

(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 6 月 22 日 (22.06.2017)



(10) 国际公布号
WO 2017/101425 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2016/089556
- (22) 国际申请日: 2016 年 7 月 10 日 (10.07.2016)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201510930307.6 2015 年 12 月 15 日 (15.12.2015) 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: 朱家星 (ZHU, Jiaxing); 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: 上海晨皓知识产权代理事务所(普通合伙) (SHANGHAI CHENHAO INTELLECTUAL PROPERTY LAW FIRM GENERAL PARTNER-SHIP); 中国上海市黄浦区制造局路 787 号二幢 202B 室, Shanghai 200011 (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: APPARATUS, METHOD AND SYSTEM FOR USE IN RETRIEVAL OF MEDIA RESOURCES

(54) 发明名称: 用于媒体资源检索的设备、方法及系统

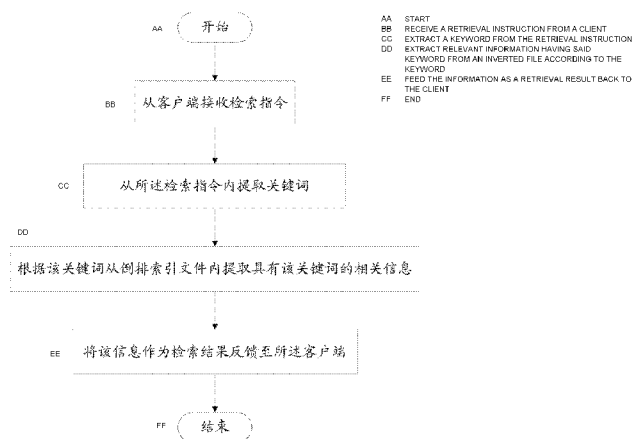


图 2

(57) Abstract: An apparatus, a method and a system for use in retrieval of media resources, said apparatus comprising: a receiving device (210) used for receiving a retrieval instruction from a client (100); a processing device (220) used for extracting a keyword from the retrieval instruction, for extracting relevant information having said keyword from an inverted file according to the keyword, and for feeding the information as a retrieval result back to the client (100), wherein the inverted file stores the relevant information that pertain to media files. By virtue of the described technical solution, all information pertaining to media files are pre-stored in the inverted file, and a server (200) may, upon receiving the retrieval instruction from the client (100), directly extract information complying with the retrieval instruction from the inverted file, and feed the information back to the client (100). Compared with existing method of matching media files by using like matching, the present mode significantly improves retrieval speed on media files and lessens workload of the server (200).

(57) 摘要:

[见续页]



WO 2017/101425 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)。

根据细则 4.17 的声明:

- 关于申请人有权要求在先申请的优先权(细则 4.17(iii))

一种用于媒体资源检索的设备、方法及系统，该设备包含：接收装置（210），用于从客户端（100）接收检索指令；处理装置（220），用于从所述检索指令内提取关键词，并根据该关键词从倒排索引文件内提取具有该关键词的相关信息，并将该信息作为检索结果反馈至所述客户端（100），其中所述倒排索引文件内存储有有关媒体文件的相关信息。通过上述技术方案，所有有关媒体文件的信息均被预先存入倒排索引文件内，服务器（200）在接收到来自客户端（100）的检索指令之后可直接从该倒排索引文件内提取与该检索指令相符合的信息，并将该信息反馈至客户端（100）。相比于现有的采用 like 匹配的方式来匹配媒体文件，该方式显著提高了对媒体文件的检索速度，并缓解了服务器（200）的工作负荷。

用于媒体资源检索的设备、方法及系统

本申请要求于 2015 年 12 月 15 日提交中国专利局、申请号为 201510930307.6 的中国专利申请的优先权，其全部内容通过引用结合在本
5 申请中。

技术领域

本发明实施例涉及信息技术领域，具体地，涉及一种用于媒体资源检索的设备、方法及系统。

10

背景技术

随着信息技术的发展，各类娱乐影像资料出现了爆发式的增长，现有的媒体资产管理系统（Media Asset Management System）存储有海量媒体文件（例如，视频、音频、图片等等）。对于如何快速从该媒体资产管理系统检索出用户想要的文件，则成为了本行业亟待解决的问题。

15

在现有的媒体资产管理系统中，媒体文件主要存储于该系统的服务器内，用户可通过客户端（例如，个人计算机、手机、平板电脑等）发送检索请求，服务器收到该检索请求之后从中提取关键词（例如，“甄嬛传”），并采用 like 匹配的方式从所存储的媒体文件中提取相应的信息。由于服务器所存储的媒体文件的数量相当之巨大，采用 like 匹配的方式来从所存储的大量媒体文件内提取相应的信息速度非常慢，服务器的工作负荷也较大，最终导致对用户的检索请求响应速度偏慢，用户体验较差。

20

发明内容

本发明实施例的目的是提供一种用于媒体资源检索的设备、方法及系

25

统，其可使得媒体资产管理系统中的服务器在收到来自客户端的检索请求之后迅速找到相应的信息并反馈给客户端，响应速度较快。

为了实现上述目的，本发明实施例提供一种用于媒体资源检索的设备，该设备包含：接收装置，用于从客户端接收检索指令；处理装置，用于从所述检索指令内提取关键词，并根据该关键词从预置的倒排索引文件内提取具有该关键词的相关信息，并将该信息作为检索结果反馈至所述客户端，其中所述倒排索引文件内存储有有关媒体文件的相关信息。

其中，所述相关信息可包含以下中的一者或多者：媒体文件名称、播放平台、付费平台以及媒体文件类型。

其中，所述接收装置还可用于接收媒体文件；以及所述处理装置还可用于从所述媒体文件提取相关信息，并将该信息存入所述倒排索引文件内。

其中，所述倒排索引文件可存储于所述处理装置的缓存内。

相应地，本发明实施例还提供一种媒体资源检索系统，该系统包含：客户端，用于发送检索指令；以及服务器，该服务器包含上述用于媒体资源检索的设备。

相应地，本发明实施例还提供一种用于媒体资源检索的方法，该方法包括：从客户端接收检索指令；从所述检索指令内提取关键词，并根据该关键词从预置的倒排索引文件内提取具有该关键词的相关信息，并将该信息作为检索结果反馈至所述客户端，其中所述倒排索引文件内存储有有关媒体文件的相关信息。

其中，所述相关信息可包含以下中的一者或多者：媒体文件名称、播放平台、付费平台以及媒体文件类型。

其中，所述方法还可包括：接收媒体文件；以及从所述媒体文件提取相关信息，并将该信息存入所述倒排索引文件内。

其中，所述倒排索引文件可存储于缓存内。

通过上述技术方案，所有有关媒体文件的信息均被预先存入倒排索引文件内，服务器在接收到来自客户端的检索指令之后可直接从该倒排索引文件内提取与该检索指令相符合的信息，并将该信息反馈至客户端。相比于现有的采用 like 匹配的方式来匹配媒体文件，该方式显著提高了对媒体文件的检索速度，并缓解了服务器的工作负荷。

本发明实施例的其它特征和优点将在随后的具体实施方式部分予以详细说明。

附图说明

附图是用来提供对本发明实施例的进一步理解，并且构成说明书的一部分，与下面的具体实施方式一起用于解释本发明实施例，但并不构成对本发明实施例的限制。在附图中：

图 1 为本发明实施例提供的媒体资源检索设备的结构示意图；以及
图 2 为本发明实施例提供的媒体资源检索方法的流程图。

具体实施方式

以下结合附图对本发明具体实施方式进行详细说明。应当理解的是，此处所描述的具体实施方式仅用于说明和解释本发明实施例，并不用于限制本发明实施例。

图 1 为本发明实施例提供的媒体资源检索设备的结构示意图。如图 1 所示，本发明实施例提供了一种媒体资源检索系统，该系统包含客户端 100，用于发送检索指令；以及服务器 200，该服务器 200 包含用于媒体资源检索的设备。该用于媒体资源检索的设备该设备包含：接收装置 210，用于从客

户端接收检索指令；处理装置 220，用于从所述检索指令内提取关键词，并根据该关键词从预置的倒排索引文件内提取具有该关键词的相关信息（包含诸如媒体文件名称、播放平台、付费平台以及媒体文件类型等等），并将该信息作为检索结果反馈至所述客户端。

5 其中，所述倒排索引文件内存储有有关媒体文件的相关信息。例如，其可存储有媒体文件的属性值（例如，文件名称、播放平台等）以及具有该属性值的媒体文件的地址。也就是说，倒排索引文件内的每一项记录均包含一个属性值和具有该属性值的各媒体文件的地址。在一般的数据存储及检索方式，均是通过遍历所存储的每一文件，确定该文件的属性，并将该属性与检索关键词进行比较，这样较为费时，而通过本发明实施例的方案，可很轻松得从所存储的所有媒体文件的属性值寻找到符合检索关键词的属性，并确定具有该属性的媒体文件的地址。以检索词为“甄嬛传”为例，处理装置可搜索倒排索引文件，判断是否存在所述媒体文件名称为“甄嬛传”的媒体文件，并根据搜索结果向客户端进行反馈。在存在媒体文件名称为“甄嬛传”的媒体文件的情况下，还可将该媒体文件的地址一同反馈给客户端，以便该客户端进行访问该媒体文件。以检索词为“MP4”为例，处理装置可搜索倒排索引文件，判断是否存在所述媒体文件类型为“MP4”的媒体文件，并根据搜索结果向客户端进行反馈。在存在媒体文件类型为“MP4”的媒体文件的情况下，还可将该媒体文件的地址一同反馈给客户端，以便该客户端进行访问该媒体文件。

20 所述预置的倒排索引文件可通过以下操作而被生成：所述接收装置可接收媒体文件；以及所述处理装置还针对所述接收装置接收的每一媒体文件，从该媒体文件提取相关信息（即，属性值，例如媒体文件名称、播放平台、付费平台以及媒体文件类型等等），并将该信息存入所述倒排索引文件内。当然，同时存储所述倒排索引文件内的还有所述媒体文件的地址。需要说明的是，所述属性值及相关信息并不限于上述列举的内容，还可例如为媒体文

25

件的码率信息等等，本发明实施例并不限于此。

在一个实施例中，所述设备可为基于 ElasticSearch 技术的搜索平台，该设备可作为实现检索功能的集群中的一节点来提供检索结果。该 ElasticSearch 技术的搜索平台能够达到实时搜索且稳定、可靠、快速的效
5 果。

在一个实施例中，所述倒排索引文件可存储于所述处理装置的缓存内。由于缓存内的数据存取速度要高于对硬盘上的数据进行存取的速度，通过该布置可进一步提升检索速度。

图 2 为本发明实施例提供的媒体资源检索方法的流程图。如图 2 所示，
10 本发明实施例还提供一种用于媒体资源检索的方法，该方法包括：从客户端接收检索指令；从所述检索指令内提取关键词，并根据该关键词从预置的倒排索引文件内提取具有该关键词的相关信息（包含诸如媒体文件名称、播放平台、付费平台以及媒体文件类型等等），并将该信息作为检索结果反馈至所述客户端，其中所述倒排索引文件内存储有有关媒体文件的相关信息。藉
15 此，由于所有有关媒体文件的信息均被预先存入倒排索引文件内，且有别于一般的文件存储方式，倒排索引文件内的每一项记录均包含一个属性值和具有该属性值的各媒体文件的地址，因此服务器在接收到来自客户端的检索指令之后可直接从该倒排索引文件内提取与该检索指令相符合的信息，并将该信息反馈至客户端。相比于现有的采用 like 匹配的方式来匹配媒体文件，该
20 方式显著提高了对媒体文件的检索速度，并缓解了服务器的工作负荷。

所述预置的倒排索引文件可通过以下操作而被生成：接收媒体文件；以及从所述媒体文件提取相关信息（即，属性值，例如媒体文件名称、播放平台、付费平台以及媒体文件类型等等），并将该信息存入所述倒排索引文件内。也就是说，服务器每存储一媒体文件，均会提取该媒体文件的属性信息
25 存入所述倒排索引文件内以用于后续检索之用。当然，同时存储所述倒排索

引文件内的还有所述媒体文件的地址。需要说明的是，所述属性值及相关信息并不限于上述列举的内容，还可例如为媒体文件的码率信息等等，本发明实施例并不限于此。

其中，所述方法可基于 ElasticSearch 技术，该技术可通过具有检索功
5 能的集群来提供检索结果，能够达到实时搜索且稳定、可靠、快速的效果。

其中，所述倒排索引文件可存储于缓存内。由于缓存内的数据存取速度要高于对硬盘上的数据进行存取的速度，通过该布置可进一步提升检索速度。

通过上述技术方案，所有有关媒体文件的信息均被预先存入倒排索引文
10 件内，服务器在接收到来自客户端的检索指令之后可直接从该倒排索引文件内提取与该检索指令相符合的信息，并将该信息反馈至客户端。相比于现有的采用 like 匹配的方式来匹配媒体文件，该方式显著提高了对媒体文件的检索速度，并缓解了服务器的工作负荷。另外，从技术角度来说，由于媒体文件文件包含很多方面的信息（诸如，码率信息），服务器的数据库在存储这
15 些信息时，为了减少中间表的生成，会产生很多冗余字段，而在采用本申请方案之后，这些信息可以直接存入倒排索引文件内，不需要再对数据库进行扩展，减小了对数据库存储容量的压力。

以上结合附图详细描述了本发明实施例的具体实施方式，但是，本发明实施例并不限于上述实施方式中的具体细节，在本发明实施例的技术构思范
20 围内，可以对本发明实施例的技术方案进行多种简单变型，这些简单变型均属于本发明实施例的保护范围。

另外需要说明的是，在上述具体实施方式中所描述的各个具体技术特征，在不矛盾的情况下，可以通过任何合适的方式进行组合。为了避免不必要的重复，本发明实施例对各种可能的组合方式不再另行说明。

25 此外，本发明实施例的各种不同的实施方式之间也可以进行任意组合，

只要其不违背本发明实施例的思想，其同样应当视为本发明实施例所公开的内容。

权 利 要 求 书

1、一种用于媒体资源检索的设备，包含：

接收装置，用于从客户端接收检索指令；

处理装置，用于从所述检索指令内提取关键词，并根据该关键词从预置的倒排索引文件内提取具有该关键词的相关信息，并将该信息作为检索结果
5 反馈至所述客户端，其中所述倒排索引文件内存储有有关媒体文件的相关信息。

2、根据权利要求 1 所述的设备，其中，所述相关信息包含以下中的一者或多者：媒体文件名称、播放平台、付费平台以及媒体文件类型。

10

3、根据权利要求 1 或 2 所述的设备，其中，

所述接收装置还用于接收媒体文件；以及

所述处理装置还用于从所述媒体文件提取相关信息，并将该信息存入所述倒排索引文件内。

15

4、根据权利要求 1~3 中任一项所述的设备，其中，所述倒排索引文件存储于所述处理装置的缓存内。

5、一种媒体资源检索系统，包含：

20 客户端，用于发送检索指令；以及

服务器，该服务器包含根据权利要求 1-4 中任一项权利要求所述的用于媒体资源检索的设备。

6、一种用于媒体资源检索的方法，包括：

25 从客户端接收检索指令；

从所述检索指令内提取关键词，并根据该关键词从预置的倒排索引文件内提取具有该关键词的相关信息，并将该信息作为检索结果反馈至所述客户端，其中所述倒排索引文件内存储有有关媒体文件的相关信息。

5 7、根据权利要求 6 所述的方法，其中，所述相关信息包含以下中的一者或多者：媒体文件名称、播放平台、付费平台以及媒体文件类型。

8、根据权利要求 6 或 7 所述的方法，其中，该方法还包括：
接收媒体文件；以及

10 从所述媒体文件提取相关信息，并将该信息存入所述倒排索引文件内。

9、根据权利要求 6~8 中任一项所述的方法，其中，所述倒排索引文件存储于缓存内。

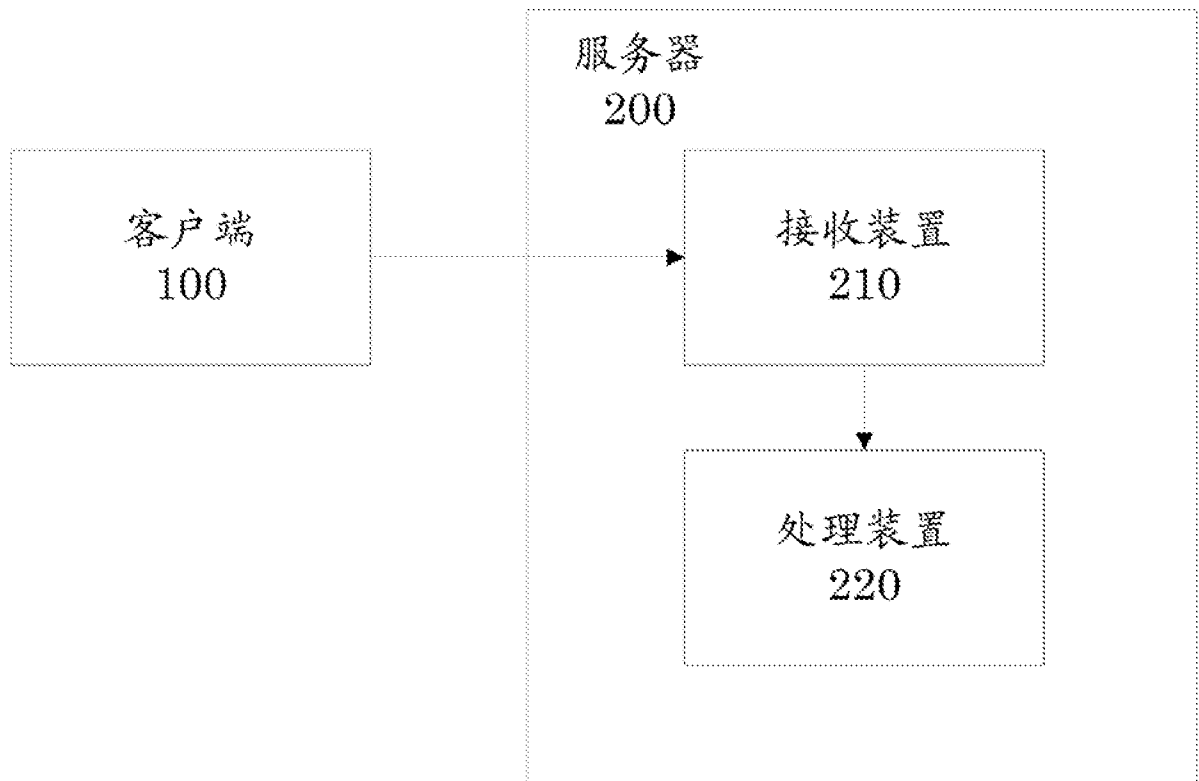


图 1

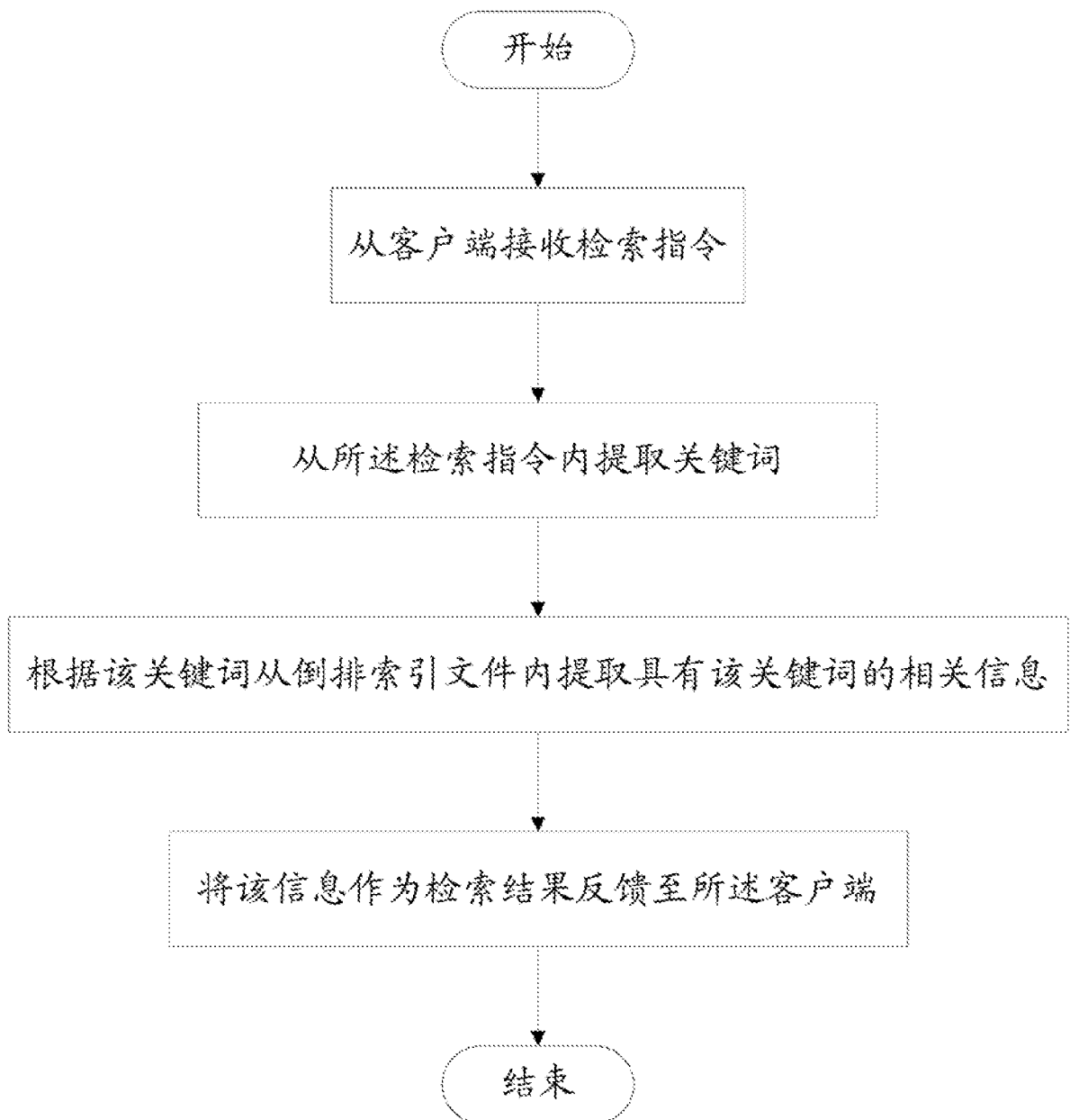


图 2

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/089556

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: search, keyword, server, client-side, reverse index

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 103744913 A (GOSUNCN TECHNOLOGY GROUP CO., LTD.), 23 April 2014 (23.04.2014), description, paragraphs [0008]-[0018]	1-9
X	CN 102761843 A (SHANGHAI TRASIN INFORMATION TECHNOLOGY CO., LTD.), 31 October 2012 (31.10.2012), claims 1-10	1-9
X	CN 101101605 A (HUAWEI TECHNOLOGIES CO., LTD.), 09 January 2008 (09.01.2008), description, page 11, paragraph 3 to page 15, paragraph 10	1-9
X	CN 101655848 A (HUAWEI TECHNOLOGIES CO., LTD.), 24 February 2010 (24.02.2010), description, page 3, paragraph 8 to page 7, paragraph 3	1-9

☐ 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 September 2016 (22.09.2016)	Date of mailing of the international search report 12 October 2016 (12.10.2016)
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 SU, Fei Telephone No.: (86-10) 62413632

International application No.
PCT/CN2016/089556

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

国际检索报告

国际申请号

PCT/CN2016/089556

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))

WPI, EPODOC, CNPAT, CNKI: 检索, 搜索, 查找, 关键词, 倒排索引, 服务器, 客户端, search, keyword, server, client-side, reverse index

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
X	CN 103744913 A (高新兴科技集团股份有限公司) 2014年 4月 23日 (2014 - 04 - 23) 说明书第[0008]-[0018]段	1-9
X	CN 102761843 A (上海洲信信息技术有限公司) 2012年 10月 31日 (2012 - 10 - 31) 权利要求1-10	1-9
X	CN 101101605 A (华为技术有限公司) 2008年 1月 9日 (2008 - 01 - 09) 说明书第11页第3段-第15页第10段	1-9
X	CN 101655848 A (华为技术有限公司) 2010年 2月 24日 (2010 - 02 - 24) 说明书第3页第8段-第7页第3段	1-9

☐ 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 9月 22日

国际检索报告邮寄日期

2016年 10月 12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

受权官员

苏菲

传真号 (86-10)62019451

电话号码 (86-10)62413632

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2016/089556

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	103744913	A	2014年 4月 23日	无			
CN	102761843	A	2012年 10月 31日	CN	102761843	B	2015年 6月 17日
CN	101101605	A	2008年 1月 9日	CN	100478962	C	2009年 4月 15日
CN	101655848	A	2010年 2月 24日	无			

表 PCT/ISA/210 (同族专利附件) (2009年7月)