层, Beijing 100085 (CN)。
陈炜于 (CHEN, Weiyu); 中国北京市海淀区上地十街 10
号百度大厦三层, Beijing 100085 (CN)。

(74) 代理人: 北京英赛嘉华知识产权代理有限公司
(INSIGHT INTELLECTUAL PROPERTY LIM-
ITED); 中国北京市海淀区知春路甲 48 号盈都大厦
A 座 19A, Beijing 100098 (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,

[见续页]

(54) Title: PAGE GENERATION METHOD AND DEVICE

(54) 发明名称: 网页生成方法和装置

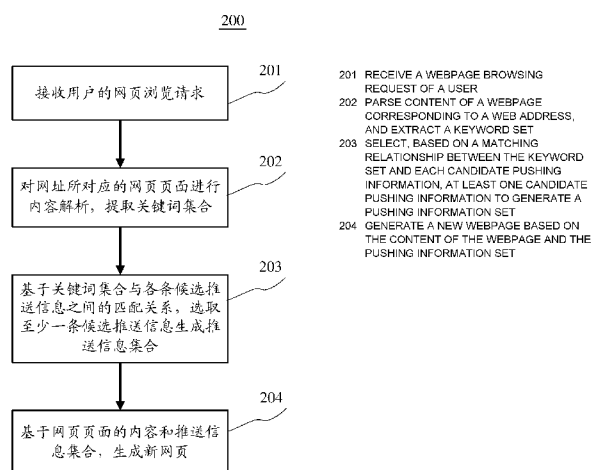


图 2

(57) Abstract: Disclosed are a page generation method and device. A specific embodiment of the method comprises: receiving a webpage browsing request of a user, wherein the webpage browsing request comprises a web address; parsing content of a webpage corresponding to the web address, and extracting a keyword set; selecting, based on a matching relationship between the keyword set and each candidate pushing information, at least one candidate pushing information to generate a pushing information set; and generating a new webpage based on the content of the webpage and the pushing information set. The embodiment realizes targeted information pushing.

(57) 摘要: 本申请公开了网页生成方法和装置。所述方法的一具体实施方式包括: 接收用户的网页浏览请求, 其中, 所述网页浏览请求包括网址; 对所述网址所对应的网页页面进行内容解析, 提取关键词集合; 基于所述关键词集合与各候选推送信息之间的匹配关系, 选取至少一条候选推送信息生成推送信息集合; 基于所述网页页面的内容和所述推送信息集合, 生成新网页。该实施方式实现了富于针对性的信息推送。

WO 2017/000402 A1

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))。

网页生成方法和装置

相关申请的交叉引用

本申请要求百度在线网络技术（北京）有限公司于 2015 年 6 月 30 日提交的、发明名称为“网页生成方法和装置”的、中国专利申请号“201510385768.X”的优先权，其全部内容作为整体并入本申请中。

技术领域

本申请涉及计算机技术领域，具体涉及互联网技术领域，尤其涉及网页生成方法和和装置。

背景技术

信息推送，又称为“网络广播”，是通过一定的技术标准或协议，在互联网上通过推送用户需要的信息来减少信息过载的一项技术。信息推送技术通过主动推送信息给用户，可以减少用户在网络上搜索所花的时间。

然而，现有的信息推送方式通常是在网页上直接加载各种推送信息，这些推送信息与所在网页的内容有明显的差异，从而，存在着网页内容相关数据利用不足，信息推送缺乏针对性的问题。

发明内容

本申请的目的在于提出一种改进的网页生成方法和装置，来解决以上背景技术部分提到的技术问题。

第一方面，本申请提供了一种网页生成方法，所述方法包括：接收用户的网页浏览请求，其中，所述网页浏览请求包括网址；对所述网址所对应的网页页面进行内容解析，提取关键词集合；基于所述关键词集合与各条候选推送信息之间的匹配关系，选取至少一条候选推送信息生成推送信息集合；基于所述网页页面的内容和所述推送信息

集合，生成新网页。

在一些实施例中，所述对所述网址所对应的网页页面进行内容解析提取关键词集合，包括：对所述网址所对应的网页页面的内容进行统计分析和/或语义分析，提取至少一个关键词；基于所述至少一个关键词，生成关键词集合。

在一些实施例中，所述基于所述至少一个关键词，生成关键词集合，包括：对于所述至少一个关键词中的单个关键词，进行扩展以生成扩展关键词，其中，所述扩展关键词包括所述单个关键词和以下中的至少一项：所述单个关键词的同义词、所述单个关键词的近义词、所述单个关键词的关联词；基于所述扩展关键词，生成关键词集合。

在一些实施例中，所述基于所述关键词集合与各条候选推送信息之间的匹配关系，选取至少一条候选推送信息生成推送信息集合，包括：对各条候选推送信息进行内容解析，分别提取候选推送信息关键词集合；将所述关键词集合分别与各个候选推送信息关键词集合进行相似度计算；基于相似度计算的结果，选取至少一条候选推送信息生成推送信息集合。

在一些实施例中，所述基于相似度计算的结果，选取至少一条候选推送信息生成推送信息集合，包括：基于相似度计算的结果以及预设的推送信息的量值，选取所述量值个数的候选推送信息生成推送信息集合。

在一些实施例中，所述基于所述网页页面的内容和所述推送信息集合，生成新网页，包括：以所述推送信息集合中的推送信息与所述网页页面的内容中的相应关键词相关联的方式，生成新网页。

在一些实施例中，所述基于所述网页页面的内容和所述推送信息集合，生成新网页，包括：以所述推送信息集合中的推送信息与所述网页页面的内容分开设定的方式，生成新网页。

第二方面，本申请提供了一种网页生成装置，所述装置包括：接收单元，配置用于接收用户的网页浏览请求，其中，所述网页浏览请求包括网址；解析单元，配置用于对所述网址所对应的网页页面进行内容解析，提取关键词集合；信息选取单元，配置用于基于所述关键

词集合与各条候选推送信息之间的匹配关系，选取至少一条候选推送信息生成推送信息集合；生成单元，配置用于基于所述网页页面的内容和所述推送信息集合，生成新网页。

在一些实施例中，所述解析单元包括：分析模块，配置用于对所述网址所对应的网页页面的内容进行统计分析和/或语义分析，提取至少一个关键词；生成模块，配置用于基于所述至少一个关键词，生成关键词集合。

在一些实施例中，所述生成模块进一步配置用于：对于所述至少一个关键词中的单个关键词，进行扩展以生成扩展关键词，其中，所述扩展关键词包括所述单个关键词和以下中的至少一项：所述单个关键词的同义词、所述单个关键词的近义词、所述单个关键词的关联词；基于所述扩展关键词，生成关键词集合。

在一些实施例中，所述信息选取单元包括：解析模块，配置用于对各条候选推送信息进行内容解析，分别提取候选推送信息关键词集合；相似度计算模块，配置用于将所述关键词集合分别与各个候选推送信息关键词集合进行相似度计算；选取模块，配置用于基于相似度计算的结果，选取至少一条候选推送信息生成推送信息集合。

在一些实施例中，所述选取模块进一步配置用于：基于相似度计算的结果以及预设的推送信息的量值，选取所述量值个数的候选推送信息生成推送信息集合。

在一些实施例中，所述生成单元进一步配置用于：以所述推送信息集合中的推送信息与所述网页页面的内容中的相应关键词相关联的方式，生成新网页。

在一些实施例中，所述生成单元进一步配置用于：以所述推送信息集合中的推送信息与所述网页页面的内容分开设置的方式，生成新网页。

本申请提供的网页生成方法和装置，通过对用户请求的网址所对应的网页页面进行内容解析以便提取关键词集合，而后基于关键词集合与各条候选推送信息之间的匹配关系来选取推送信息，最后基于网页页面的内容和推送信息来生成新网页，从而有效利用了网页的内容

数据，实现了富于针对性的信息推送。

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，本申请的其它特征、目的和优点将会变得更明显：

图 1 是本申请可以应用于其中的示例性系统架构图；

图 2 是根据本申请的网页生成方法的一个实施例的流程图；

图 3 是根据本申请的网页生成方法的一个应用场景的示意图；

图 4 是根据本申请的网页生成方法的又一个实施例的流程图；

图 5 是根据本申请的网页生成装置的一个实施例的结构示意图；

图 6 是适于用来实现本申请实施例的终端设备或服务器的计算机系统的结构示意图。

具体实施方式

下面结合附图和实施例对本申请作进一步的详细说明。可以理解的是，此处所描述的具体实施例仅仅用于解释相关发明，而非对该发明的限定。另外还需要说明的是，为了便于描述，附图中仅示出了与有关发明相关的部分。

需要说明的是，在不冲突的情况下，本申请中的实施例及实施例中的特征可以相互组合。下面将参考附图并结合实施例来详细说明本申请。

图 1 示出了可以应用本申请的网页生成方法或网页生成装置的实施例的示例性系统架构 100。

如图 1 所示，系统架构 100 可以包括终端设备 101、102、103，网络 104 和服务器 105。网络 104 用以在终端设备 101、102、103 和服务器 105 之间提供通信链路的介质。网络 104 可以包括各种连接类型，例如有线、无线通信链路或者光纤电缆等等。

用户可以使用终端设备 101、102、103 通过网络 104 与服务器 105 交互，以接收或发送消息等。终端设备 101、102、103 上可以安装有各种通讯客户端应用，例如网页浏览器应用、购物类应用、搜索类应

用、即时通信工具、邮箱客户端、社交平台软件等。

终端设备 101、102、103 可以是具有显示屏并且支持网页浏览的各种电子设备，包括但不限于智能手机、平板电脑、电子书阅读器、MP3 播放器(Moving Picture Experts Group Audio Layer III, 动态影像专家压缩标准音频层面 3)、MP4(Moving Picture Experts Group Audio Layer IV, 动态影像专家压缩标准音频层面 4)播放器、膝上型便携计算机和台式计算机等等。

服务器 105 可以是提供各种服务的服务器，例如对终端设备 101、102、103 上显示的网页提供支持的后台网页服务器。后台网页服务器可以对接收到的网页页面请求等数据进行分析等处理，并将处理结果（例如网页页面数据）反馈给终端设备。

需要说明的是，本申请实施例所提供的网页生成方法一般由服务器 105 执行，相应地，网页生成装置一般设置于服务器 105 中。

应该理解，图 1 中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要，可以具有任意数目的终端设备、网络和服务服务器。

继续参考图 2，示出了根据本申请的网页生成方法的一个实施例的流程 200。所述的网页生成方法，包括以下步骤：

步骤 201，接收用户的网页浏览请求。

在本实施例中，网页生成方法运行于其上的电子设备（例如图 1 所示的服务器）可以通过有线连接方式或者无线连接方式从用户利用其进行网页浏览的终端接收网页浏览请求，其中，上述网页浏览请求包括了用户期望浏览的网页的地址，即网址。实践中，网址一般由统一资源定位符（Uniform Resource Locator，URL）来表示。需要指出的是，上述无线连接方式可以包括但不限于 3G/4G 连接、WiFi 连接、蓝牙连接、WiMAX 连接、Zigbee 连接、UWB（ultra wideband）连接、以及其他现在已知或将来开发的无线连接方式。

通常，用户利用终端上安装的网页浏览器来浏览网页，这时，用户可以通过直接输入网址或者点击网页浏览器中呈现的网页中的链接来向网页服务器发起网页浏览请求。在本实施例中，上述网页可以包括 html 格式、xhtml 格式、asp 格式、php 格式、jsp 格式、shtml 格式、

nsp 格式、xml 格式的网页或者其它未来将开发的格式的网页（只要这种格式的网页文件可以用浏览器打开并浏览其包含的图片、动画、文字等内容）。

步骤 202，对网址所对应的网页页面进行内容解析，提取关键词集合。

在本实施例中，基于步骤 201 中得到的网址，上述电子设备（例如如图 1 所示的服务器）可以首先获取上述网址所对应的网页页面；之后再利用各种分析手段对上述网页页面的内容进行分析，从而提取出一个或多个关键词。

在本实施例的一些可选的实现方式中，对上述网页页面的内容的分析方式可以是统计分析方式。例如，可以对上述内容中存在的各个词语的出现频率进行统计和排序，之后，再选取出现频率排序靠前的一个或多个词语作为待提取的关键词。

在本实施例的一些可选的实现方式中，对上述网页页面的内容的分析方式还可以是语义分析方式。作为示例，可以对网页页面的内容进行全切分方法等处理，把内容分割成词；再对所得到的词进行重要性计算（例如采用词频-逆向文件频率方法（Term Frequency-Inverse Document Frequency, TF-IDF）），基于重要性计算的结果来得到关键词。

利用全切分方法，可以首先切分出与语言词库匹配的所有可能的词，再运用统计语言模型确定最优的切分结果。以用户输入信息“南京市长江大桥”为例，可以首先进行语言词库匹配，找到匹配的所有词——南京，市，长江，大桥，南京市，长江大桥，市长，江大桥，江大，桥；这些词以词网格（word lattices）形式表示，接着基于词网格做路径搜索，再基于统计语言模型（例如 N-Gram 模型，）找到最优路径。如果结果显示“南京市 长江 大桥”的语言模型得分最高，则“南京市 长江 大桥”即为“南京市长江大桥”的最优切分。在这里所述的 N-Gram 模型是常用的一种语言模型，对中文而言，可以称之为汉语语言模型（CLM, Chinese Language Model）。该 N-Gram 模型基于这样一种假设，第 N 个词的出现只与前面 N-1 个词相关，而与其

它任何词都不相关，整句的概率就是各个词出现概率的乘积，而这些概率可以通过直接从语料中统计 N 个词同时出现的次数得到。

利用全切分方法将内容分割成词之后，可以采用词频-逆向文件频率方法对这些词进行重要性计算，再基于重要性来选取词作为关键词或者对这些词进行重要性评分。词频-逆向文件频率方法的主要思想是，如果某个词或短语在一篇文章中出现的频率（Term Frequency, TF）高，并且在其他文章中很少出现，则认为此词或者短语具有很好的类别区分能力，适合用来分类。而逆向文件频率（Inverse Document Frequency, IDF）主要是指，如果包含某个词或短语的文档越少，则 IDF 越大，则说明该词或短语具有很好的类别区分能力。由此，使用词频-逆向文件频率方法，可以计算某个词或短语在某篇文章里面的重要性。

需要说明的是，上述语义分析方式的各种方法是目前广泛研究和应用的公知技术，在此不再赘述。

步骤 203，基于关键词集合与各条候选推送信息之间的匹配关系，选取至少一条候选推送信息生成推送信息集合。

在本实施例中，网页生成方法运行于其上的电子设备上可以预先存储多条候选推送信息，这些候选推送信息可以用于与上述网页页面的内容相结合，以便作为一个整体在浏览器上呈现。

在本实施例中，上述电子设备可以将上述关键词集合与各条候选推送信息的内容分别进行逐一匹配，并根据每条候选推送信息的内容所包括的关键词的数目来确定该条候选推送信息与关键词集合的匹配关系。例如，如果某条候选推送信息的内容中包括了上述关键词集合中的所有关键词，则可以确定该条候选推送信息与关键词集合存在完全匹配关系；而如果某条候选推送信息的内容中包括了上述关键词集合中的部分关键词，则可以确定该条候选推送信息与关键词集合存在部分匹配关系；而如果某条候选推送信息的内容中不包括上述关键词集合中的任何关键词，则可以确定该条候选推送信息与关键词集合存在不匹配关系。根据匹配关系，上述电子设备可以从多条候选推送信息中选取至少一条候选推送信息，并由此生成推送信息集合。例如，

上述电子设备可以选取与关键词集合符合完全匹配关系的候选推送信息，作为待与上述网页页面的内容结合的推送信息。

步骤 204，基于网页页面的内容和推送信息集合，生成新网页。

在本实施例中，上述电子设备可以将上述网页页面的内容（即上述网址对应的页面的页面内容）与推送信息集合相结合作为新网页的内容，并由此生成新网页。

在本实施例的一些可选的实现方式中，上述电子设备可以以上述推送信息集合中的推送信息与上述网页页面的内容中的相应关键词相关联的方式，生成新网页。作为示例，对于网页页面的内容中的某个关键词而言，上述电子设备可以首先确定其在网页页面中的位置；接着，从推送信息集合中查找与该关键词相匹配的推送信息；最后，将查找到的推送信息与该关键词设置为关联显示，例如，如果用户的鼠标悬停在网页页面的上述关键词上时，将弹出相应的推送信息。

在本实施例的一些可选的实现方式中，上述电子设备可以以上述推送信息集合中的推送信息与上述网页页面的内容分开设置的方式，生成新网页。作为示例，上述电子设备可以将推送信息与上述网页页面的内容设置为在新网页的不同显示区域进行显示，这时，用户可以集中查看推送信息且在其观看上述网页页面的内容时不会被干扰。

继续参见图 3，图 3 是根据本实施例的网页生成方法的应用场景的一个示意图。在图 3 的应用场景中，用户首先发起一个知识类网页的浏览请求；之后，网页服务器可以后台获取上述网页页面的内容，并提取出关键词“自动驾驶”；然后，上述网页服务器从其预存储的候选推送信息中找到与关键词“自动驾驶”相匹配的一条或多条信息作为推送信息；最后，上述网页服务器可以采取推送信息与关键词“自动驾驶”相关联的方式，生成新网页。当用户浏览新网页时，如果用户点击了“自动驾驶”这个词条，就会如图 3 所示，弹出推送信息。

本申请的上述实施例提供的方法通过将网页页面的内容和推送信息相关联，实现了富于针对性的信息推送。

进一步参考图 4，其示出了网页生成方法的又一个实施例的流程

400。该网页生成方法的流程 400，包括以下步骤：

步骤 401，接收用户的网页浏览请求。

在本实施例中，网页生成方法运行于其上的电子设备（例如图 1 所示的服务器）可以通过有线连接方式或者无线连接方式从用户利用其进行网页浏览的终端接收网页浏览请求，其中，上述网页浏览请求包括了用户期望浏览的网页的地址，即网址。

步骤 402，对网址所对应的网页页面的内容进行统计分析和/或语义分析，提取至少一个关键词。

在本实施例中，对上述网页页面的内容的分析方式可以是统计分析方式。例如，可以对上述内容中存在的各个词语的出现频率进行统计和排序，之后，再选取出现频率排序靠前的一个或多个词语作为待提取的关键词。或者，对上述网页页面的内容的分析方式还可以是语义分析方式。作为示例，可以对网页页面的内容进行全切分方法等处理，把内容分割成词；再对所得到的词进行重要性计算，基于重要性计算的结果来得到关键词。本领域技术人员可以理解的是，也可以综合采用统计分析和语义分析这两种手段来提取关键词。

步骤 403，对于至少一个关键词中的单个关键词，进行扩展以生成扩展关键词。

在本实施例中，可以对上述至少一个关键词中的单个关键词，都进行扩展以生成扩展关键词，其中，扩展关键词包括了上述单个关键词和以下中的至少一项：该单个关键词的同义词，例如，关键词“孩子”可以具有同义词“儿童”；该关键词的近义词，例如，关键词“中药”可以具有近义词“草药”，“出席”可以具有近义词“参加”；该关键词的关联词，例如，关键词“感冒”可以具有“发烧”或“流感”之类的关联词。

步骤 404，基于扩展关键词，生成关键词集合。

在本实施例中，利用步骤 403，可以针对每一个关键词生成其扩展关键词；之后，可以汇总上述至少一个关键词中的每一个关键词的扩展关键词，生成关键词集合（其包括了上述至少一个关键词的所有扩展关键词）。

步骤 405, 基于关键词集合与各条候选推送信息之间的匹配关系, 选取至少一条候选推送信息生成推送信息集合。

在本实施例中, 本步骤可以如下进行:

首先, 对各条候选推送信息进行内容解析, 分别提取候选推送信息关键词集合。作为示例, 可以采取与步骤 402 相同的内容分析方式对各条候选推送信息进行内容解析, 由此可以针对每条候选推送信息都提取对应的候选推送信息关键词集合。

其次, 将步骤 404 中得到的关键词集合分别与各个候选推送信息关键词集合进行相似度计算。在本实施例中, 可以采用余弦相似度 (cosine similarity) 算法、Jaccard 系数之类的公知的文本相似度计算方法来进行相似度计算。以 Jaccard 系数方法为例, 步骤 404 中得到的关键词集合与候选推送信息关键词集合之间的相似度=关键词集合与候选推送信息关键词集合之间共有的词的数目/关键词集合与候选推送信息关键词集合一起包括的词的数目。

最后, 基于相似度计算的结果, 选取至少一条候选推送信息生成推送信息集合。在本实施例中, 可以首先基于相似度计算的结果, 对各条候选推送信息进行排序得到候选推送信息序列 (例如采取相似度由高到低的顺序); 之后, 可以根据量值条件 (需要的候选推送信息的数量) 或者阈值条件 (例如相似度值要大于预设阈值), 从上述序列中选取至少一条候选推送信息生成推送信息集合。

步骤 406, 基于网页页面的内容和推送信息集合, 生成新网页。

在本实施例中, 上述电子设备可以将上述网页页面的内容与推送信息集合相结合作为新网页的内容, 并由此生成新网页。

从图 4 中可以看出, 与图 2 对应的实施例相比, 本实施例中的网页生成方法的流程 400 突出了对关键词进行扩展的步骤。由此, 本实施例描述的方案可以引入更多的关键词相关数据, 从而实现更全面的候选推送信息的选取和更有效的网页生成。

进一步参考图 5, 作为对上述各图所示方法的实现, 本申请提供了一种网页生成装置的一个实施例, 该装置实施例与图 2 所示的方法

实施例相对应，该装置具体可以应用于各种电子设备中。

如图 5 所示，本实施例所述的网页生成装置 500 包括：接收单元 501、解析单元 502、信息选取单元 503 和生成单元 504。其中，接收单元 501 配置用于接收用户的网页浏览请求，其中，上述网页浏览请求包括网址；解析单元 502 配置用于对上述网址所对应的网页页面进行内容解析，提取关键词集合；信息选取单元 503 配置用于基于上述关键词集合与各条候选推送信息之间的匹配关系，选取至少一条候选推送信息生成推送信息集合；而生成单元 504 配置用于基于上述网页页面的内容和上述推送信息集合，生成新网页。

在本实施例中，网页生成装置 500 的接收单元 501 可以通过有线连接方式或者无线连接方式从用户利用其进行网页浏览的终端接收网页浏览请求，其中，上述网页浏览请求包括了用户期望浏览的网页的地址，即网址。

在本实施例中，基于接收单元 501 得到的网址，上述解析单元 502 可以首先获取上述网址所对应的网页页面；之后再利用各种分析手段对上述网页页面的内容进行分析，从而提取出一个或多个关键词。

在本实施例中，网页生成装置 500 上可以预先存储多条候选推送信息，这些候选推送信息可以与上述网页页面的内容相结合，以便作为一个整体在浏览器上呈现。由此，网页生成装置 500 的信息选取单元 503 可以将上述关键词集合与各条候选推送信息的内容分别进行逐一匹配，并根据每条候选推送信息的内容所包括的关键词的数目来确定该条候选推送信息与关键词集合的匹配关系。根据匹配关系，上述信息选取单元 503 可以从多条候选推送信息中选取至少一条候选推送信息，并由此生成推送信息集合。

在本实施例中，上述生成单元 504 可以将上述网页页面的内容（即上述网址对应的页面的页面内容）与推送信息集合相结合作为新网页的内容，并由此生成新网页。

本领域技术人员可以理解，上述网页生成装置 500 还包括一些其他公知结构，例如处理器、存储器等，为了不必要地模糊本公开的实施例，这些公知的结构在图 5 中未示出。

下面参考图 6，其示出了适于用来实现本申请实施例的终端设备或服务器的计算机系统 600 的结构示意图。

如图 6 所示，计算机系统 600 包括中央处理单元（CPU）601，其可以根据存储在只读存储器（ROM）602 中的程序或者从存储部分 608 加载到随机访问存储器（RAM）603 中的程序而执行各种适当的动作和处理。在 RAM 603 中，还存储有系统 600 操作所需的各种程序和数据。CPU 601、ROM 602 以及 RAM 603 通过总线 604 彼此相连。输入/输出（I/O）接口 605 也连接至总线 604。

以下部件连接至 I/O 接口 605：包括键盘、鼠标等的输入部分 606；包括诸如阴极射线管（CRT）、液晶显示器（LCD）等以及扬声器等的输出部分 607；包括硬盘等的存储部分 608；以及包括诸如 LAN 卡、调制解调器等网络接口卡的通信部分 609。通信部分 609 经由诸如因特网的网络执行通信处理。驱动器 610 也根据需要连接至 I/O 接口 605。可拆卸介质 611，诸如磁盘、光盘、磁光盘、半导体存储器等等，根据需要安装在驱动器 610 上，以便于从其上读出的计算机程序根据需要被安装入存储部分 608。

特别地，根据本公开的实施例，上文参考流程图描述的过程可以被实现为计算机软件程序。例如，本公开的实施例包括一种计算机程序产品，其包括有形地包含在机器可读介质上的计算机程序，所述计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中，该计算机程序可以通过通信部分 609 从网络上被下载和安装，和/或从可拆卸介质 611 被安装。

附图中的流程图和框图，图示了按照本申请各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上，流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分，所述模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意，在有些作为替换的实现中，方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如，两个接连地表示的方框实际上可以基本并行地执行，它们

有时也可以按相反的顺序执行，这依所涉及的功能而定。也要注意的，框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以用执行规定的功能或操作的专用的基于硬件的系统来实现，或者可以用专用硬件与计算机指令的组合来实现。

描述于本申请实施例中所涉及到的单元可以通过软件的方式实现，也可以通过硬件的方式来实现。所描述的单元也可以设置在处理器中，例如，可以描述为：一种处理器包括接收单元、解析单元、信息选取单元和生成单元。其中，这些单元的名称在某种情况下并不构成对该单元本身的限定，例如，接收单元还可以被描述为“接收用户的网页浏览请求的单元”。

作为另一方面，本申请还提供了一种计算机可读存储介质，该计算机可读存储介质可以是上述实施例中所述装置中所包含的计算机可读存储介质；也可以是单独存在，未装配入终端中的计算机可读存储介质。所述计算机可读存储介质存储有一个或者一个以上程序，所述程序被一个或者一个以上的处理器用来执行描述于本申请的网页生成方法。

以上描述仅为本申请的较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解，本申请中所涉及的发明范围，并不限于上述技术特征的特定组合而成的技术方案，同时也应涵盖在不脱离所述发明构思的情况下，由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本申请中公开的(但不限于)具有类似功能的技术特征进行互相替换而形成的技术方案。

权 利 要 求 书

1、一种网页生成方法，其特征在于，所述方法包括：

接收用户的网页浏览请求，其中，所述网页浏览请求包括网址；

对所述网址所对应的网页页面进行内容解析，提取关键词集合；

基于所述关键词集合与各条候选推送信息之间的匹配关系，选取至少一条候选推送信息生成推送信息集合；

基于所述网页页面的内容和所述推送信息集合，生成新网页。

2、根据权利要求 1 所述的网页生成方法，其特征在于，所述对所述网址所对应的网页页面进行内容解析提取关键词集合，包括：

对所述网址所对应的网页页面的内容进行统计分析和/或语义分析，提取至少一个关键词；

基于所述至少一个关键词，生成关键词集合。

3、根据权利要求 2 所述的网页生成方法，其特征在于，所述基于所述至少一个关键词，生成关键词集合，包括：

对于所述至少一个关键词中的单个关键词，进行扩展以生成扩展关键词，其中，所述扩展关键词包括所述单个关键词和以下中的至少一项：所述单个关键词的同义词、所述单个关键词的近义词、所述单个关键词的关联词；

基于所述扩展关键词，生成关键词集合。

4、根据权利要求 1-3 之一所述的网页生成方法，其特征在于，所述基于所述关键词集合与各条候选推送信息之间的匹配关系，选取至少一条候选推送信息生成推送信息集合，包括：

对各条候选推送信息进行内容解析，分别提取候选推送信息关键词集合；

将所述关键词集合分别与各个候选推送信息关键词集合进行相似度计算；

基于相似度计算的结果，选取至少一条候选推送信息生成推送信息集合。

5、根据权利要求4所述的网页生成方法，其特征在于，所述基于相似度计算的结果，选取至少一条候选推送信息生成推送信息集合，包括：

基于相似度计算的结果以及预设的推送信息的量值，选取所述量值个数的候选推送信息生成推送信息集合。

6、根据权利要求1-5之一所述的网页生成方法，其特征在于，所述基于所述网页页面的内容和所述推送信息集合，生成新网页，包括：

以所述推送信息集合中的推送信息与所述网页页面的内容中的相应关键词相关联的方式，生成新网页。

7、根据权利要求1-6之一所述的网页生成方法，其特征在于，所述基于所述网页页面的内容和所述推送信息集合，生成新网页，包括：

以所述推送信息集合中的推送信息与所述网页页面的内容分开设置的方式，生成新网页。

8、一种网页生成装置，其特征在于，所述装置包括：

接收单元，配置用于接收用户的网页浏览请求，其中，所述网页浏览请求包括网址；

解析单元，配置用于对所述网址所对应的网页页面进行内容解析，提取关键词集合；

信息选取单元，配置用于基于所述关键词集合与各条候选推送信息之间的匹配关系，选取至少一条候选推送信息生成推送信息集合；

生成单元，配置用于基于所述网页页面的内容和所述推送信息集合，生成新网页。

9、根据权利要求8所述的网页生成装置，其特征在于，所述解析

单元包括:

分析模块,配置用于对所述网址所对应的网页页面的内容进行统计分析和/或语义分析,提取至少一个关键词;

生成模块,配置用于基于所述至少一个关键词,生成关键词集合。

10、根据权利要求 9 所述的网页生成装置,其特征在于,所述生成模块进一步配置用于:

对于所述至少一个关键词中的单个关键词,进行扩展以生成扩展关键词,其中,所述扩展关键词包括所述单个关键词和以下中的至少一项:所述单个关键词的同义词、所述单个关键词的近义词、所述单个关键词的关联词;

基于所述扩展关键词,生成关键词集合。

11、根据权利要求 8-10 之一所述的网页生成装置,其特征在于,所述信息选取单元包括:

解析模块,配置用于对各条候选推送信息进行内容解析,分别提取候选推送信息关键词集合;

相似度计算模块,配置用于将所述关键词集合分别与各个候选推送信息关键词集合进行相似度计算;

选取模块,配置用于基于相似度计算的结果,选取至少一条候选推送信息生成推送信息集合。

12、根据权利要求 11 所述的网页生成装置,其特征在于,所述选取模块进一步配置用于:

基于相似度计算的结果以及预设的推送信息的量值,选取所述量值个数的候选推送信息生成推送信息集合。

13、根据权利要求 8-12 之一所述的网页生成装置,其特征在于,所述生成单元进一步配置用于:

以所述推送信息集合中的推送信息与所述网页页面的内容中的相

应关键词相关联的方式，生成新网页。

14、根据权利要求 8-13 之一所述的网页生成装置，其特征在于，所述生成单元进一步配置用于：

以所述推送信息集合中的推送信息与所述网页页面的内容分开设置的方式，生成新网页。

15、一种设备，其特征在于，包括：

一个或者多个处理器；

存储器；

一个或者多个程序，所述一个或者多个程序存储在所述存储器中，当被所述一个或多个处理器执行时：

接收用户的网页浏览请求，其中，所述网页浏览请求包括网址；

对所述网址所对应的网页页面进行内容解析，提取关键词集合；

基于所述关键词集合与各条候选推送信息之间的匹配关系，选取至少一条候选推送信息生成推送信息集合；

基于所述网页页面的内容和所述推送信息集合，生成新网页。

16、一种非易失性计算机存储介质，所述计算机存储介质存储有一个或多个程序，当所述一个或者多个程序被一个设备执行时，使得所述设备：

接收用户的网页浏览请求，其中，所述网页浏览请求包括网址；

对所述网址所对应的网页页面进行内容解析，提取关键词集合；

基于所述关键词集合与各条候选推送信息之间的匹配关系，选取至少一条候选推送信息生成推送信息集合；

基于所述网页页面的内容和所述推送信息集合，生成新网页。

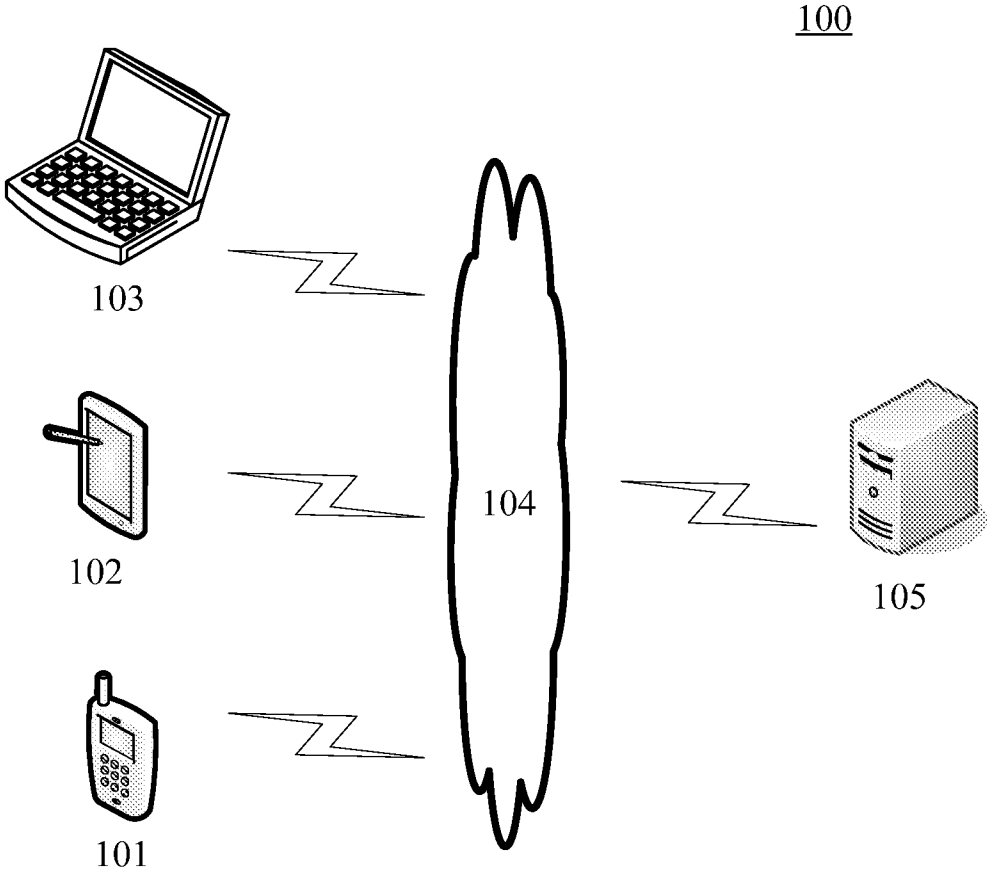


图 1

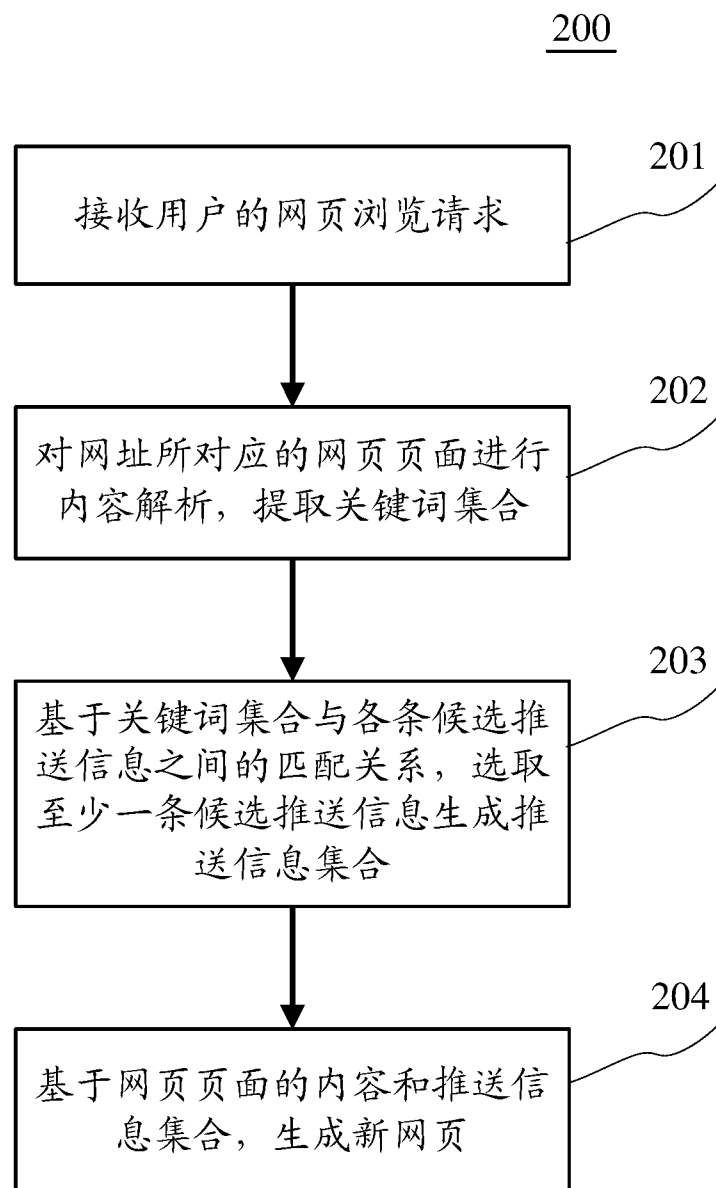


图 2

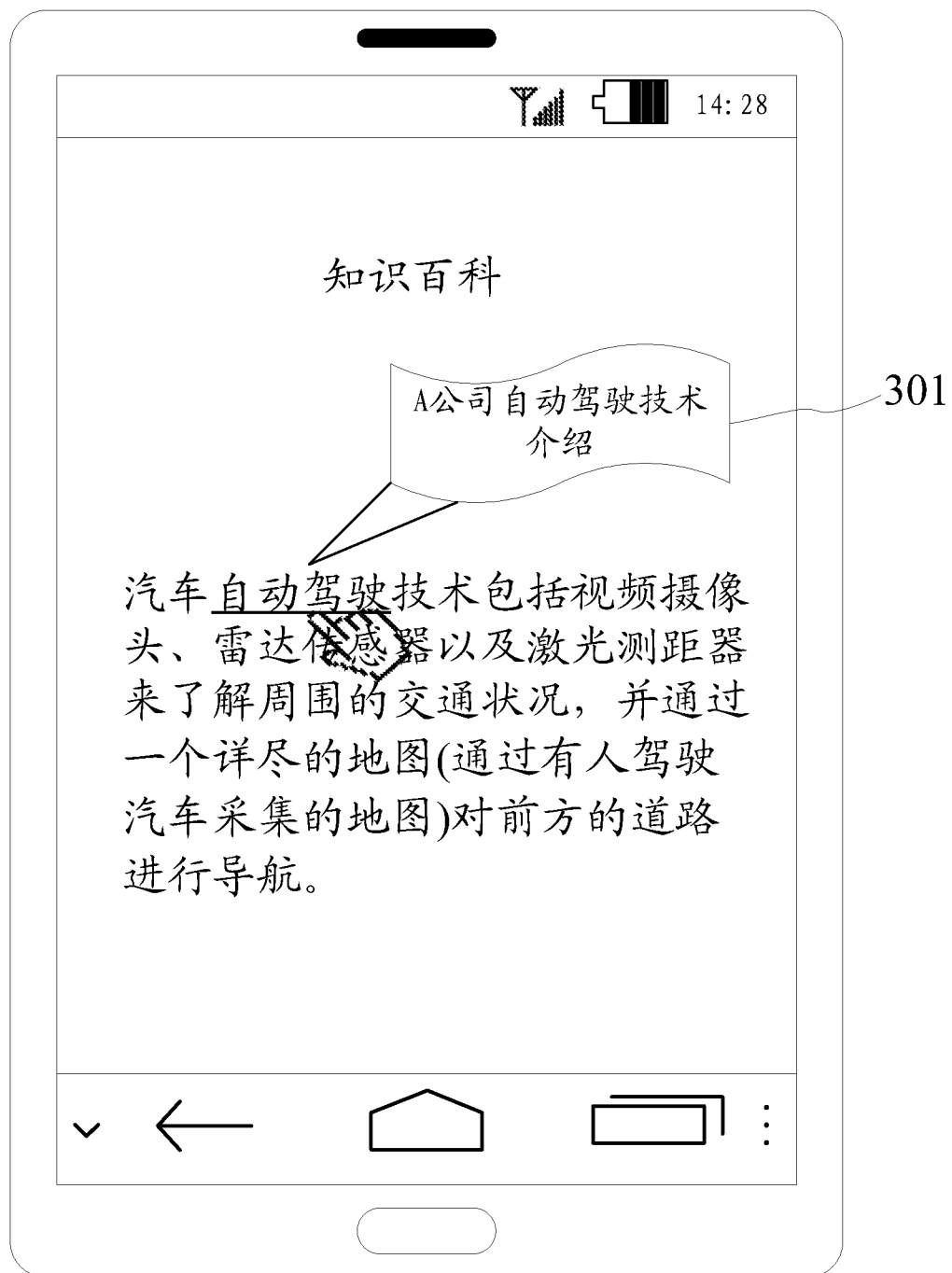


图 3

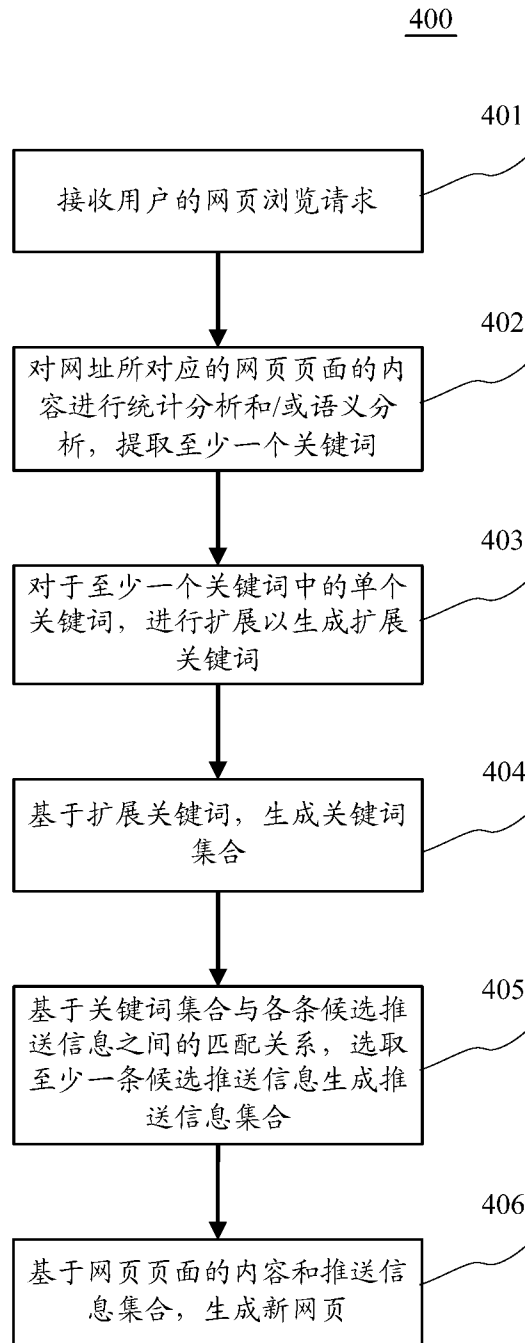


图 4

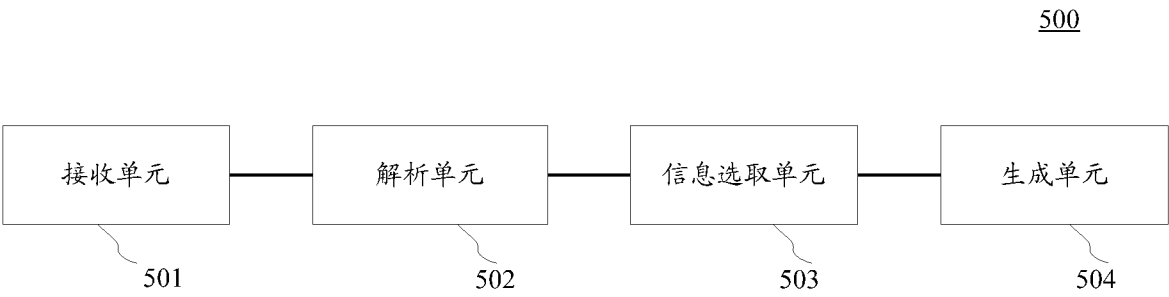


图 5

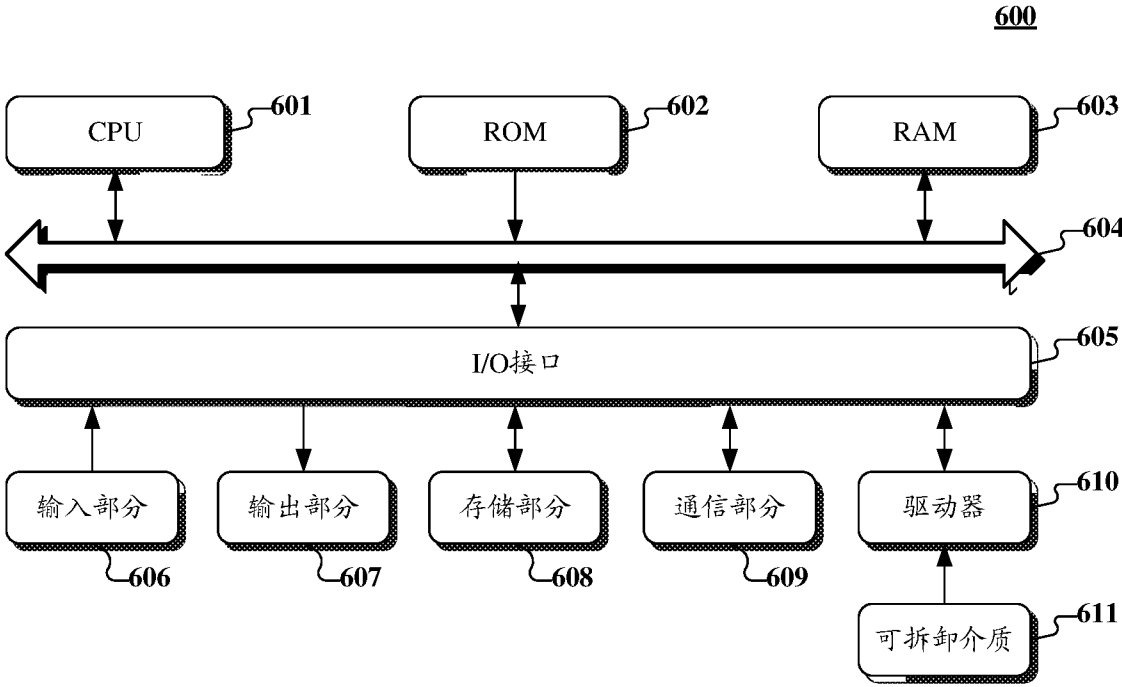


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2015/090703

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; G06Q; H04L

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)

CPRSABS; VEN; CNKI; IEEE: web page, push???, page?, keyword?, content

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 102646248 A (SHEN, Wence), 22 August 2012 (22.08.2012), description, paragraphs 50-110	1-16
Y	CN 103530339 A (BEIJING BAIDU NETWORK TECHNOLOGY CO., LTD.), 22 January 2014 (22.01.2014), description, paragraphs 26-140	1-16
A	CN 101071424 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.), 14 November 2007 (14.11.2007), the whole document	1-16
A	CN 101866341 A (HUAWEI TECHNOLOGIES CO., LTD.), 20 October 2010 (20.10.2010), the whole document	1-16

☐ 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 22 February 2016 (22.02.2016)	Date of mailing of the international search report 03 March 2016 (03.03.2016)
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 XU, Lin Telephone No.: (86-10) 62089885

International application No.
PCT/CN2015/090703

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

国际检索报告

国际申请号

PCT/CN2015/090703

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; G06Q; H04L

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))

CPRSABS;VEN;CNKI;IEEE:推送, 网页, 页面, 关键词, 关键字, 内容, push???, page?, keyword?, content,

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
Y	CN 102646248 A (沈文策) 2012年 8月 22日 (2012 - 08 - 22) 说明书第50-110段	1-16
Y	CN 103530339 A (北京百度网讯科技有限公司) 2014年 1月 22日 (2014 - 01 - 22) 说明书第26-140段	1-16
A	CN 101071424 A (腾讯科技深圳有限公司) 2007年 11月 14日 (2007 - 11 - 14) 全文	1-16
A	CN 101866341 A (华为技术有限公司) 2010年 10月 20日 (2010 - 10 - 20) 全文	1-16

☐ 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 2月 22日

国际检索报告邮寄日期

2016年 3月 3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

授权官员

徐琳

传真号 (86-10)62019451

电话号码 (86-10)62089885

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2015/090703

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	102646248	A	2012年 8月 22日	CN 102646248 B	2016年 1月 13日
CN	103530339	A	2014年 1月 22日	无	
CN	101071424	A	2007年 11月 14日	CN 101071424 B	2010年 8月 25日
CN	101866341	A	2010年 10月 20日	无	

表 PCT/ISA/210 (同族专利附件) (2009年7月)