

(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 29 日 (29.06.2017)



(10) 国际公布号
WO 2017/107449 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2016/088955
- (22) 国际申请日: 2016 年 7 月 6 日 (06.07.2016)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201510977935.X 2015 年 12 月 23 日 (23.12.2015) 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: 赵锦城 (ZHAO, Jincheng); 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: 北京合智同创知识产权代理有限公司 (BEIJING HEADSTAY INTELLECTUAL PROPERTY INC.); 中国北京市海淀区大钟寺 13 号院 1 号楼华杰大厦 5A1-2, 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,

[见续页]

(54) Title: METHOD AND DEVICE FOR CAPTURING WEBPAGE VIDEO

(54) 发明名称: 网页视频抓取的方法及网页视频抓取的装置

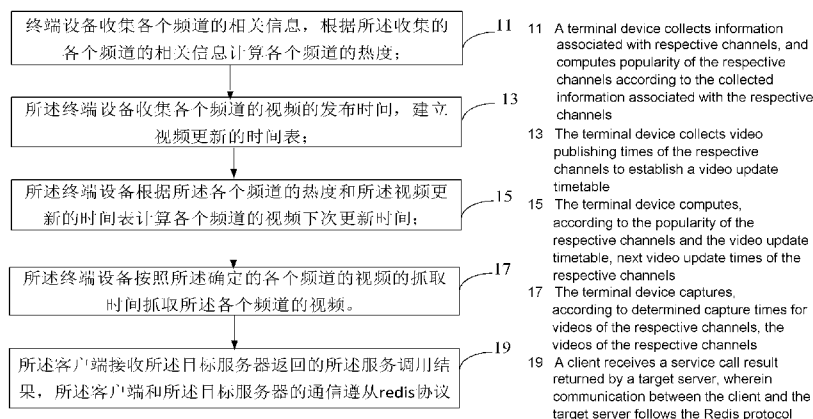


图 1

(57) Abstract: A method for capturing a webpage video comprises: collecting information associated with respective channels, and computing popularity of the respective channels according to the collected information associated with the respective channels (11); collecting video publishing times of the respective channels to establish a video update timetable (13); computing, according to the popularity of the respective channels and the video update timetable, next video update times of the respective channels (15); determining, according to the next video update times of the respective channels, capture times for videos of the respective channels; and capturing the videos of the respective channels according to the determined capture times for the videos of the respective channels (17). The solution captures videos from webpages only at determined video capture times, thereby significantly saving network resources.

(57) 摘要: 一种网页视频抓取方法, 包括: 收集各个频道的相关信息, 根据所述收集的各个频道的相关信息计算各个频道的热度 (11); 收集各个频道的视频的发布时间, 建立视频更新的时间表 (13); 根据所述各个频道的热度和所述视频更新的时间表计算各个频道的视频下次更新时间 (15); 根据所述各个频道的视频下次更新时间确定所述各个频道的视频的抓取时间; 按照所述确定的各个频道的视频的抓取时间抓取所述各个频道的视频 (17)。该方案在抓取网页视频时只在确定好的视频抓取时间去网页上抓取视频, 极大的节约了网络资源。

WO 2017/107449 A1



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)。

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))

本国际公布:

- 包括国际检索报告(条约第 21 条(3))。

网页视频抓取的方法及网页视频抓取的装置

本申请要求于 2015 年 12 月 23 日提交中国专利局、申请号为 201510977935.X，发明名称为“网页视频抓取的方法及网页视频抓取的终端设备”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

技术领域

本申请实施例涉及网络搜索技术领域，尤其涉及一种网页视频抓取的方法及网页视频抓取的装置。

背景技术

1994 年基于 WEB 的搜索引擎出现以来，搜索引擎便得到了极大的发展。搜索引擎解决了海量互联网资源的快速定位和检索，在人们日常生活和工作中发挥了越来越大的作用。作为搜索引擎的重要一部分，网页信息抓取技术，即抓取技术的发展也是越来越成熟。在计算机领域，越来越多的网络抓取技术都以开源的方式提供给开发人员和企业应用。

随着视频网站在国内的普及和蓬勃发展，人们的互联网生活几乎与视频网站已经密不可分。人们不仅十分喜欢视频这种直观的信息获取方式，而且爱浏览视频，并且很多人都亲自参与视频的拍摄和制作，并发布视频于视频站点与网友们共享和交流。

随着互联网技术的发展，搜索引擎技术在其中扮演着越来越重要的角色。网络爬虫决定搜索引擎的数量来源，是搜索引擎系统的重要组成部分。YouTube 是设立在美国的一个视频分享网站，该网站鼓励用户上传视频、观看视频及分享视频或短片，至今 YouTube 已经成为影音网站的翘楚。

为丰富全网视频搜索的内容，对于 YouTube 上视频数据的抓取目前广泛采用定时轮询的方案。然而从现有 YouTube 中遍历几十万的频道，定时抓取视频数据资源耗费太大。具体到各个频道，不同的频道有不同的播放总数、视频总量及订阅数，而且不同的频道更新周期不同，由此，定时轮询存在着严重的时效性及资源浪费的问题。

有鉴于此，本申请实施例提供一种网页视频抓取的方法及网页视频抓取的装置，可以实现在抓取网页视频时时效性高，而且在抓取网页视频时节约网络资源。

发明内容

本申请实施例提供一种网页视频抓取的方法及一种网页视频抓取的设备，可以全部或者部分解决现有技术中在抓取网页视频时使用的定时轮询的方法存在严重的时效性差及资源浪费的技术问题。

5 本申请实施例提供一种网页视频抓取的方法，包括：

收集各个频道的相关信息，根据所述收集的各个频道的相关信息计算各个频道的热度；

收集各个频道的视频的发布时间，建立视频更新的时间表；

10 根据所述各个频道的热度和所述视频更新的时间表计算各个频道的视频下次更新时间；

根据所述各个频道的视频下次更新时间确定所述各个频道的视频的抓取时间；

按照所述确定的各个频道的视频的抓取时间抓取所述各个频道的视频。

15 本申请实施例还提供一种在其上记录有配置为执行上述方法的程序的计算机可读记录介质。

本申请实施例提供一种网页视频抓取装置，包括：

记录模块，配置为收集各个频道的相关信息，根据所述收集的各个频道的相关信息计算各个频道的热度，并收集各个频道的视频的发布时间，建立视频更新的时间表；

20 评估模块，配置为根据所述各个频道的热度和所述视频更新的时间表计算各个频道的视频下次更新时间，并根据所述各个频道的视频下次更新时间确定所述各个频道的视频的抓取时间；

视频抓取模块，配置为按照所述确定的各个频道的视频的抓取时间抓取所述各个频道的视频。

25 本申请实施例提供的一种网页视频抓取的方法及一种网页视频抓取装置，针对用户的各个频道的热度和各个频道的视频的更新时间，计算各个视频的下次更新时间，并根据所述各个频道的视频的下次更新时间确定所述各个频道视频的抓取时间，进而根据所述各个视频的抓取时间来抓取所述各个频道的视频，视频抓取的时效性高，而且不会存在现有技术的轮询的技术方案，本申请实施例所提供的方案在抓取网页视频时只在确定好的视频抓取时
30

间去网页上抓取视频，该方案极大的节约了网络资源。

附图说明

为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作一简单地介绍，显而易见地，下面描述中的附图是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 为本申请一种网页视频抓取的方法实施例的流程图；

图 2 为本申请一种网页视频抓取装置实施例的结构示意图。

具体实施方式

为使本申请实施例的目的、技术方案和优点更加清楚，下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

请参考图 1，本申请一种网页视频抓取的方法实施例示意图，应用于服务器端。本申请实施例提供一种网页视频抓取的方法包括：

步骤 11，收集各个频道的相关信息，根据所述收集的各个频道的相关信息计算各个频道的热度；

步骤 13，所述收集各个频道的视频的发布时间，建立视频更新的时间表；

步骤 15，根据所述各个频道的热度和所述视频更新的时间表计算各个频道的视频下次更新时间；

步骤 17，根据所述各个频道的视频下次更新时间确定所述各个频道的视频的抓取时间；

步骤 19，按照所述确定的各个频道的视频的抓取时间抓取所述各个频道的视频。

本申请实施例提供的一种网页视频抓取的方法，针对发布视频的用户在视频网站的热度（即所述用户在所述频道的热度）和发布视频的用户在各个频道的视频的更新时间，计算各个视频的下次更新时间，并根据所述各个频道的视频的下次更新时间确定所述各个频道视频的抓取时间，进而根据所述各个视频的抓取时间来抓取所述各个频道的视频，视频抓取的时效性高，而

且不会存在现有技术的轮询的技术方案，本申请实施例所提供的方案在抓取网页视频时只在确定好的视频抓取时间去网页上抓取视频，该方案极大的节约了网络资源。

具体地，根据本申请一种网页视频抓取的方法实施例所公开的技术方案，
5 所述收集各个频道的相关信息包括：收集各个频道的视频数、订阅数和播放总数；则所述根据所述收集的各个频道的相关信息计算各个频道的热度包括根据所述收集的各个频道的视频数、订阅数和播放总数计算各个频道的热度。

具体地，根据本申请一种网页视频抓取的方法实施例所公开的技术方案，
10 所述收集各个频道的视频的发布时间，建立视频更新的时间表，包括：收集各个频道的视频的发布时间，计算各个频道的视频每次更新的时间间隔和更新的时间段，建立所述视频更新的时间表。所述视频更新的时间表可以参考如下表 1。

表 1

频道	发布时间
娱乐	2015 年 12 月 1 日 19:00
新闻	2015 年 12 月 3 日 9:00

具体地，根据本申请一种网页视频抓取的方法实施例所公开的技术方案，
15 所述根据所述各个频道的热度和所述视频更新的时间表计算各个频道的视频下次更新时间，包括：根据所述各个频道的热度、所述视频每次更新的时间间隔和更新的时间段计算各个频道的视频下次更新时间。

根据本申请也在网页视频抓取的方法实施例所公开的技术方案，所述各个频道的热度表示发布视频的用户在发布视频的网站的热度，即所述发布视频的用户在发布视频的网站受欢迎的程度，即热门程度，热度越高代表所述
20 发布视频的用户在所述网站越是受欢迎。热度的计算可以通过多次试验进行验证，针对某频道对所述收集的视频数、订阅数和播放总数赋予不同的权重来计算各个频道的热度。例如，通过分析用户习惯，各个权重由高到低依次为：订阅数、播放总数和视频数。

25 下面以一个比较热门的用户频道作为例子进行介绍。第一步：抓取其视频总数(video_num)、用户订阅数(fans_num)及播放总数(play_total)，设置其热度为 $\text{fans_num} \times 0.7 + \text{play_total} \times 0.2 + \text{video_num} \times 0.1$ ；第二步：抓取改用户频道

下的各个视频及其更新时间，如果视频数过多，则截取其最近的 1 万条视频信息，根据视频与更新时间的对应关系，建立视频更新时间表；第三步：、通过分析视频更新时间表，分析用户习惯和计算相邻两视频间的更新间隔，计算出视频下次的更新间隔；第四步：通过计算出的更新间隔除以该用户频道
5 的热度得出最终的更新时间间隔，并加上当前时间得到用户频道的最终更新时间。

根据本申请一种网页视频抓取的方法实施例所公开的技术方案，所述方法还包括：收集各个频道视频的视频抓取时间，在所述视频更新的时间表记录所述视频的抓取时间；计算所述视频的发布时间与所述视频的抓取时间的
10 间隔；则，所述根据所述各个频道的热度和所述视频更新的时间表计算各个频道的视频下次更新时间，还包括：根据所述各个频道的热度、所述视频每次更新的时间间隔、更新的时间段及所述视频的发布时间与所述视频的抓取时间的间隔计算各个频道的视频下次更新时间。本申请实施例所述的方案，计算视频的发布时间和抓取时间的间隔，并根据所述发布时间和抓取时间的
15 间隔的大小来计算视频的下次更新时间，即本申请实施例所公布的方案中在后续确定视频的下次更新时间时充分考虑了之前的视频的抓取时间是否合适这个因素，由此，不断迭代，自动调整视频的下次更新时间，进而确定视频的抓取时间，使得根据本申请实施例所公开的技术方案所进行的视频的抓取时间越来越准确，视频抓取的时效性越来越高，而且由于所述视频的抓取时间控制的准确，不会产生网络资源的浪费，进一步解决网络资源。
20

具体地，根据本申请一种网页视频抓取的方法实施例所公开的技术方案，在所述按照所述确定的各个频道的视频的视频的抓取时间抓取所述各个频道的视频之后，所述方法包括存储所述抓取的各个频道的视频的信息。

具体地，根据本申请一种网页视频抓取的方法实施例所公开的技术方案，
25 在所述存储所述抓取的各个频道的视频的信息之后，所述方法包括向用户推荐所述抓取的各个频道的视频。

根据本申请一种网页视频抓取的方法实施例所公开的技术方案，在视频抓取之后，所述可以存储抓取的视频的关键信息，所述关键信息可以包括视频的链接、视频的抓取时间、视频的热度（播放次数、点赞数、踩数）等。
30 所述发现有用户登录视频网页后，可以向所述登录的用户推荐已经存储的视

频。向所述用户推荐视频的具体过程可以是将所述存储的视频按照抓取时间进行排序后推荐给所述登录的用户，也可以是按照视频的热度进行排序后推荐给所述登录的用户，还可以是根据所述视频的抓取时间和热度二者的综合排序来向所述登录的用户推荐已经存储的视频的链接。向登录的用户推荐视频，充分考虑了用户的需求，可以增强用户体验，更方便用户。

本申请实施例还提供一种在其上记录有配置为执行上述方法的程序的计算机可读记录介质。

所述计算机可读记录介质包括用于以计算机（例如计算机）可读的形式存储或传送信息的任何机制。例如，机器可读介质包括只读存储器（ROM）、随机存取存储器（RAM）、磁盘存储介质、光存储介质、闪速存储介质、电、光、声或其他形式的传播信号（例如，载波、红外信号、数字信号等）等。

请参考图 2，本申请一种网页视频抓取装置实施例的结构示意图，应用于服务器端。本申请实施例提供一种网页视频抓取装置包括记录模块 21、评估模块 23 和视频抓取模块 25，其中：

所述记录模块配置为收集各个频道的相关信息，根据所述收集的各个频道的相关信息计算各个频道的热度，并收集各个频道的视频的发布时间，建立视频更新的时间表；

所述评估模块配置为根据所述各个频道的热度和所述视频更新的时间表计算各个频道的视频下次更新时间，并根据所述各个频道的视频下次更新时间确定所述各个频道的视频的抓取时间；

所述视频抓取模块配置为按照所述确定的各个频道的视频的抓取时间抓取所述各个频道的视频。

本申请实施例提供一种网页视频抓取装置，针对用户的各个频道的热度和各个频道的视频的更新时间，计算各个视频的下次更新时间，并根据所述各个频道的视频的下次更新时间确定所述各个频道视频的抓取时间，进而根据所述各个视频的抓取时间来抓取所述各个频道的视频，视频抓取的时效性高，而且不会存在现有技术的轮询的技术方案，本申请实施例所提供的方案在抓取网页视频时只在确定好的视频抓取时间去网页上抓取视频，该方案极大的节约了网络资源。

具体地，根据本申请一种网页视频抓取装置实施例所公开的技术方案，

所述记录模块 21 收集各个频道的相关信息包括：所述记录模块 21 收集各个频道的视频数、订阅数和播放总数；所述根据所述收集的各个频道的相关信息计算各个频道的热度包括所述根据所述收集的各个频道的视频数、订阅数和播放总数计算各个频道的热度。

5 具体地，根据本申请一种网页视频抓取装置实施例所公开的技术方案，所述记录模块 21 收集各个频道的视频的发布时间，建立视频更新的时间表，包括：所述记录模块 21 收集各个频道的视频的发布时间，计算各个频道的视频每次更新的时间间隔和更新的时间段，建立所述视频更新的时间表。

具体地，根据本申请一种网页视频抓取装置实施例所公开的技术方案，
10 所述评估模块 23 根据所述各个频道的热度和所述视频更新的时间表计算各个频道的视频下次更新时间，包括：所述评估模块 23 根据所述各个频道的热度、所述视频每次更新的时间间隔和更新的时间段计算各个频道的视频下次更新时间。

根据本申请一种网页视频抓取装置实施例所公开的技术方案，所述记录
15 模块 21 具体配置为收集各个频道视频的视频的抓取时间，在所述视频更新的时间表记录所述视频的抓取时间；并计算所述视频的发布时间与所述视频的抓取时间的间隔；则，所述评估模块 23 根据所述各个频道的热度和所述视频更新的时间表计算各个频道的视频下次更新时间，还包括：所述评估模块 23 根据所述各个频道的热度、所述视频每次更新的时间间隔、更新的时间段及
20 所述视频的发布时间与所述视频的抓取时间的间隔计算各个频道的视频下次更新时间。

具体地，根据本申请一种网页视频抓取装置实施例所公开的技术方案，在所述视频抓取模块 25 按照所述确定的各个频道的视频的抓取时间抓取所述各个频道的视频之后，所述方法包括所述视频抓取模块 25 存储所述抓取的
25 各个频道的视频的信息。

具体地，根据本申请一种网页视频抓取装置实施例所公开的技术方案，在所述视频抓取模块 25 存储所述抓取的各个频道的视频的信息之后，所述视频抓取模块 25 向用户推荐所述抓取的各个频道的视频。

本申请实施例所述的一种网页视频抓取的方法可以通过硬件处理器来实现
30 相关功能。所述硬件处理器在执行本申请实施例提供的一种网页视频抓取

的方法，可以包括如下处理：步骤 11，收集各个频道的相关信息，根据所述收集的各个频道的相关信息计算各个频道的热度；步骤 13，收集各个频道的视频的发布时间，建立视频更新的时间表；步骤 15，根据所述各个频道的热度和所述视频更新的时间表计算各个频道的视频下次更新时间；步骤 17，根据所述各个频道的视频下次更新时间确定所述各个频道的视频的抓取时间；步骤 19，所述按照所述确定的各个频道的视频的抓取时间抓取所述各个频道的视频。

本申请实施例提供的一种网页视频抓取装置，针对用户的各个频道的热度和各个频道的视频的更新时间，计算各个视频的下次更新时间，并根据所述各个频道的视频的下次更新时间确定所述各个频道视频的抓取时间，进而根据所述各个视频的抓取时间来抓取所述各个频道的视频，视频抓取的时效性高，而且不会存在现有技术的轮询的技术方案，本申请实施例所提供的方案在抓取网页视频时只在确定好的视频抓取时间去网页上抓取视频，该方案极大的节约了网络资源。

以上所描述的装置实施例仅仅是示意性的，其中所述作为分离部件说明的模块可以是或者也可以不是物理上分开的，作为模块显示的部件可以是或者也可以不是物理模块，即可以位于一个地方，或者也可以分布到多个网络模块上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性的劳动的情况下，即可以理解并实施。

通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到各实施方式可借助软件加必需的通用硬件平台的方式来实现，当然也可以通过硬件。基于这样的理解，上述技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品可以存储在计算机可读存储介质中，如 ROM/RAM、磁碟、光盘等，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备等）执行各个实施例或者实施例的某些部分所述的方法。

最后应说明的是：以上实施例仅用以说明本申请的技术方案，而非对其限制；尽管参照前述实施例对本申请进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或

者对其部分技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围。

权利要求书

1、一种网页视频抓取的方法，其特征在于，包括：

收集各个频道的相关信息，根据所述收集的各个频道的相关信息计算各个频道的热度；

5 收集各个频道的视频的发布时间，建立视频更新的时间表；

根据所述各个频道的热度和所述视频更新的时间表计算各个频道的视频下次更新时间；

根据所述各个频道的视频下次更新时间确定所述各个频道的视频的抓取时间；

10 按照所述确定的各个频道的视频的抓取时间抓取所述各个频道的视频。

2、根据权利要求1所述的方法，其特征在于，

所述收集各个频道的相关信息包括：所述收集各个频道的视频数、订阅数和播放总数；

15 所述根据所述收集的各个频道的相关信息计算各个频道的热度包括所述根据所述收集的各个频道的视频数、订阅数和播放总数，计算各个频道的热度。

3、根据权利要求1所述的方法，其特征在于，

20 所述收集各个频道的视频的发布时间，建立视频更新的时间表，包括：收集各个频道的视频的发布时间，计算各个频道的视频每次更新的时间间隔和更新的时间段，建立所述视频更新的时间表。

4、根据权利要求1-3任一权利要求所述的方法，其特征在于，

所述根据所述各个频道的热度和所述视频更新的时间表计算各个频道的视频下次更新时间，包括：根据所述各个频道的热度、所述视频每次更新的时间间隔和更新的时间段计算各个频道的视频下次更新时间。

25 5、根据权利要求4所述的方法，其特征在于，

所述方法还包括：收集各个频道视频的视频的抓取时间，在所述视频更新的时间表记录所述视频的抓取时间；所述计算所述视频的发布时间与所述视频的抓取时间的间隔；

30 所述根据所述各个频道的热度和所述视频更新的时间表计算各个频道的视频下次更新时间，还包括：根据所述各个频道的热度、所述视频每次更新

的时间间隔、更新的时间段及所述视频的发布时间与所述视频的抓取时间的间隔计算各个频道的视频下次更新时间。

6、根据权利要求 1-3 和 5 任一权利要求所述的方法，其特征在于，

在所述按照所述确定的各个频道的视频的抓取时间抓取所述各个频道的视频之后，所述方法包括存储所述抓取的各个频道的视频的信息。

7、根据权利要求 4 所述的方法，其特征在于，

在所述按照所述确定的各个频道的视频的抓取时间抓取所述各个频道的视频之后，所述方法包括所述存储所述抓取的各个频道的视频的信息。

8、根据权利要求 7 所述的方法，其特征在于，

在所述存储所述抓取的各个频道的视频的信息之后，所述方法包括向用户推荐所述抓取的各个频道的视频。

9、一种网页视频抓取装置，其特征在于，包括：

记录模块，配置为收集各个频道的相关信息，根据所述收集的各个频道的相关信息计算各个频道的热度，并收集各个频道的视频的发布时间，建立视频更新的时间表；

评估模块，配置为根据所述各个频道的热度和所述视频更新的时间表计算各个频道的视频下次更新时间，并根据所述各个频道的视频下次更新时间确定所述各个频道的视频的抓取时间；

视频抓取模块，配置为按照所述确定的各个频道的视频的抓取时间抓取所述各个频道的视频。

10、根据权利要求 9 所述的装置，其特征在于，

所述记录模块收集各个频道的相关信息包括：所述记录模块收集各个频道的视频数、订阅数和播放总数；

所述根据所述收集的各个频道的相关信息计算各个频道的热度包括所述根据所述收集的各个频道的视频数、订阅数和播放总数计算各个频道的热度。

11、根据权利要求 9 所述的装置，其特征在于，

所述记录模块收集各个频道的视频的发布时间，建立视频更新的时间表，包括：所述记录模块收集各个频道的视频的发布时间，计算各个频道的视频每次更新的时间间隔和更新的时间段，建立所述视频更新的时间表。

12、根据权利要求 9-11 任一权利要求所述的装置，其特征在于，

所述评估模块根据所述各个频道的热度和所述视频更新的时间表计算各个频道的视频下次更新时间，包括：所述评估模块根据所述各个频道的热度、所述视频每次更新的时间间隔和更新的时间段计算各个频道的视频下次更新时间。

5 13、根据权利要求 12 所述的装置，其特征在于，

所述记录模块具体配置为收集各个频道视频的视频的抓取时间，在所述视频更新的时间表记录所述视频的抓取时间；并计算所述视频的发布时间与所述视频的抓取时间的间隔；

10 所述评估模块根据所述各个频道的热度和所述视频更新的时间表计算各个频道的视频下次更新时间，还包括：所述评估模块根据所述各个频道的热度、所述视频每次更新的时间间隔、更新的时间段及所述视频的发布时间与所述视频的抓取时间的间隔计算各个频道的视频下次更新时间。

14、根据权利要求 9-11 和 13 任一权利要求所述的装置，其特征在于，

15 在所述视频抓取模块按照所述确定的各个频道的视频的抓取时间抓取所述各个频道的视频之后，还包括所述视频抓取模块存储所述抓取的各个频道的视频的信息。

15、根据权利要求 12 所述的，其特征在于，

20 在所述视频抓取模块按照所述确定的各个频道的视频的抓取时间抓取所述各个频道的视频之后，所述视频抓取模块存储所述抓取的各个频道的视频的信息。

16、根据权利要求 15 所述的，其特征在于，

在所述视频抓取模块存储所述抓取的各个频道的视频的信息之后，所述视频抓取模块向用户推荐所述抓取的各个频道的视频。

25 17、一种在其上记录有配置为执行权利要求 1-8 任一项所述方法的程序的计算机可读记录介质。

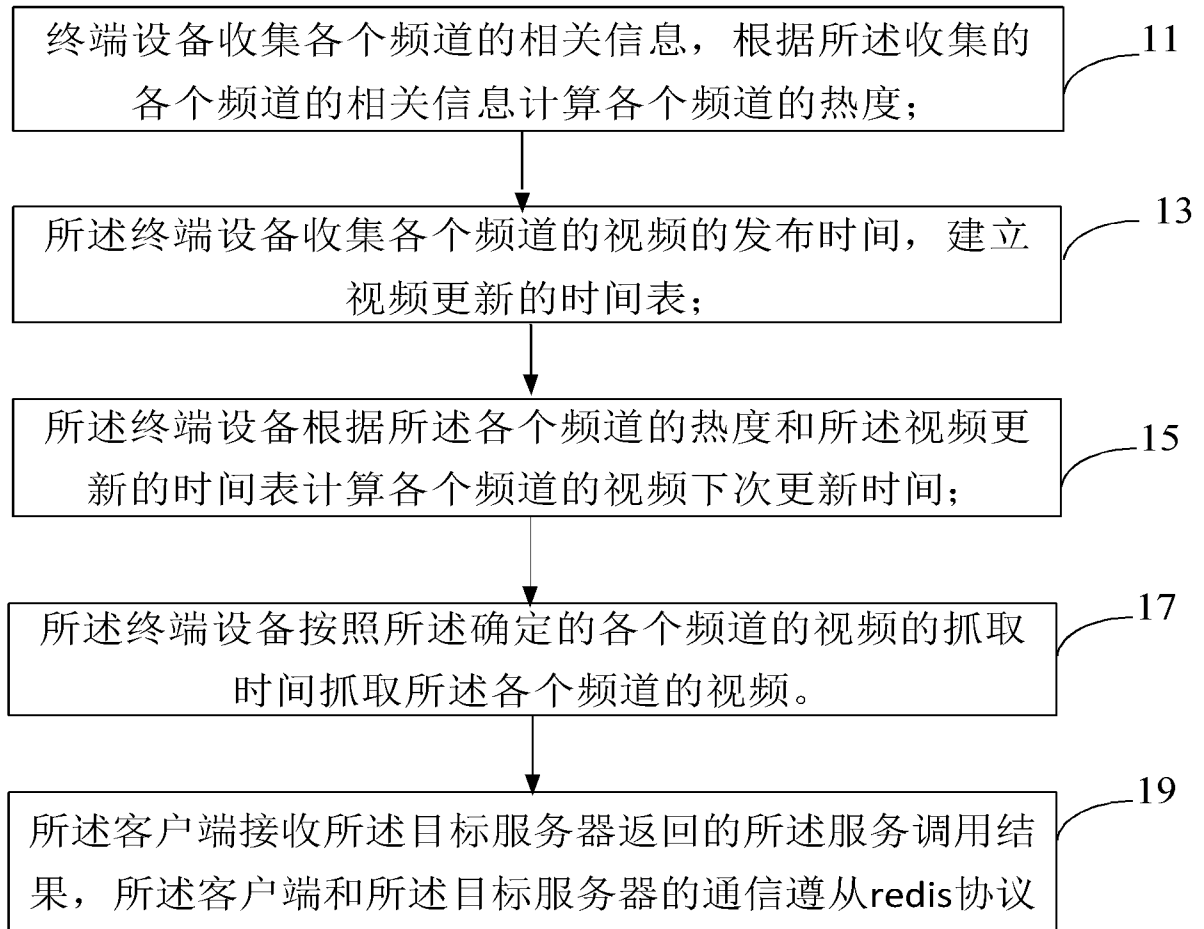


图 1



图 2

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/088955

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, CNKI, WPI, EPODOC: hot, retrieve, video?, crawl+, grab+, captur+, website, webpage, web, internet, updat+, detect+, monitor+, channel, search+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PY	CN 105528376 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.), 27 April 2016 (27.04.2016), description, paragraphs [0068]-[0141]	1-17
Y	CN 104361005 A (BEIJING ZHONGSOU NETWORK TECHNOLOGIES CO., LTD.), 18 February 2015 (18.02.2015), description, paragraphs [0039] and [0064]-[0095]	1-17
Y	CN 104063448 A (EAST CHINA NORMAL UNIVERSITY), 24 September 2014 (24.09.2014), description, paragraphs [0012]-[0016]	1-17
A	CN 102033929 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.), 27 April 2011 (27.04.2011), the whole document	1-17
A	CN 104765766 A (BEIJING QIYI CENTURY SCIENCE AND TECHNOLOGY CO., LTD.), 08 July 2015 (08.07.2015), the whole document	1-17
A	US 2013013583 A1 (YU, Lei et al.), 10 January 2013 (10.01.2013), the whole document	1-17

☐ 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 August 2016 (15.08.2016)	Date of mailing of the international search report 14 September 2016 (14.09.2016)
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 WANG, Xiaoyan Telephone No.: (86-10) 010-62413916

International application No.
PCT/CN2016/088955

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, CNKI, WPI, EPDOC: 视频, 抓取, 网页, 网站, 热度, 更新, 检测, 监视, 监测, 频道, 搜索, 检索, video?, crawl+, grab+, captur+, website, webpage, web, internet, updat+, detect+, monitor+, channel, search+		
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
PY	CN 105528376 A (腾讯科技深圳有限公司) 2016年 4月 27日 (2016 - 04 - 27) 说明书第[0068]-[0141]段	1-17
Y	CN 104361005 A (北京中搜网络技术股份有限公司) 2015年 2月 18日 (2015 - 02 - 18) 说明书第[0039]、[0064]-[0095]段	1-17
Y	CN 104063448 A (华东师范大学) 2014年 9月 24日 (2014 - 09 - 24) 说明书第[0012]-[0016]段	1-17
A	CN 102033929 A (百度在线网络技术北京有限公司) 2011年 4月 27日 (2011 - 04 - 27) 全文	1-17
A	CN 104765766 A (北京奇艺世纪科技有限公司) 2015年 7月 8日 (2015 - 07 - 08) 全文	1-17
A	US 2013013583 A1 (YU, LEI 等) 2013年 1月 10日 (2013 - 01 - 10) 全文	1-17
☐ 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		国际检索报告邮寄日期
2016年 8月 15日		2016年 9月 14日
ISA/CN的名称和邮寄地址		授权官员
中华人民共和国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		王晓燕 电话号码 (86-10)010-62413916

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2016/088955

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	105528376	A	2016年 4月 27日	无	
CN	104361005	A	2015年 2月 18日	无	
CN	104063448	A	2014年 9月 24日	无	
CN	102033929	A	2011年 4月 27日	无	
CN	104765766	A	2015年 7月 8日	无	
US	2013013583	A1	2013年 1月 10日	US 2015058998 A1	2015年 2月 26日