

(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  
2015年4月2日 (02.04.2015)



(10) 国际公布号  
WO 2015/043389 A1

- (51) 国际专利分类号:  
G06F 17/30 (2006.01) H04L 29/08 (2006.01)  
H04L 29/12 (2006.01)
- (21) 国际申请号: PCT/CN2014/086519
- (22) 国际申请日: 2014年9月15日 (15.09.2014)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:  
201310462461.6 2013年9月30日 (30.09.2013) CN  
201310462768.6 2013年9月30日 (30.09.2013) CN  
201310462214.6 2013年9月30日 (30.09.2013) CN
- (71) 申请人: 北京奇虎科技有限公司 (BEIJING QIHOO TECHNOLOGY COMPANY LIMITED) [CN/CN]; 中国北京市西城区新街口外大街28号D座112室(德胜园区), Beijing 100088 (CN)。奇智软件(北京)有限公司 (QIZHI SOFTWARE (BEIJING) COMPANY LIMITED) [CN/CN]; 中国北京市朝阳区酒仙桥路6号院2号楼B座2层、3层301-306室, Beijing 100015 (CN)。
- (72) 发明人: 崔代超 (CUI, Daichao); 中国北京市朝阳区酒仙桥路6号院2号楼, Beijing 100015 (CN)。
- (74) 代理人: 北京润泽恒知识产权代理有限公司 (BEIJING RISEHIGH INTELLECTUAL PROPERTY LAW FIRM); 中国北京市海淀区中关村南大街31号神舟大厦702, Beijing 100081 (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。

[见续页]

(54) Title: PARTICIPLE INFORMATION PUSH METHOD AND DEVICE BASED ON VIDEO SEARCH

(54) 发明名称: 一种基于视频搜索的分词信息推送方法和装置

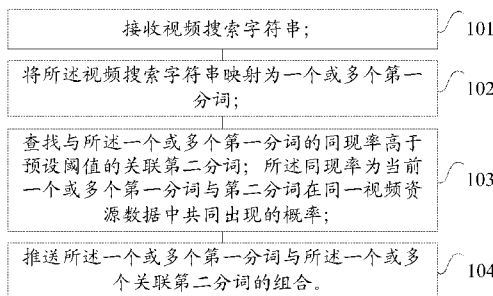


图1 / Fig. 1

- 101 RECEIVING A VIDEO SEARCH CHARACTER STRING;
- 102 MAPPING THE VIDEO SEARCH CHARACTER STRING INTO ONE OR MORE FIRST PARTICIPLES;
- 103 SEARCHING FOR ASSOCIATED SECOND PARTICIPLES OF WHICH THE CO-OCCURRENCE RATE TOGETHER WITH THAT OF THE ONE OR MORE FIRST PARTICIPLES IS HIGHER THAN A PRESET THRESHOLD VALUE, THE CO-OCCURRENCE RATE BEING THE PROBABILITY THAT THE CURRENT ONE OR MORE FIRST PARTICIPLES AND THE SECOND PARTICIPLES APPEAR IN THE SAME VIDEO RESOURCE DATA TOGETHER
- 104 PUSHING A COMBINATION OF THE ONE OR MORE FIRST PARTICIPLES AND THE ONE OR MORE ASSOCIATED SECOND PARTICIPLES.

(57) Abstract: Disclosed are a participle information push method and device based on a video search, a push method for an on-line playing entry object based on a video search and a push method for an associated resource address based on a video search. The method comprises: mapping feature text information about a received video search character string or acquired video resource data into one or more first participles; and searching for associated second participles of which the co-occurrence rate together with that of the one or more first participles is higher than a preset threshold value, the co-occurrence rate being the probability that the current one or more first participles and the second participles appear in the same video resource data together, thereby pushing a combination of the one or more first participles and the one or more associated second participles or an entry object of a video resource data related thereto or a network connection address associated with the video resource data. In the present application, data of the video resources which are seldom searched by users but have multiple relative resources in a video database are pushed, good-quality resources in the video database are deeply dug, resource digging efficiency is improved. An index table will be continuously expanded along with the constant accumulation of Internet video contents, and the number and scope of contents produced by various video stations will be far beyond the number of words already searched by users, thereby being beneficial to expand the recall rate.

(57) 摘要:

[见续页]

WO 2015/043389 A1



(84) **指定国** (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ,

CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

**本国际公布:**

— 包括国际检索报告(条约第 21 条(3))。

---

本申请公开了一种基于视频搜索的分词信息推送方法和装置、基于视频搜索的在线播放入口对象的推送方法以及基于视频搜索的关联资源地址的推送方法, 其中将接收的视频搜索字符串或者获取的视频资源数据的特征本文本信息映射为一个或多个第一分词; 查找与所述一个或多个第一分词的同现率高于预设阈值的关联第二分词; 所述同现率为当前一个或多个第一分词与第二分词在同一视频资源数据中共同出现的概率, 从而推送所述一个或多个第一分词与所述一个或多个关联第二分词的组合或与之相关的视频数据资源的入口对象或关联视频资源数据的网络连接地址。本申请将虽然比较少有用户搜索的但视频库汇总已有较多相关资源的视频资源数据推送出来, 从而实现深度挖掘视频库中的优质资源, 提高了资源挖掘的效率; 索引表会随着互联网视频内容的不断积累而不断扩大, 各大视频站生产出来的内容数量和广度会远远超过用户已经搜索过的词数, 有利于扩大召回率。

# 说明书

## 一种基于视频搜索的分词信息推送方法和装置

### 技术领域

本发明涉及互联网的技术领域，尤其涉及一种基于视频搜索的分词信息推送方法和一种基于视频搜索的分词信息推送装置。

### 5 背景技术

视频搜索引擎是有别于综合搜索的一种垂直搜索技术。视频搜索引擎抓取互联网中的视频类的结果并建立索引，由于它可以向搜索者提供纯粹的视频类结果，从而可以大大节省网民寻找视频的时间。

10 根据视频搜索的相关统计数据显示，娱乐、游戏、影视、新闻、动漫等类型的视频是用户的主要搜索对象。这表明用户对于视频搜索本身具有泛需求的性质。用户往往不带有很强的目的性，搜索结果并非“非彼不可”，而是带有一定扩展性，只要目标在用户所喜欢的范畴内即可。因此，往往会在搜索结果之外对用户进行相关推荐是。

15 但是，现有的视频搜索引擎在相关推荐方面做得还有不足：部分视频搜索引擎没有相关推荐，有相关推荐的视频搜索引擎只是根据用户的搜索历史数据、通过人工整理得到关联体系等简单方式实现推荐。这种推荐系统基于用户已有的搜索习惯，召回率较低，另外由于用户的搜索范围一般会比现有互联网中的资源范围要小很多，不能充分挖掘互联网中的优质视频。

20 另一种搜索推荐方法是依靠人工整理出一个资源关联体系或从其他知识体系中得到这样的体系，应用到推荐系统中。例如在某搜索引擎搜索“广场舞”时，会得到“交谊舞”、“肚皮舞”、“健身操”等的推荐词，搜索“dota”时会得到“穿越火线”、“魔兽世界”等的推荐词，但是这种体系召回率较低，在长尾的搜索中一般不能给出推荐。

### 发明内容

鉴于上述问题，提出了本发明以便提供一种克服上述问题或者至少部分地解决上述问题的一种基于视频搜索的分词信息推送方法和相应的一种基于视频搜索的分词信息推送装置。

25 根据本发明的一个方面，提供了一种基于视频搜索的分词信息推送方法，包括：  
接收视频搜索字符串；  
将所述视频搜索字符串映射为一个或多个第一分词；  
30 查找与所述一个或多个第一分词的同现率高于预设阈值的关联第二分词；所述同现率为当前一个或多个第一分词与第二分词在同一视频资源数据中共同出现的概率；  
推送所述一个或多个第一分词与所述一个或多个关联第二分词的组合。

根据本发明的另一方面，提供了一种基于视频搜索的在线播放入口对象的推送方法，包括：

35 接收视频搜索字符串；  
将所述视频搜索字符串映射为一个或多个第一分词；  
查找与所述一个或多个第一分词的同现率高于预设阈值的关联第二分词；所述同现率为当前一个或多个第一分词与第二分词在同一视频资源数据中共同出现的概率；  
40 获取与所述一个或多个第一分词和所述关联第二分词匹配的一个或多个视频数据资源的网络地址；

根据所述一个或多个视频数据资源网络地址构造在线播放所述视频数据资源的入口对象；  
推送所述一个或多个在线播放视频数据资源的入口对象。

45 根据本发明的另一方面，提供了一种基于视频搜索的关联资源地址的推送方法，包括：  
当接收到第一视频资源数据的加载或播放请求时，获取所述第一视频资源数据的特征文本信息；

将所述特征文本信息映射为一个或多个第一分词；  
查找与所述一个或多个第一分词的同现率高于预设阈值的关联第二分词；所述同现率为当前一个或多个第一分词与第二分词在同一视频资源数据中共同出现的概率；

获取与所述一个或多个第一分词和所述关联第二分词匹配的第二视频资源数据的网络链接地址；

推送所述第二视频资源数据的网络链接地址。

- 5 根据本发明的另一方面，提供了一种基于视频搜索的分词信息推送装置，包括：  
视频搜索字符串接收模块，适于接收视频搜索字符串；  
第一分词映射模块，适于将所述视频搜索字符串映射为一个或多个第一分词；  
第二分词查找模块，适于查找与所述一个或多个第一分词的同现率高于预设阈值的关联  
10 第二分词；所述同现率为当前一个或多个分词与第二分词在同一视频资源数据中共同出现的概率；  
组合推送模块，适于推送所述一个或多个第一分词与所述一个或多个关联第二分词的组合。

15 根据本发明的又一个方面，提供了一种计算机程序，其包括计算机可读代码，当所述计算机可读代码在计算设备上运行时，导致所述计算设备执行根据权利要求 1-7 中的任一个所述的基于视频搜索的分词信息推送方法。

根据本发明的再一个方面，提供了一种计算机可读介质，其中存储了如权利要求 17 所述的计算机程序。

本发明的有益效果为：

20 本发明可以根据现有已发布内容进行推送，使搜索引擎摆脱对用户搜索习惯的依赖，将虽然比较少有用户搜索的但视频库汇总已有较多相关资源的视频资源数据推送出来，从而实现深度挖掘视频库中的优质资源，提高了资源挖掘的效率；此外，索引表会随着互联网视频内容的不断积累而不断扩大，各大视频站生产出来的内容数量和广度会远远超过用户已经搜索过的词数，有利于扩大召回率。

25 本发明通过推送第一分词和第二分词的组合，用户可以基于此组合直接进行更多层次的搜索，使用户简单搜索即可获得更多的结果，无需多次提交搜索，从而减轻了访问服务器的负担，减少了网络资源的占用，并提升了用户体验。

30 本发明可以根据现有已发布内容进行推送，使搜索引擎摆脱对用户搜索习惯的依赖，将虽然比较少有用户搜索的但视频库汇总已有较多相关资源的视频资源数据推送出来，从而实现深度挖掘视频库中的优质资源，提高了资源挖掘的效率；此外，索引表会随着互联网视频内容的不断积累而不断扩大，各大视频站生产出来的内容数量和广度会远远超过用户已经搜索过的词数，有利于扩大召回率。

35 本发明可以根据现有已发布内容进行推送，使搜索引擎摆脱对用户搜索习惯的依赖，将虽然比较少有用户搜索的但视频库汇总已有较多相关资源的视频资源数据推送出来，从而实现深度挖掘视频库中的优质资源，提高了资源挖掘的效率；此外，索引表会随着互联网视频内容的不断积累而不断扩大，各大视频站生产出来的内容数量和广度会远远超过用户已经搜索过的词数，有利于扩大召回率。

上述说明仅是本发明技术方案的概述，为了能够更清楚了解本发明的技术手段，而可依照说明书的内容予以实施，并且为了让本发明的上述和其它目的、特征和优点能够更明显易懂，以下特举本发明的具体实施方式。

#### 40 附图说明

通过阅读下文优选实施方式的详细描述，各种其他的优点和益处对于本领域普通技术人员将变得清楚了。附图仅用于示出优选实施方式的目的，而并不认为是对本发明的限制。而且在整个附图中，用相同的参考符号表示相同的部件。在附图中：

45 图 1 示意性示出了根据本发明一个实施例的一种基于视频搜索的分词信息推送方法实施例的步骤流程图；

图 2 示意性示出了根据本发明一个实施例的一种基于视频搜索的在线播放入口对象的推送方法实施例的步骤流程图；

图 3 示意性示出了根据本发明一个实施例的一种基于视频搜索的关联资源地址的推送实施例的步骤流程图；

图 4 示意性示出了根据本发明一个实施例的一种基于视频搜索的分词信息推送装置实施例的结构框图；

图 5 示意性地示出了用于执行根据本发明的方法的计算设备的框图；以及

图 6 示意性地示出了用于保持或者携带实现根据本发明的方法的程序代码的存储单元。

## 5 具体实施方式

下面结合附图和具体的实施方式对本发明作进一步的描述。

参照图 1，示出了根据本发明一个实施例的一种基于视频搜索的分词信息推送方法实施例的步骤流程图，具体可以包括如下步骤：

步骤 101，接收视频搜索字符串；

需要说明的是，视频搜索字符串可以是用户输入的视频搜索信息，可以用于请求搜索与之相关的视频数据资源。

在实际应用中，视频搜索字符串可以是单词，即包括一个语义独立的词，例如中秋、端午、国庆等等；视频搜索字符串也可以是复合词，即包括两个或两个以上语义独立的词，例如中秋月饼、端午粽子、国庆西藏旅游等等。

步骤 102，将所述视频搜索字符串映射为一个或多个第一分词；

需要说明的是，被映射的分词可以是预先设置的，可以用于计算不同分词之间的同现率。

映射的规则也可以是预先设置的一个或多个，可以包括去除视频搜索字符串的脏词、修饰词、语气助词、宽泛词等无实际意义的词语；可以包括设定停止词，即一些常见的词，为拆分词组时停止的标准，例如的、我、你等等；还可以包括关联关系的对应，将同一事物的多种表达对应为一种表达，例如将八月十五、中秋节、月饼节等关联为中秋；还可以包括其他映射规则，本发明实施例对此不加以限制。

英文是以词为单位的，词和词之间是靠空格隔开，而中文是以字为单位，句子中所有的字连起来才能描述一个意思。例如，英文句子 I am a student，用中文则为：“我是一个学生”。计算机可以很简单通过空格知道 student 是一个单词，但是不能很容易明白“学”、“生”两个字合起来才表示一个词。把中文的汉字序列切分成有意义的词，就是中文分词。例如，我是一个学生，分词的结果是：我、是、一个、学生。

下面介绍一些常用的分词方法：

1、基于字符串匹配的分词方法：是指按照一定的策略将待分析的汉字串与一个预置的机器词典中的词条进行匹配，若在词典中找到某个字符串，则匹配成功（识别出一个词）。实际使用的分词系统，都是把机械分词作为一种初分手段，还需通过利用各种其它的语言信息来进一步提高切分的准确率。

2、基于特征扫描或标志切分的分词方法：是指优先在待分析字符串中识别和切分出一些带有明显特征的字，以这些字作为断点，可将原字符串分为较小的串再来进机械分词，从而减少匹配的错误率；或者将分词和词类标注结合起来，利用丰富的词类信息对分词决策提供帮助，并且在标注过程中又反过来对分词结果进行检验、调整，从而提高切分的准确率。

3、基于理解的分词方法：是指通过让计算机模拟人对句子的理解，达到识别词的效果。其基本思想就是在分词的同时进行句法、语义分析，利用句法信息和语义信息来处理歧义现象。它通常包括三个部分：分词子系统、句法语义子系统、总控部分。在总控部分的协调下，分词子系统可以获得有关词、句子等的句法和语义信息来对分词歧义进行判断，即它模拟了人对句子的理解过程。这种分词方法需要使用大量的语言知识和信息。

4、基于统计的分词方法：是指，中文信息中由于字与字相邻共现的频率或概率能够较好的反映成词的可信度，所以可以对语料中相邻共现的各个字的组合的频度进行统计，计算它们的互现信息，以及计算两个汉字 X、Y 的相邻共现概率。互现信息可以体现汉字之间结合关系的紧密程度。当紧密程度高于某一个阈值时，便可认为此字组可能构成了一个词。这种方法只需对语料中的字组频度进行统计，不需要切分词典。

在本发明的一种优选实施例中，所述步骤 102 具体可以包括如下子步骤：

子步骤 S11，提取所述视频搜索字符串所映射的一个分词；

对于视频搜索字符串为单词的情形，可以按照预设的映射规则直接提取其对应的分词。例如，视频搜索字符串为“中秋节”、“我的中秋节”或者“中秋节了”等，映射的第一分词都

可以为“中秋”。当然，视频搜索字符串也可以与其映射的第一分词是同一个词，例如视频搜索字符串为“中秋”，映射的第一分词也可以“中秋”。

或者，

5 子步骤 S12，当接收到的视频搜索字符串为复合词时，将所述视频搜索字符串拆分为多个搜索子词；

子步骤 S13，提取所述多个搜索子词所映射的多个分词。

10 对于视频搜索字符串为复合词的情形，可以按照预设的映射规则进行分词，得到搜索子词，然后分别提取搜索子词对应的分词。例如，接收到的视频搜索字符串为“中秋节月饼”，可以将其拆分为“中秋节”和“月饼”两个搜索子词，然后将“中秋节”映射为“中秋”，将“月饼”映射为“月饼”，得到“中秋”和“月饼”两个第一分词。

步骤 103，查找与所述一个或多个第一分词的同现率高于预设阈值的关联第二分词；

所述同现率为当前一个或多个第一分词与第二分词在同一视频资源数据中共同出现的概率；

15 需要说明的是，第二分词可以是在全部预设的分词中，除第一分词以外的分词。关联第二分词可以是与第一分词的同现率高于预设阈值的第二分词。

在实际应用中，视频资源数据可以包括特征文本信息，该特征文本信息可以用于记载该视频资源数据的相关信息，也可以用于提取分词。

在本发明的一种优选实施例中，所述特征文本信息可以包括视频标题、视频关键词和/或视频描述。

20 例如，在一段名为《【拍客】东莞暴雨后变威尼斯，千余辆车水浸抛锚—在线播放—XX网，视频高清在线观看》的视频资源数据中，其特征文本信息可以如下：

视频标题 (Title): 【拍客】东莞暴雨后变威尼斯，千余辆车水浸抛锚—在线播放—XX网，  
视频高清在线观看；

视频关键词 (Keywords): YY 记者 生活资讯 东莞 水浸；

25 视频描述 (Description): 昨天上午的一场暴雨，让东莞部分地区的街坊瞬间感到好像来到了威尼斯。行驶中的小车在暴雨中遭到水浸抛锚，有的街坊家中也是一片汪洋。

具体而言，同现率可以为当前一个或多个分词与第二分词在同一视频资源数据的特征文本信息中共同出现的概率，具体可以包括一个第一分词和第二分词的同现率，多个分词和第二分词的同现率。

30 在本发明的一种优选实施例中，所述步骤 103 具体可以包括如下子步骤：

子步骤 S21，当所述视频搜索字符串被映射为一个第一分词时，提取所述第一分词对应的预置索引表；其中，所述索引表包括所述第一分词所属的视频资源数据的信息，以及，所述视频资源数据中的所有分词；所述视频资源数据中的所有分词为通过抓取视频资源数据，提取所述视频资源数据的特征文本信息，对所述特征文本信息进行分词生成；

35 在具体实现中，可以预先采用搜索引擎通过爬虫抓取各个网站平台上的视频资源数据，然后建立索引库：提取视频资源数据的特征文本信息进行分词处理，并建立每个分词对应的索引表，该索引表中可以存储视频资源数据的信息（可以是 ID、内网地址、外网地址等等视频标识，也可以是一条由当前分词和其他分词所组成的记录）、视频资源数据中的所有分词（包括第一分词和除第一分词外的第二分词）。

40 在本发明的一种优选实施例中，所述特征文本信息可以包括视频标题、视频关键词和/或视频描述。

例如，“中秋”的索引表可以如下：

	中秋
视频标识	第二分词

视频 A	月亮 月饼 阳台 板栗 葡萄 灯笼 家
视频 B	冰淇淋 时代广场 街舞
视频 C	月饼 铁观音 柚子 葡萄 苹果 柿子 提子
视频 D	月饼 田螺 芋头 菱角
视频 E	花灯 舞龙
... ..	... ..

其中，第一分词为“中秋”，视频资源数据的信息包括视频标识。当然，视频资源数据的信息也可以不包括视频标识，而只有第一分词和第二分词所成的记录（即每一行的第二分词作为一条记录）。

当然，上述索引表只是作为示例，在实施本发明实施例时，可以根据实际情况设置其他索引表，本发明实施例对此不加以限制。另外，除了上述索引表外，本领域技术人员还可以根据实际需要采用其他索引表，本发明实施例对此也不加以限制。

需要说明的是，可以周期或者不定时抓取各个平台上的视频资源数据，然后更新索引建库，即更新各索引表。

子步骤 S22，计算所述第一分词与所述索引表中各个第二分词的同现率，所述同现率为所述索引表中各个第二分词出现的次数与所述索引表中视频资源数据的信息总数的比值；其中，所述第二分词为所述视频资源数据中的所有分词中除所述第一分词以外的分词；

由于索引表中各个第二分词出现的次数与其所属的视频资料数据的数量一样，同现率也可以表示为所述索引表中各个第二分词所属的视频资料数据的数量与所述索引表中视频资源数据的信息总数的比值。

例如，分词“广场舞”的索引表中总共有 100 条视频资源数据的信息，分词“兵哥哥”的索引表中总共有 200 条视频资源数据的信息，“广场舞”和“兵哥哥”同时出现在这两个索引表中的视频资源数据的信息共 10 条，则对于“广场舞”而言，“广场舞”与“兵哥哥”的同现率为  $10/100=10\%$ ，而对于“兵哥哥”而言，“兵哥哥”与“广场舞”的同现率为  $10/200=5\%$ 。

子步骤 S23，提取所述同现率高于预设阈值的第二分词作为关联第二分词。

具体实现中，预设阈值可以由本领域技术人员根据实际情况而设定的，本发明实施例对此不加以限制。本发明实施例中所提取的关联第二分词可以为空，也可以为一个或多个。

在本发明的一种优选实施例中，所述步骤 103 具体可以包括如下子步骤：

子步骤 S31，当所述视频搜索字符串被映射为多个第一分词时，分别提取所述多个第一分词对应的多个预置索引表；各个索引表中包括所述第一分词所属的视频资源数据的信息，以及，所述视频资源数据中的所有分词；所述视频资源数据中的所有分词为通过抓取视频资源数据，提取所述视频资源数据的特征文本信息，对所述特征文本信息进行分词生成；

在具体实现中，可以预先采用搜索引擎通过爬虫抓取各个平台上的视频资源数据，然后建立索引建库：提取视频资源数据的特征文本信息进行分词处理，并建立每个分词对应的索引表，该索引表中可以存储视频资源数据的信息（可以是 ID、内网地址、外网地址等等视频标识，也可以是一条由当前分词和其他分词所组成的记录）、视频资源数据中的所有分词（包括第一分词和除第一分词外的第二分词）。

在本发明的一种优选实施例中，所述特征文本信息可以包括视频标题、视频关键词和/或视频描述。

子步骤 S32，提取与所述多个第一分词共同出现的第二分词作为候选分词；其中，所述第二分词为所述视频资源数据中的所有分词中除所述第一分词以外的分词；

具体而言，当前有多个第一分词，即有多个数量对应的索引表，候选分词需要在各个索引表中出现，即候选分词分别与当前各第一分词都共同在同一索引表中出现。

子步骤 S33，分别在各个索引表中计算所述第一分词与所述候选分词的同现率，所述同现率为所述索引表中候选分词出现的次数与所述索引表中视频资源数据的信息总数的比值；

例如，可以将视频搜索字符串“中秋节月饼”映射为第一分词“中秋”和“月饼”，提取了其中一个候选分词为“月亮”，则可以分别计算“中秋”和“月亮”的同现率（假设为 70%）、“月饼”和“月亮”同现率（假设为 60%）。

子步骤 S34，分别为所述多个第一分词与所述候选分词的同现率配置对应的多个权重；

权重可以根据各第一分词间的索引表中视频资源数据的信息总数比例的进行确定，其中，

索引表中视频资源数据的信息总数越多其权重越大。例如，在“中秋”的索引表中视频资源数据的信息总数为 900，而在“月饼”的索引表中视频资源数据的信息总数为 100，则“中秋”和“月亮”的同现率的权重可以为 0.9，“月饼”和“月亮”同现率的权重可以为 0.1。

当然，上述权重只是作为示例，在实施本发明实施例时，可以根据实际情况设置其他权重，例如按照当前社会热点（新闻排名、微博排名等）设置对应的权重、按照用户的本地和/或网上操作行为（视频播放、新闻阅读等）设置对应的权重等等，本发明实施例对此不加以限制。另外，除了上述权重外，本领域技术人员还可以根据实际需要采用其他权重，本发明实施例对此也不加以限制。

子步骤 S35，分别计算多个配置了权重的同现率的平均值，作为所述多个第一分词与所述候选分词的同现率；

本发明实施例中，可以以多个同现率的加权平均值作为最终的同现率。

例如，“中秋”、“月饼”和“月亮”的同现率可以为  $(70\% \times 0.9 + 60\% \times 0.1) / 2 = 34.5\%$ 。

子步骤 S36，提取所述同现率高于预设阈值的候选分词作为关联第二分词。

具体实现中，预设阈值可以由本领域技术人员根据实际情况而设定的，本发明实施例对此不加以限制。本发明实施例中所提取的关联第二分词可以为空，也可以为一个或多个。

在本发明的一种优选实施例中，所述步骤 103 具体可以包括如下子步骤：

子步骤 S41，当所述视频搜索字符串被映射为多个第一分词时，分别提取所述多个第一分词对应的多个预置索引表；其中，各个索引表中包括所述第一分词所属的视频资源数据的信息，以及，所述视频资源数据中的所有分词；所述视频资源数据中的所有分词为通过抓取视频资源数据，提取所述视频资源数据的特征文本信息，对所述特征文本信息进行分词生成；

在具体实现中，可以预先采用搜索引擎通过爬虫抓取各个平台上的视频资源数据，然后建立索引建库：提取视频资源数据的特征文本信息进行分词处理，并建立每个分词对应的索引表，该索引表中可以存储视频资源数据的信息（可以是 ID、内网地址、外网地址等等视频标识，也可以是一条由当前分词和其他分词所组成的记录）、视频资源数据中的所有分词（包括第一分词和除第一分词外的第二分词）。

在本发明的一种优选实施例中，所述特征文本信息可以包括视频标题、视频关键词和/或视频描述。

子步骤 S42，采用所述多个索引表确定主分词，所述主分词为视频资源数据的信息总数最多的索引表对应的第一分词；

为了提高用户体验，对于视频资源数据相差比较悬殊的多个第一分词，可以忽略视频资源数据的信息总量少的第一分词。例如，对于视频搜索字符串“中秋节月饼”所映射的第一分词“中秋”和“月饼”，在“中秋”的索引表中视频资源数据的信息总数为 900，而在“月饼”的索引表中视频资源数据的信息总数为 100，则可以设置“中秋”作为主分词。

子步骤 S43，计算所述主分词与其对应的索引表中各个第二分词的同现率，所述同现率为所述索引表中各个第二分词出现的次数与所述索引表中视频资源数据的信息总数的比值；其中，所述第二分词为所述视频资源数据中的所有分词中除所述第一分词以外的分词；

本发明实施例中，可以以主分词的同现率的作为最终的同现率。

子步骤 S44，提取所述同现率高于预设阈值的第二分词作为关联第二分词。

具体实现中，预设阈值可以由本领域技术人员根据实际情况而设定的，本发明实施例对此不加以限制。本发明实施例中所提取的关联第二分词可以为空，也可以为一个或多个。

步骤 104，推送所述一个或多个第一分词与所述一个或多个关联第二分词的组合。

具体而言，在子步骤 S23 之后，可以在网页的输入框的下拉菜单等位置推送当前一个第一分词与一个或多个分词的组合。例如视频搜索字符串为“dota”，与其同现率较高的词为：“搞笑”、“蛋疼”、“2009”、“海啸”、“第一视角”和“经典”，同现率分别为 40%、35%、30%、25%、20%和 10%，则将依次推送组合“dota 搞笑”、“dota 蛋疼”、“dota2009”、“dota 海啸”、“dota 第一视角”和“dota 经典”。

在子步骤 S36 之后，可以在网页的输入框的下拉菜单等位置推送当前多个第一分词与一个或多个分词的组合。例如视频搜索字符串为“广场舞兵哥哥”，将其映射为第一分词“广场舞”和“兵哥哥”，提取与这两个第一分词同时出现的第二分词，例如第二分词“教学”，其可以作为关联第二分词，则最终推送组合“广场舞兵哥哥教学”。



在本发明的一种优选实施例中，步骤 104 具体可以包括如下子步骤：

子步骤 S51，推送所述主分词和所述关联第二分词。

在子步骤 S44 之后，可以在网页的输入框的下拉菜单等位置推送当前主分词与一个或多个分词的组合。例如，对于视频搜索字符串“中秋节月饼”所映射的第一分词“中秋”和“月  
5 饼”，可以设置“中秋”作为主分词，得到关联第二分词“月亮”，则可以推送组合“中秋月  
亮”。

用户可以通过点击下拉菜单中的推送组合，搜索新的视频资源数据。

本发明可以根据现有已发布内容进行推送，使搜索引擎摆脱对用户搜索习惯的依赖，将虽  
10 然比较少有用户搜索的但视频库汇总已有较多相关资源的视频资源数据推送出来，从而实现深  
度挖掘视频库中的优质资源，提高了资源挖掘的效率；此外，索引表会随着互联网视频内容的  
不断积累而不断扩大，各大视频站生产出来的内容数量和广度会远远超过用户已经搜索过的词  
数，有利于扩大召回率。

本发明通过推送第一分词和第二分词的组合，用户可以基于此组合直接进行更多层次的  
15 搜索，使用户简单搜索即可获得更多的结果，无需多次提交搜索，从而减轻了访问服务器的负  
担，减少了网络资源的占用，并提升了用户体验。

参照图 2，示出了根据本发明一个实施例的一种基于视频搜索的在线播放入口对象的推送  
方法实施例的步骤流程图，具体可以包括如下步骤：

步骤 201，接收视频搜索字符串；

20 步骤 202，将所述视频搜索字符串映射为一个或多个第一分词；

在本发明的一种优选实施例中，所述步骤 202 具体可以包括如下子步骤：

子步骤 S61，提取所述视频搜索字符串所映射的一个分词；

或者，

子步骤 S62，当接收到的视频搜索字符串为复合词时，将所述视频搜索字符串拆分为多个  
25 搜索子词；

子步骤 S63，提取所述多个搜索子词所映射的多个分词。

步骤 203，查找与所述一个或多个第一分词的同现率高于预设阈值的关联第二分词；

在本发明的一种优选实施例中，所述步骤 203 具体可以包括如下子步骤：

子步骤 S71，当所述视频搜索字符串被映射为一个第一分词时，提取所述第一分词对应的  
30 预置索引表；其中，所述索引表包括所述第一分词所属的视频资源数据的信息，以及，所述视  
频资源数据中的所有分词；所述视频资源数据中的所有分词为通过抓取视频资源数据，提取所  
述视频资源数据的特征文本信息，对所述特征文本信息进行分词生成；

子步骤 S72，计算所述第一分词与所述索引表中各个第二分词的同现率，所述同现率为所  
35 述索引表中各个第二分词出现的次数与所述索引表中视频资源数据的信息总数的比值；其中，  
所述第二分词为所述视频资源数据中的所有分词中除所述第一分词以外的分词；

子步骤 S73，提取所述同现率高于预设阈值的第二分词作为关联第二分词。

在本发明的一种优选实施例中，所述步骤 203 具体可以包括如下子步骤：

子步骤 S81，当所述视频搜索字符串被映射为多个第一分词时，分别提取所述多个第一分  
40 词对应的多个预置索引表；各个索引表中包括所述第一分词所属的视频资源数据的信息，以  
及，所述视频资源数据中的所有分词；所述视频资源数据中的所有分词为通过抓取视频资源数  
据，提取所述视频资源数据的特征文本信息，对所述特征文本信息进行分词生成；

在本发明的一种优选实施例中，所述特征文本信息可以包括视频标题、视频关键词和/或  
视频描述。

子步骤 S82，提取与所述多个第一分词共同出现的第二分词作为候选分词；其中，所述第  
45 二分词为所述视频资源数据中的所有分词中除所述第一分词以外的分词；

子步骤 S83，分别在各个索引表中计算所述第一分词与所述候选分词的同现率，所述同现  
率为所述索引表中候选分词出现的次数与所述索引表中视频资源数据的信息总数的比值；

子步骤 S84，分别为所述多个第一分词与所述候选分词的同现率配置对应的多个权重；

子步骤 S85，分别计算多个配置了权重的同现率的平均值，作为所述多个第一分词与所述  
50 候选分词的同现率；

子步骤 S86, 提取所述同现率高于预设阈值的候选分词作为关联第二分词。

在本发明的一种优选实施例中, 所述步骤 203 具体可以包括如下子步骤:

子步骤 S91, 当所述视频搜索字符串被映射为多个第一分词时, 分别提取所述多个第一分词对应的多个预置索引表; 其中, 各个索引表中包括所述第一分词所属的视频资源数据的信息, 以及, 所述视频资源数据中的所有分词; 所述视频资源数据中的所有分词为通过抓取视频资源数据, 提取所述视频资源数据的特征文本信息, 对所述特征文本信息进行分词生成;

在本发明的一种优选实施例中, 所述特征文本信息可以包括视频标题、视频关键词和/或视频描述。

子步骤 S92, 采用所述多个索引表确定主分词, 所述主分词为视频资源数据的信息总数最多的索引表对应的第一分词;

子步骤 S93, 计算所述主分词与其对应的索引表中各个第二分词的同现率, 所述同现率为所述索引表中各个第二分词出现的次数与所述索引表中视频资源数据的信息总数的比值; 其中, 所述第二分词为所述视频资源数据中的所有分词中除所述第一分词以外的分词;

子步骤 S94, 提取所述同现率高于预设阈值的第二分词作为关联第二分词。

步骤 204, 获取与所述一个或多个第一分词和所述关联第二分词匹配的一个或多个视频数据资源的网络地址;

在子步骤 S73 之后, 可以获得当前一个第一分词与一个或多个分词的组合。例如视频搜索字符串为“dota”, 与其同现率较高的词为: “搞笑”、“蛋疼”、“2009”、“海啸”、“第一视角”和“经典”, 同现率分别为 40%、35%、30%、25%、20%和 10%, 则获得的组合依次为“dota 搞笑”、“dota 蛋疼”、“dota2009”、“dota 海啸”、“dota 第一视角”和“dota 经典”。

在子步骤 S86 之后, 可以获得当前多个第一分词与一个或多个分词的组合。例如视频搜索字符串为“广场舞兵哥哥”, 将其映射为第一分词“广场舞”和“兵哥哥”, 提取与这两个第一分词同时出现的第二分词, 例如第二分词“教学”, 其可以作为关联第二分词, 则得到最终的组合“广场舞兵哥哥教学”。

在本发明的一种优选实施例中, 步骤 204 具体可以包括如下子步骤:

子步骤 S101, 获取与所述主分词和所述关联第二分词匹配的一个或多个视频数据资源的网络地址。

在子步骤 S94 之后, 可以获得当前主分词与一个或多个分词的组合。例如, 对于视频搜索字符串“中秋节月饼”所映射的第一分词“中秋”和“月饼”, 可以设置“中秋”作为主分词, 得到关联第二分词“月亮”, 则最终获得组合“中秋月亮”。

本发明实施例中, 可以基于一个或多个第一分词和第二分词的组合进行匹配的视频数据资源的搜索, 当搜索到时, 记录其网络地址, 具体可以是内网地址, 也可以是外网地址。

步骤 205, 根据所述一个或多个视频数据资源网络地址构造在线播放所述视频数据资源的入口对象;

入口对象可以为网页中链接到在线播放 URL 的图标或按钮。具体实现中, 可以在当前页面中配置一个图标或按钮, 在扩展窗口中与该视频数据资源网络地址相关联, 当用户点击该图标或按钮, 该视频数据资源网络地址被触发时, 可以从数据库的 URL 下加载对应的视频数据资源。

步骤 206, 推送所述一个或多个在线播放视频数据资源的入口对象。

实际应用中, 入口对象可以放置在当前页面的任一位置, 用户可以通过触发入口对象而触发该入口对象对应的视频数据资源的网络地址, 进而加载所述视频数据资源。

例如, 用户在搜索框中输入视频搜索字符串“钢铁”, 其本身可以作为第一分词, 获取第一分词和关联第二分词的组合“钢铁侠 3”匹配的视频数据资源, 该视频资源的入口对象为一图标, 该图标上写有“立即观看”提示用户, 当用户点击该图标时, 可以转到“钢铁侠 3”的播放页面。

本发明可以根据现有已发布内容进行推送, 使搜索引擎摆脱对用户搜索习惯的依赖, 将虽然比较少有用户搜索的但视频库汇总已有较多相关资源的视频资源数据推送出来, 从而实现深度挖掘视频库中的优质资源, 提高了资源挖掘的效率; 此外, 索引表会随着互联网视频内容的不断积累而不断扩大, 各大视频站生产出来的内容数量和广度会远远超过用户已经搜索过的词数, 有利于扩大召回率。

本发明通过推送在线播放视频数据资源的入口对象，用户可以基于此入口对象直接进行获取更多的视频搜索结果，使用户简单搜索即可获得更多的结果，无需多次提交搜索，从而减轻了访问服务器的负担，减少了网络资源的占用，并提升了用户体验。

5 参照图 3，示出了根据本发明一个实施例的一种基于视频搜索的关联资源地址的推送实施例的步骤流程图，具体可以包括如下步骤：

步骤 301，当接收到第一视频资源数据的加载或播放请求时，获取所述第一视频资源数据的特征本本文信息；

需要说明的是，第一视频资源数据可以位于终端设备上，也可以位于网络上，特征本本文信息可以是视频资源数据所携带的信息。

在本发明的一种优选实施例中，所述步骤 301 具体可以包括如下子步骤：

子步骤 S111，当接收到第一视频数据的播放请求时，接收当前终端发送的所述第一视频资源数据的特征本本文信息；

15 当第一视频资源数据位于终端设备上时，可以由终端设备提取第一视频资源数据的特征本本文信息，然后上传到对应的服务器侧。

或者，

子步骤 S112，当接收到第一视频数据加载请求时，提取本地预置的所述视频资源数据的特征本本文信息。

20 当第一视频资源数据位于网络上时，可以由服务器侧提取第一视频资源数据的特征本本文信息。

在本发明的一种优选实施例中，所述特征本本文信息可以包括视频标题、视频关键词和/或视频描述。

步骤 302，将所述特征本本文信息映射为一个或多个第一分词；

在本发明的一种优选实施例中，所述步骤 302 具体可以包括如下子步骤：

25 子步骤 S121，提取所述特征本本文信息所映射的一个分词；

或者，

子步骤 S122，当接收到的特征本本文信息为复合词时，将所述特征本本文信息拆分为多个搜索子词；

子步骤 S123，提取所述多个搜索子词所映射的多个分词。

30 步骤 303，查找与所述一个或多个第一分词的同现率高于预设阈值的关联第二分词；

所述同现率为当前一个或多个第一分词与第二分词在同一视频资源数据中共同出现的概率；

具体而言，同现率可以为当前一个或多个分词与第二分词在同一视频资源数据的特征本本文信息中共同出现的概率，具体可以包括一个第一分词和第二分词的同现率，多个分词和第二分词的同现率。

35 需要说明的是，第二分词可以是在全部预设的分词中，除第一分词以外的分词。关联第二分词可以是与第一分词的同现率高于预设阈值的第二分词。

在实际应用中，视频资源数据可以包括特征本本文信息，该特征本本文信息可以用于记载该视频资源数据的相关信息，也可以用于提取分词。

在本发明的一种优选实施例中，所述步骤 303 具体可以包括如下子步骤：

40 子步骤 S131，当所述特征本本文信息被映射为一个第一分词时，提取所述第一分词对应的预置索引表；其中，所述索引表包括所述第一分词所属的视频资源数据的信息，以及，所述视频资源数据中的所有分词；所述视频资源数据中的所有分词为通过抓取视频资源数据，提取所述视频资源数据的特征本本文信息，对所述特征本本文信息进行分词生成；

45 在本发明的一种优选实施例中，所述特征本本文信息可以包括视频标题、视频关键词和/或视频描述。

子步骤 S132，计算所述第一分词与所述索引表中各个第二分词的同现率，所述同现率为所述索引表中各个第二分词出现的次数与所述索引表中视频资源数据的信息总数的比值；其中，所述第二分词为所述视频资源数据中的所有分词中除所述第一分词以外的分词；

子步骤 S133，提取所述同现率高于预设阈值的第二分词作为关联第二分词。

50 在本发明的一种优选实施例中，所述步骤 303 具体可以包括如下子步骤：

子步骤 S141, 当所述特征文本信息被映射为多个第一分词时, 分别提取所述多个第一分词对应的多个预置索引表; 各个索引表中包括所述第一分词所属的视频资源数据的信息, 以及, 所述视频资源数据中的所有分词; 所述视频资源数据中的所有分词为通过抓取视频资源数据, 提取所述视频资源数据的特征文本信息, 对所述特征文本信息进行分词生成;

5 在本发明的一种优选实施例中, 所述特征文本信息可以包括视频标题、视频关键词和/或视频描述。

子步骤 S142, 提取与所述多个第一分词共同出现的第二分词作为候选分词; 其中, 所述第二分词为所述视频资源数据中的所有分词中除所述第一分词以外的分词;

子步骤 S143, 分别在各个索引表中计算所述第一分词与所述候选分词的同现率, 所述同现率为所述索引表中候选分词出现的次数与所述索引表中视频资源数据的信息总数的比值;

子步骤 S144, 分别为所述多个第一分词与所述候选分词的同现率配置对应的多个权重;

子步骤 S145, 分别计算多个配置了权重的同现率的平均值, 作为所述多个第一分词与所述候选分词的同现率;

子步骤 S146, 提取所述同现率高于预设阈值的候选分词作为关联第二分词。

15 在本发明的一种优选实施例中, 所述步骤 303 具体可以包括如下子步骤:

子步骤 S151, 当所述特征文本信息被映射为多个第一分词时, 分别提取所述多个第一分词对应的多个预置索引表; 其中, 各个索引表中包括所述第一分词所属的视频资源数据的信息, 以及, 所述视频资源数据中的所有分词; 所述视频资源数据中的所有分词为通过抓取视频资源数据, 提取所述视频资源数据的特征文本信息, 对所述特征文本信息进行分词生成;

20 在本发明的一种优选实施例中, 所述特征文本信息可以包括视频标题、视频关键词和/或视频描述。

子步骤 S152, 采用所述多个索引表确定主分词, 所述主分词为视频资源数据的信息总数最多的索引表对应的第一分词;

子步骤 S153, 计算所述主分词与其对应的索引表中各个第二分词的同现率, 所述同现率为所述索引表中各个第二分词出现的次数与所述索引表中视频资源数据的信息总数的比值; 其中, 所述第二分词为所述视频资源数据中的所有分词中除所述第一分词以外的分词;

本发明实施例中, 可以以主分词的同现率的作为最终的同现率。

子步骤 S154, 提取所述同现率高于预设阈值的第二分词作为关联第二分词。

30 步骤 304, 获取与所述一个或多个第一分词和所述关联第二分词匹配的第二视频资源数据的网络链接地址;

具体而言, 在子步骤 S133 之后, 可以获得当前一个第一分词与一个或多个关联第二分词的组合。例如特征文本信息为“dota”, 与其同现率较高的词为: “搞笑”、“蛋疼”、“2009”、“海涛”、“第一视角”和“经典”, 同现率分别为 40%、35%、30%、25%、20%和 10%, 则获得的组合依次为“dota 搞笑”、“dota 蛋疼”、“dota2009”、“dota 海涛”、“dota 第一视角”和“dota 经典”。

在子步骤 S146 之后, 可以获得当前多个第一分词与一个或多个关联第二分词的组合。例如特征文本信息为“广场舞兵哥哥”, 将其映射为第一分词“广场舞”和“兵哥哥”, 提取与这两个第一分词同时出现的第二分词, 例如第二分词“教学”, 其可以作为关联第二分词, 则最终获得组合“广场舞兵哥哥教学”。

40 在本发明的一种优选实施例中, 步骤 304 具体可以包括如下子步骤:

子步骤 S161, 获取所述主分词和所述关联第二分词的第二视频资源数据的网络链接地址。

在子步骤 S154 之后, 可以获得当前主分词与一个或多个关联第二分词的组合。例如, 对于特征文本信息“中秋节月饼”所映射的第一分词“中秋”和“月饼”, 可以设置“中秋”作为主分词, 得到关联第二分词“月亮”, 则最终获得组合“中秋月亮”。

45 本发明实施例中, 可以基于一个或多个第一分词和第二分词的组合进行匹配的视频数据资源的搜索, 当搜索到时, 记录其网络连接地址, 具体可以是内网地址, 也可以是外网地址。

步骤 305, 推送所述第二视频资源数据的网络链接地址。

实际应用中, 第二视频资源数据的网络链接地址可以放置在当前页面的任一位置, 也可以通过嵌入图标或按钮等方式进行推送, 用户可以通过触发第二视频资源数据的网络链接地址进而加载所述视频数据资源。

50

本发明可以根据现有已发布内容进行推送,使搜索引擎摆脱对用户搜索习惯的依赖,将虽然比较少有用户搜索的但视频库汇总已有较多相关资源的视频资源数据推送出来,从而实现深度挖掘视频库中的优质资源,提高了资源挖掘的效率;此外,索引表会随着互联网视频内容的不断积累而不断扩大,各大视频站生产出来的内容数量和广度会远远超过用户已经搜索过的词数,有利于扩大召回率。

本发明通过获取第一分词和第二分词的匹配的第二视频资源数据的网络连接地址,用户可以基于此地址直接进行视频数据资源的获取,使用户简单搜索即可获得更多的结果,无需多次提交搜索,从而减轻了访问服务器的负担,减少了网络资源的占用,并提升了用户体验。

对于方法实施例,为了简单描述,故将其都表述为一系列的动作组合,但是本领域技术人员应该知悉,本发明实施例并不受所描述的动作顺序的限制,因为依据本发明实施例,某些步骤可以采用其他顺序或者同时进行。其次,本领域技术人员也应该知悉,说明书中所描述的实施例均属于优选实施例,所涉及的动作并不一定是本发明实施例所必须的。

参照图 4,示出了根据本发明一个实施例的一种基于视频搜索的分词信息推送装置实施例的结构框图,具体可以包括如下模块:

视频搜索字符串接收模块 401,适于接收视频搜索字符串;

第一分词映射模块 402,适于将所述视频搜索字符串映射为一个或多个第一分词;

第二分词查找模块 403,适于查找与所述一个或多个第一分词的同现率高于预设阈值的关联第二分词;所述同现率为当前一个或多个分词与第二分词在同一视频资源数据中共同出现的概率;

推送模块 404,适于推送所述一个或多个第一分词与所述一个或多个关联第二分词的组合。

在本发明的一种优选实施例中,所述第一分词映射模块 402 还可以适于:

提取所述视频搜索字符串所映射的一个分词;

或者,

当接收到的视频搜索字符串为复合词时,将所述视频搜索字符串拆分为多个搜索子词;提取所述多个搜索子词所映射的多个分词。

在本发明的一种优选实施例中,所述第二分词查找模块 403 还可以适于:

当所述视频搜索字符串被映射为一个第一分词时,提取所述第一分词对应的预置索引表;其中,所述索引表包括所述第一分词所属的视频资源数据的信息,以及,所述视频资源数据中的所有分词;所述视频资源数据中的所有分词为通过抓取视频资源数据,提取所述视频资源数据的特征文本信息,对所述特征文本信息进行分词生成;

计算所述第一分词与所述索引表中各个第二分词的同现率,所述同现率为所述索引表中各个第二分词出现的次数与所述索引表中视频资源数据的信息总数的比值;其中,所述第二分词为所述视频资源数据中的所有分词中除所述第一分词以外的分词;

提取所述同现率高于预设阈值的第二分词作为关联第二分词。

在本发明的一种优选实施例中,所述第二分词查找模块 403 还可以适于:

当所述视频搜索字符串被映射为多个第一分词时,分别提取所述多个第一分词对应的多个预置索引表;各个索引表中包括所述第一分词所属的视频资源数据的信息,以及,所述视频资源数据中的所有分词;所述视频资源数据中的所有分词为通过抓取视频资源数据,提取所述视频资源数据的特征文本信息,对所述特征文本信息进行分词生成;

提取与所述多个第一分词共同出现的第二分词作为候选分词;其中,所述第二分词为所述视频资源数据中的所有分词中除所述第一分词以外的分词;

分别在各个索引表中计算所述第一分词与所述候选分词的同现率,所述同现率为所述索引表中候选分词出现的次数与所述索引表中视频资源数据的信息总数的比值;

分别为所述多个第一分词与所述候选分词的同现率配置对应的多个权重;

分别计算多个配置了权重的同现率的平均值，作为所述多个第一分词与所述候选分词的同现率；

提取所述同现率高于预设阈值的候选分词作为关联第二分词。

在本发明的一种优选实施例中，所述第二分词查找模块 403 还可以适于：

5 当所述视频搜索字符串被映射为多个第一分词时，分别提取所述多个第一分词对应的多个预置索引表；其中，各个索引表中包括所述第一分词所属的视频资源数据的信息，以及，所述视频资源数据中的所有分词；所述视频资源数据中的所有分词为通过抓取视频资源数据，提取所述视频资源数据的特征文本信息，对所述特征文本信息进行分词生成；

10 采用所述多个索引表确定主分词，所述主分词为视频资源数据的信息总数最多的索引表对应的第一分词；

计算所述主分词与其对应的索引表中各个第二分词的同现率，所述同现率为所述索引表中各个第二分词出现的次数与所述索引表中视频资源数据的信息总数的比值；其中，所述第二分词为所述视频资源数据中的所有分词中除所述第一分词以外的分词；

提取所述同现率高于预设阈值的第二分词作为关联第二分词。

15 在本发明的一种优选实施例中，所述特征文本信息包括视频标题、视频关键词和/或视频描述。

在本发明的一种优选实施例中，所述组合推送模块 404 还可以适于：

推送所述主分词和所述关联第二分词的组合。

20 本发明的各个部件实施例可以以硬件实现，或者以在一个或者多个处理器上运行的软件模块实现，或者以它们的组合实现。本领域的技术人员应当理解，可以在实践中使用微处理器或者数字信号处理器 (DSP) 来实现根据本发明实施例的基于视频搜索的分词信息推送设备中的一些或者全部部件的一些或者全部功能。本发明还可以实现为用于执行这里所描述的方法的一部分或者全部的设备或者装置程序 (例如，计算机程序和计算机程序产品)。这样的实现本发明的程序可以存储在计算机可读介质上，或者可以具有一个或者多个信号的形式。这样的信号可以从因特网网站上下载得到，或者在载体信号上提供，或者以任何其他形式提供。

25 例如，图 5 示出了可以实现根据本发明的基于视频搜索的分词信息推送的计算设备，例如用户终端设备或应用服务器。该计算设备传统上包括处理器 510 和以存储器 520 形式的计算机程序产品或者计算机可读介质。存储器 520 可以是诸如闪存、EEPROM (电可擦除可编程只读存储器)、EPROM、硬盘或者 ROM 之类的电子存储器。存储器 520 具有用于执行上述方法中的任何方法步骤的程序代码 531 的存储空间 530。例如，用于程序代码的存储空间 530 可以包括分别用于实现上面的方法中的各种步骤的各个程序代码 531。这些程序代码可以从一个或者多个计算机程序产品中读出或者写入到这一个或者多个计算机程序产品中。这些计算机程序产品包括诸如硬盘，紧致盘 (CD)、存储卡或者软盘之类的程序代码载体。这样的计算机程序产品通常为如参考图 6 所述的便携式或者固定存储单元。该存储单元可以具有与图 5 的计算设备中的存储器 520 类似布置的存储段、存储空间等。程序代码可以例如以适当形式进行压缩。通常，存储单元包括计算机可读代码 531'，即可以由例如诸如 510 之类的处理器读取的代码，这些代码当由计算设备运行时，导致该计算设备执行上面所描述的方法中的各个步骤。

30 本文中所称的“一个实施例”、“实施例”或者“一个或者多个实施例”意味着，结合实施例描述的特定特征、结构或者特性包括在本发明的至少一个实施例中。此外，请注意，这里“在一个实施例中”的词语例子不一定全指同一个实施例。

40 在此处所提供的说明书中，说明了大量具体细节。然而，能够理解，本发明的实施例可以在没有这些具体细节的情况下被实践。在一些实例中，并未详细示出公知的方法、结构和技术，以便不模糊对本说明书的理解。

45 应该注意的是上述实施例对本发明进行说明而不是对本发明进行限制，并且本领域技术人员在不脱离所附权利要求的范围的情况下可设计出替换实施例。在权利要求中，不应将位于括号之间的任何参考符号构造成对权利要求的限制。单词“包含”不排除存在未列在权利要求中的元件或步骤。位于元件之前的单词“一”或“一个”不排除存在多个这样的元件。本发明可以借助于包括有若干不同元件的硬件以及借助于适当编程的计算机来实现。在列举了若干装置的单元权利要求中，这些装置中的若干个可以是同一个硬件项来具体体现。单词第一、第二、以及第三等的使用不表示任何顺序。可将这些单词解释为名称。

50

此外，还应当注意，本说明书中使用的语言主要是为了可读性和教导的目的而选择的，而不是为了解释或者限定本发明的主题而选择的。因此，在不偏离所附权利要求书的范围和精神的情况下，对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。对于本发明的范围，对本发明所做的公开是说明性的，而非限制性的，本发明的范围由所附权利要求书限定。

5

## 权 利 要 求 书

- 1、一种基于视频搜索的分词信息推送方法，包括：  
接收视频搜索字符串；  
将所述视频搜索字符串映射为一个或多个第一分词；  
查找与所述一个或多个第一分词的同现率高于预设阈值的关联第二分词；所述同现率为  
5 当前一个或多个第一分词与第二分词在同一视频资源数据中共同出现的概率；  
推送所述一个或多个第一分词与所述一个或多个关联第二分词的组合。
- 2、如权利要求 1 所述的方法，其特征在于，所述将所述视频搜索字符串映射为一个或多个  
10 个第一分词的步骤包括：  
提取所述视频搜索字符串所映射的一个分词；  
或者，  
当接收到的视频搜索字符串为复合词时，将所述视频搜索字符串拆分为多个搜索子词；  
提取所述多个搜索子词所映射的多个分词。
- 3、如权利要求 1 所述的方法，其特征在于，所述查找与所述一个或多个第一分词的同现  
15 率高于预设阈值的关联第二分词的步骤包括：  
当所述视频搜索字符串被映射为一个第一分词时，提取所述第一分词对应的预置索引  
表；其中，所述索引表包括所述第一分词所属的视频资源数据的信息，以及，所述视频资源数  
据中的所有分词；所述视频资源数据中的所有分词为通过抓取视频资源数据，提取所述视频资  
源数据的特征文本信息，对所述特征文本信息进行分词生成；  
20 计算所述第一分词与所述索引表中各个第二分词的同现率，所述同现率为所述索引表中  
各个第二分词出现的次数与所述索引表中视频资源数据的信息总数的比值；其中，所述第二  
分词为所述视频资源数据中的所有分词中除所述第一分词以外的分词；  
提取所述同现率高于预设阈值的第二分词作为关联第二分词。
- 4、如权利要求 1 所述的方法，其特征在于，所述查找与所述一个或多个第一分词的同现  
25 率高于预设阈值的关联第二分词的步骤包括：  
当所述视频搜索字符串被映射为多个第一分词时，分别提取所述多个第一分词对应的多  
个预置索引表；各个索引表中包括所述第一分词所属的视频资源数据的信息，以及，所述视频  
资源数据中的所有分词；所述视频资源数据中的所有分词为通过抓取视频资源数据，提取所述  
视频资源数据的特征文本信息，对所述特征文本信息进行分词生成；  
30 提取与所述多个第一分词共同出现的第二分词作为候选分词；其中，所述第二分词为所  
述视频资源数据中的所有分词中除所述第一分词以外的分词；  
分别在各个索引表中计算所述第一分词与所述候选分词的同现率，所述同现率为所述索  
引表中候选分词出现的次数与所述索引表中视频资源数据的信息总数的比值；  
分别为所述多个第一分词与所述候选分词的同现率配置对应的多个权重；  
35 分别计算多个配置了权重的同现率的平均值，作为所述多个第一分词与所述候选分词的  
同现率；  
提取所述同现率高于预设阈值的候选分词作为关联第二分词。
- 5、如权利要求 1 所述的方法，其特征在于，所述查找与所述一个或多个第一分词的同现  
率高于预设阈值的关联第二分词的步骤包括：  
40 当所述视频搜索字符串被映射为多个第一分词时，分别提取所述多个第一分词对应的多  
个预置索引表；其中，各个索引表中包括所述第一分词所属的视频资源数据的信息，以及，所  
述视频资源数据中的所有分词；所述视频资源数据中的所有分词为通过抓取视频资源数据，提  
取所述视频资源数据的特征文本信息，对所述特征文本信息进行分词生成；  
采用所述多个索引表确定主分词，所述主分词为视频资源数据的信息总数最多的索引表  
45 对应的第一分词；  
计算所述主分词与其对应的索引表中各个第二分词的同现率，所述同现率为所述索引表  
中各个第二分词出现的次数与所述索引表中视频资源数据的信息总数的比值；其中，所述第  
二分词为所述视频资源数据中的所有分词中除所述第一分词以外的分词；  
提取所述同现率高于预设阈值的第二分词作为关联第二分词。
- 6、如权利要求 3 或 4 或 5 所述的方法，其特征在于，所述特征文本信息包括视频标题、



视频关键词和/或视频描述。

7、如权利要求 5 所述的方法，其特征在于，所述推送所述一个或多个第一分词与所述一个或多个关联第二分词的组合的步骤包括：

推送所述主分词和所述关联第二分词的组合。

5

8、一种基于视频搜索的在线播放入口对象的推送方法，包括：

接收视频搜索字符串；

将所述视频搜索字符串映射为一个或多个第一分词；

10 查找与所述一个或多个第一分词的同现率高于预设阈值的关联第二分词；所述同现率为当前一个或多个第一分词与第二分词在同一视频资源数据中共同出现的概率；

获取与所述一个或多个第一分词和所述关联第二分词匹配的一个或多个视频数据资源的网络地址；

根据所述一个或多个视频数据资源网络地址构造在线播放所述视频数据资源的入口对象；

推送所述一个或多个在线播放视频数据资源的入口对象。

15

9、一种基于视频搜索的关联资源地址的推送方法，包括：

当接收到第一视频资源数据的加载或播放请求时，获取所述第一视频资源数据的特征文本信息；

将所述特征文本信息映射为一个或多个第一分词；

20 查找与所述一个或多个第一分词的同现率高于预设阈值的关联第二分词；所述同现率为当前一个或多个第一分词与第二分词在同一视频资源数据中共同出现的概率；

获取与所述一个或多个第一分词和所述关联第二分词匹配的第二视频资源数据的网络链接地址；

推送所述第二视频资源数据的网络链接地址。

25

10、一种基于视频搜索的分词信息推送装置，包括：

视频搜索字符串接收模块，适于接收视频搜索字符串；

第一分词映射模块，适于将所述视频搜索字符串映射为一个或多个第一分词；

30 第二分词查找模块，适于查找与所述一个或多个第一分词的同现率高于预设阈值的关联第二分词；所述同现率为当前一个或多个分词与第二分词在同一视频资源数据中共同出现的概率；

组合推送模块，适于推送所述一个或多个第一分词与所述一个或多个关联第二分词的组合。

35

11、如权利要求 10 所述的装置，其特征在于，所述第一分词映射模块还适于：

提取所述视频搜索字符串所映射的一个分词；

或者，

当接收到的视频搜索字符串为复合词时，将所述视频搜索字符串拆分为多个搜索子词；提取所述多个搜索子词所映射的多个分词。

40

12、如权利要求 10 所述的装置，其特征在于，所述第二分词查找模块还适于：

当所述视频搜索字符串被映射为一个第一分词时，提取所述第一分词对应的预置索引表；其中，所述索引表包括所述第一分词所属的视频资源数据的信息，以及，所述视频资源数据中的所有分词；所述视频资源数据中的所有分词为通过抓取视频资源数据，提取所述视频资源数据的特征文本信息，对所述特征文本信息进行分词生成；

45 计算所述第一分词与所述索引表中各个第二分词的同现率，所述同现率为所述索引表中各个第二分词出现的次数与所述索引表中视频资源数据的信息总数的比值；其中，所述第二分词为所述视频资源数据中的所有分词中除所述第一分词以外的分词；

提取所述同现率高于预设阈值的第二分词作为关联第二分词。

50

13、如权利要求 10 所述的装置，其特征在于，所述第二分词查找模块还适于：

当所述视频搜索字符串被映射为多个第一分词时，分别提取所述多个第一分词对应的多个预置索引表；各个索引表中包括所述第一分词所属的视频资源数据的信息，以及，所述视频资

源数据中的所有分词；所述视频资源数据中的所有分词为通过抓取视频资源数据，提取所述视频资源数据的特征文本信息，对所述特征文本信息进行分词生成；

提取与所述多个第一分词共同出现的第二分词作为候选分词；其中，所述第二分词为所述视频资源数据中的所有分词中除所述第一分词以外的分词；

5 分别在各个索引表中计算所述第一分词与所述候选分词的同现率，所述同现率为所述索引表中候选分词出现的次数与所述索引表中视频资源数据的信息总数的比值；

分别为所述多个第一分词与所述候选分词的同现率配置对应的多个权重；

分别计算多个配置了权重的同现率的平均值，作为所述多个第一分词与所述候选分词的同现率；

10 提取所述同现率高于预设阈值的候选分词作为关联第二分词。

14、如权利要求 10 所述的装置，其特征在于，所述第二分词查找模块还适于：

当所述视频搜索字符串被映射为多个第一分词时，分别提取所述多个第一分词对应的多个预置索引表；其中，各个索引表中包括所述第一分词所属的视频资源数据的信息，以及，所述视频资源数据中的所有分词；所述视频资源数据中的所有分词为通过抓取视频资源数据，提取  
15 所述视频资源数据的特征文本信息，对所述特征文本信息进行分词生成；

采用所述多个索引表确定主分词，所述主分词为视频资源数据的信息总数最多的索引表对应的第一分词；

计算所述主分词与其对应的索引表中各个第二分词的同现率，所述同现率为所述索引表中各个第二分词出现的次数与所述索引表中视频资源数据的信息总数的比值；其中，所述第二分  
20 词为所述视频资源数据中的所有分词中除所述第一分词以外的分词；

提取所述同现率高于预设阈值的第二分词作为关联第二分词。

15、如权利要求 12 或 13 或 14 所述的装置，所述特征文本信息包括视频标题、视频关键词和/或视频描述。

16、如权利要求 14 所述的装置，其特征在于，所述组合推送模块还适于：

25 推送所述主分词和所述关联第二分词的组合。

17、一种计算机程序，包括计算机可读代码，当所述计算机可读代码在计算设备上运行时，导致所述计算设备执行根据权利要求 1-7 中的任一个所述的基于视频搜索的分词信息推送方法。

18、一种计算机可读介质，其中存储了如权利要求 17 所述的计算机程序。

# 说明书附图

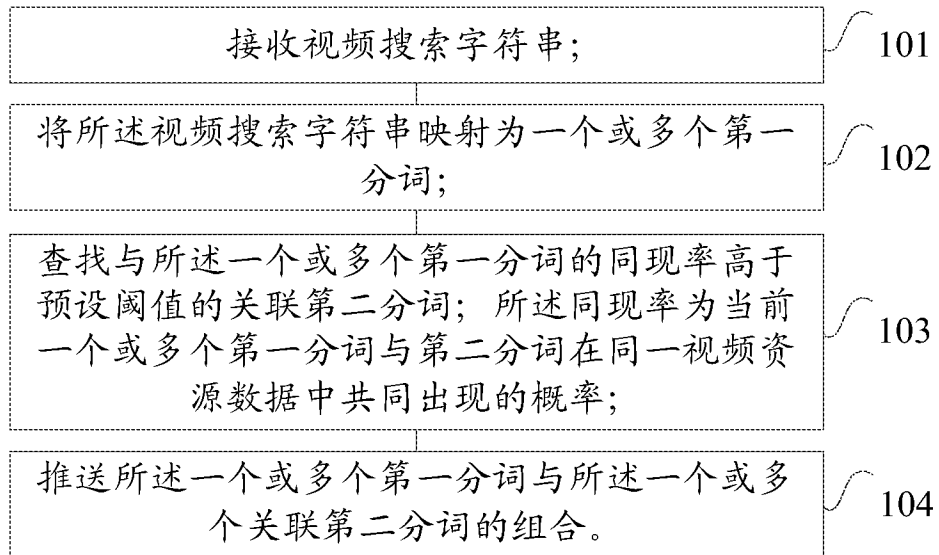


图 1

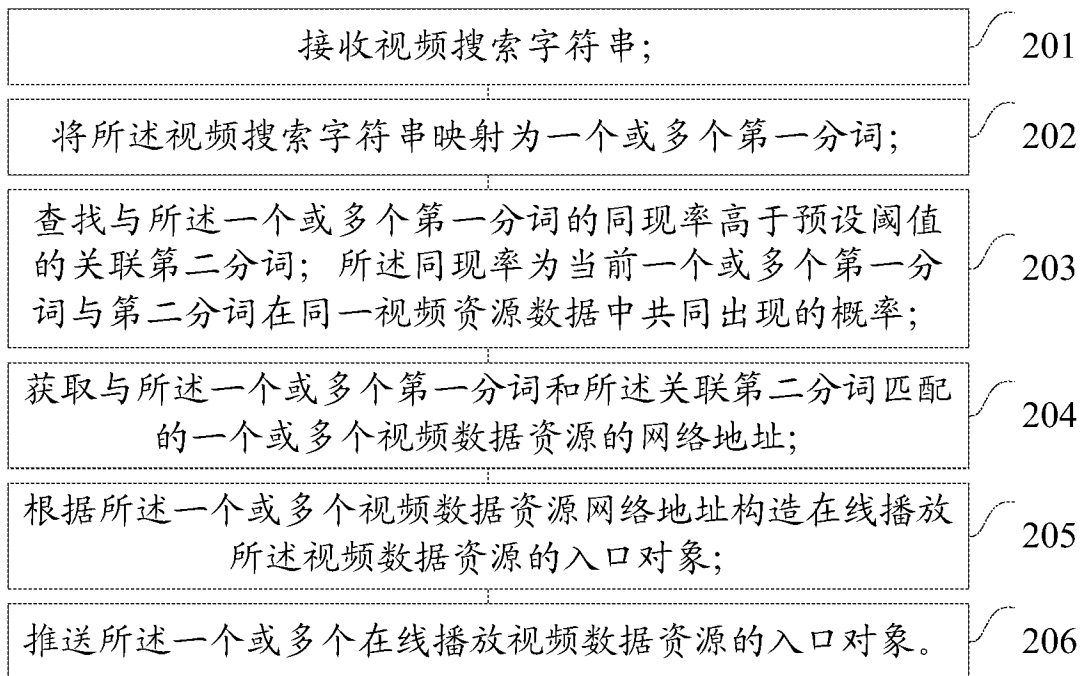


图 2

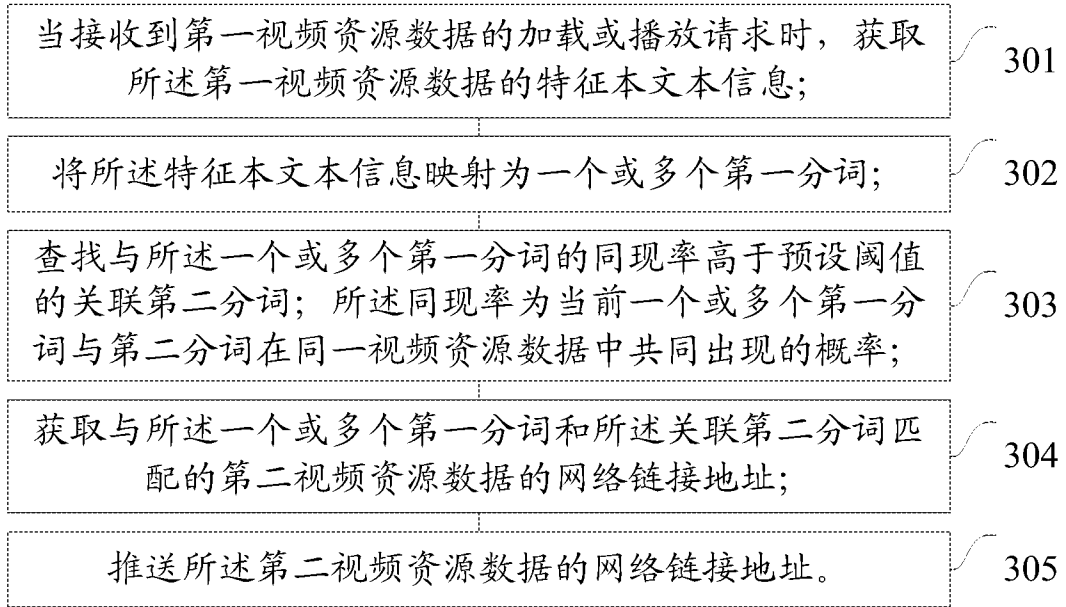


图 3

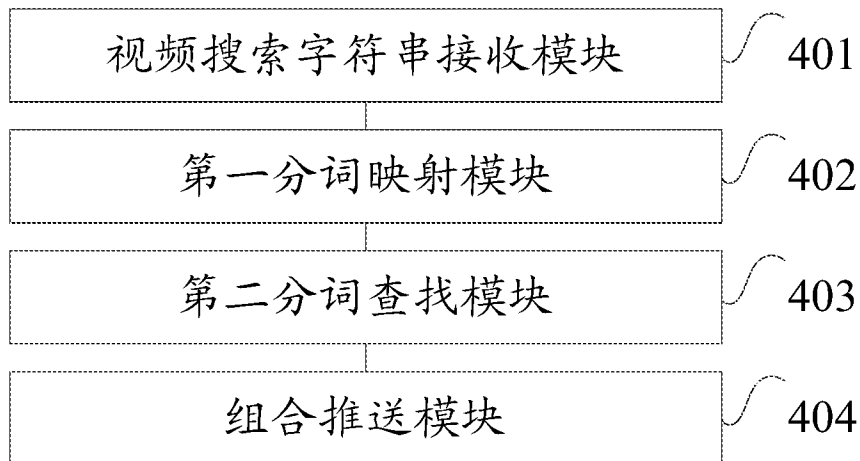


图 4

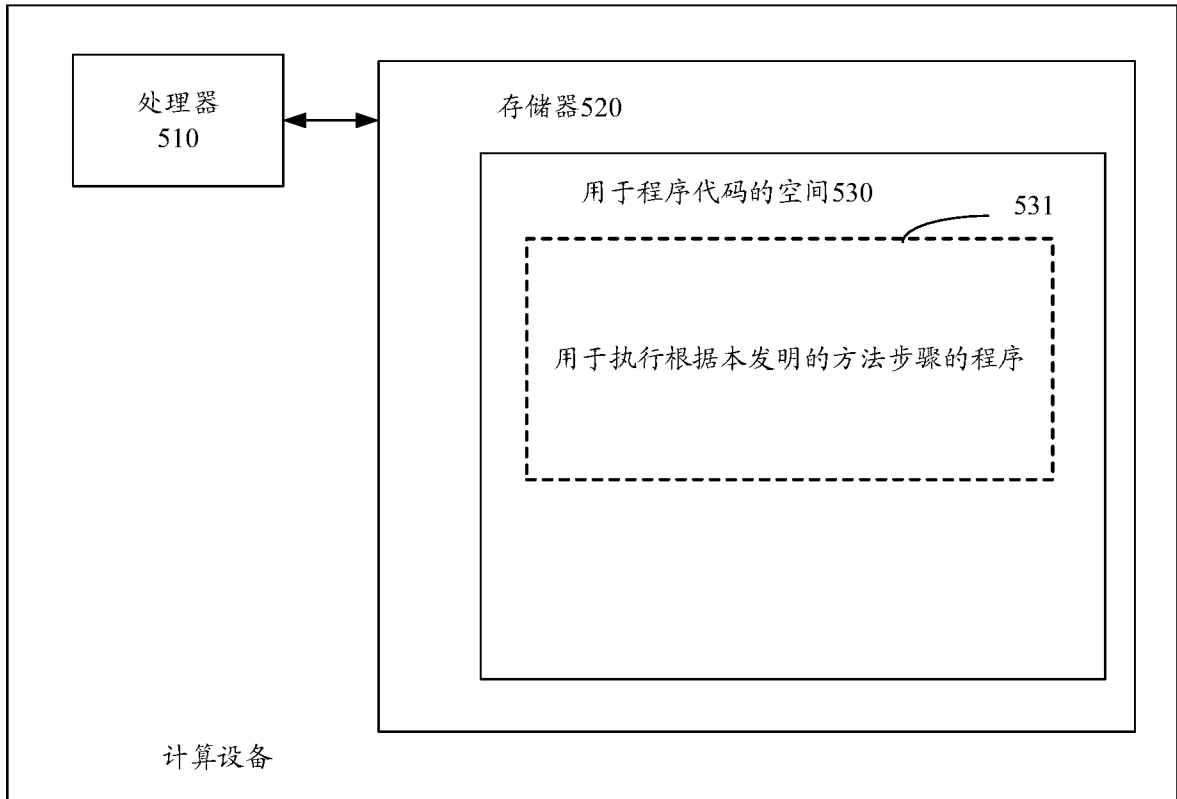


图 5

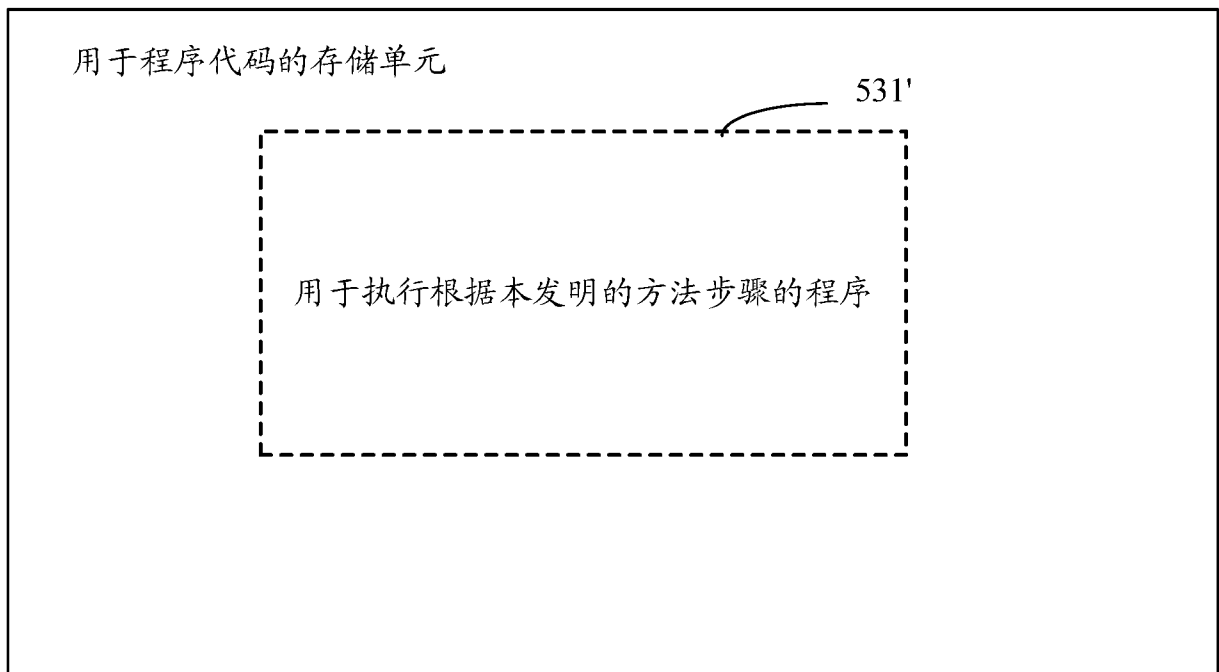


图 6

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2014/086519

## A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30 (2006.01) i; H04L 29/12 (2006.01) i; H04L 29/08 (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)

H04L 29; G06F 17

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: video, movie, search, keyword, divide, word, push, recommend, probability, common, address, play, load,  
weight, average, appear

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 101957828 A (ALIBABA GROUP HOLDINGS LTD.) 26 January 2011 (26.01.2011) description, paragraphs [0105]-[0157], and figures 1 and 2	1, 2, 10, 11, 17, 18
Y	CN 101957828 A (ALIBABA GROUP HOLDINGS LTD.) 26 January 2011 (26.01.2011) description, paragraphs [0105]-[0157], and figures 1 and 2	8, 9
A	CN 101957828 A (ALIBABA GROUP HOLDINGS LTD.) 26 January 2011 (26.01.2011) description, paragraphs [0105]-[0157], and figures 1 and 2	3-7, 12-16
Y	CN 101236567 A (SHANGHAI SHENGYUE ELECTRONIC TECHNOLOGY CO., LTD.) 06 August 2008 (06.08.2008) description, page 9	8
Y	CN 103164405 A (SHENGYUE INFORMATION TECHNOLOGY (SHANGHAI) CO., LTD.) 19 June 2013 (19.06.2013) description, paragraphs [0093]-[0111], and figures 3 and 4	9
X	CN 102326144 A (ATIGEO LLC) 18 January 2012 (18.01.2012) claims 1-3, 12, 22	1, 2, 10, 11, 17, 18

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>“&amp;” document member of the same patent family</p>
---	---

Date of the actual completion of the international search  
22 November 2014

Date of mailing of the international search report  
22 December 2014

Name and mailing address of the ISA  
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  
SHEN, Minjie  
Telephone No. (86-10) 62413417

**INTERNATIONAL SEARCH REPORT**International application No.  
PCT/CN2014/086519

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 102326144 A (ATIGEO LLC.) 18 January 2012 (18.01.2012) claims 1-3, 12, 22	8, 9
PX	CN 103488787 A (BEIJING QIHOO TECHNOLOGY CO., LTD. et al.) 01 January 2014 (01.01.2014) the whole document	1-18
PX	CN 103500214 A (BEIJING QIHOO TECHNOLOGY CO., LTD. et al.) 08 January 2014 (08.01.2014) the whole document	1-18
PX	CN 103491205 A (BEIJING QIHOO TECHNOLOGY CO., LTD. et al.) 01 January 2014 (01.01.2014) the whole document	1-18
A	US 2004064447 A1 (SIMSKE, STEVEN J. et al.) 01 April 2004 (01.04.2004) the whole document	1-18

**INTERNATIONAL SEARCH REPORT**  
Information on patent family members

International application No.  
PCT/CN2014/086519

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 101957828 A	26 January 2011	WO 2011011046 A1	27 January 2011
		EP 2457151 A1	30 May 2012
		HK 1149823 A1	23 August 2013
		US 2011016111 A1	20 January 2011
		JP 2012533818 A	27 December 2012
CN 101236567 A	06 August 2008	None	
CN 103164405 A	19 June 2013	None	
CN 102326144 A	18 January 2012	US 2009216696 A1	27 August 2009
		CA 2805391 A	03 September 2009
		WO 2009108726 A1	03 September 2009
		CA 2716062 A	03 September 2009
		WO 2010068931 A1	17 June 2010
		US 2010153324 A1	17 June 2010
		CA 2746675 A	17 June 2010
		EP 2260373 A1	15 December 2010
		CN 102016787 A	13 April 2011
		JP 2011514995 A	12 May 2011
		EP 2377011 A1	19 October 2011
		MX 2011006340 A	28 October 2011
		JP 2012512465 A	31 May 2012
US 2004064447 A1	01 April 2004	US 2013066887 A1	14 March 2013
		JP 2013218727 A	24 October 2013
		US 2013325769 A1	05 December 2013
		GB 2393541 A	31 March 2004
		DE 10328833 A1	15 April 2004
CN 103488787 A	01 January 2014	None	
CN 103500214 A	08 January 2014	None	
CN 103491205 A	01 January 2014	None	
US 2004064447 A1	01 April 2004	GB 2417115 A	15 February 2006



<p>A. 主题的分类</p> <p>G06F 17/30(2006.01)i; H04L 29/12(2006.01)i; H04L 29/08(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>H04L29; G06F17</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>WPI, EPODOC, CNPAT, 视频, 电影, 搜索, 分词, 推送, 推荐, 概率, 出现, 共同, 关键词, 地址, 播放, 加载, 权重, 平均, video, movie, search, keyword, divide, word, push, recommend, probability, common, address, play, load, weight, average</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>X</td> <td>CN 101957828 A (阿里巴巴集团控股有限公司) 2011年 1月 26日 (2011 - 01 - 26) 说明书第105至157段及图1至2</td> <td>1-2, 10-11, 17-18</td> </tr> <tr> <td>Y</td> <td>CN 101957828 A (阿里巴巴集团控股有限公司) 2011年 1月 26日 (2011 - 01 - 26) 说明书第105至157段及图1至2</td> <td>8-9</td> </tr> <tr> <td>A</td> <td>CN 101957828 A (阿里巴巴集团控股有限公司) 2011年 1月 26日 (2011 - 01 - 26) 说明书第105至157段及图1至2</td> <td>3-7, 12-16</td> </tr> <tr> <td>Y</td> <td>CN 101236567 A (上海升岳电子科技有限公司) 2008年 8月 06日 (2008 - 08 - 06) 说明书第9页</td> <td>8</td> </tr> <tr> <td>Y</td> <td>CN 103164405 A (盛乐信息技术上海有限公司) 2013年 6月 19日 (2013 - 06 - 19) 说明书第93至111段及图3至4</td> <td>9</td> </tr> <tr> <td>X</td> <td>CN 102326144 A (阿迪吉欧有限责任公司) 2012年 1月 18日 (2012 - 01 - 18) 权利要求1-3、12、22</td> <td>1-2, 10-11, 17-18</td> </tr> <tr> <td>Y</td> <td>CN 102326144 A (阿迪吉欧有限责任公司) 2012年 1月 18日 (2012 - 01 - 18) 权利要求1-3、12、22</td> <td>8-9</td> </tr> </tbody> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 101957828 A (阿里巴巴集团控股有限公司) 2011年 1月 26日 (2011 - 01 - 26) 说明书第105至157段及图1至2	1-2, 10-11, 17-18	Y	CN 101957828 A (阿里巴巴集团控股有限公司) 2011年 1月 26日 (2011 - 01 - 26) 说明书第105至157段及图1至2	8-9	A	CN 101957828 A (阿里巴巴集团控股有限公司) 2011年 1月 26日 (2011 - 01 - 26) 说明书第105至157段及图1至2	3-7, 12-16	Y	CN 101236567 A (上海升岳电子科技有限公司) 2008年 8月 06日 (2008 - 08 - 06) 说明书第9页	8	Y	CN 103164405 A (盛乐信息技术上海有限公司) 2013年 6月 19日 (2013 - 06 - 19) 说明书第93至111段及图3至4	9	X	CN 102326144 A (阿迪吉欧有限责任公司) 2012年 1月 18日 (2012 - 01 - 18) 权利要求1-3、12、22	1-2, 10-11, 17-18	Y	CN 102326144 A (阿迪吉欧有限责任公司) 2012年 1月 18日 (2012 - 01 - 18) 权利要求1-3、12、22	8-9
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																								
X	CN 101957828 A (阿里巴巴集团控股有限公司) 2011年 1月 26日 (2011 - 01 - 26) 说明书第105至157段及图1至2	1-2, 10-11, 17-18																								
Y	CN 101957828 A (阿里巴巴集团控股有限公司) 2011年 1月 26日 (2011 - 01 - 26) 说明书第105至157段及图1至2	8-9																								
A	CN 101957828 A (阿里巴巴集团控股有限公司) 2011年 1月 26日 (2011 - 01 - 26) 说明书第105至157段及图1至2	3-7, 12-16																								
Y	CN 101236567 A (上海升岳电子科技有限公司) 2008年 8月 06日 (2008 - 08 - 06) 说明书第9页	8																								
Y	CN 103164405 A (盛乐信息技术上海有限公司) 2013年 6月 19日 (2013 - 06 - 19) 说明书第93至111段及图3至4	9																								
X	CN 102326144 A (阿迪吉欧有限责任公司) 2012年 1月 18日 (2012 - 01 - 18) 权利要求1-3、12、22	1-2, 10-11, 17-18																								
Y	CN 102326144 A (阿迪吉欧有限责任公司) 2012年 1月 18日 (2012 - 01 - 18) 权利要求1-3、12、22	8-9																								
<p><input checked="" type="checkbox"/> 其余文件在C栏的续页中列出。</p> <p><input checked="" type="checkbox"/> 见同族专利附件。</p>																										
<p>* 引用文件的具体类型:</p> <p>“A” 认为不特别相关的表示了现有技术一般状态的文件</p> <p>“E” 在国际申请日的当天或之后公布的在先申请或专利</p> <p>“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)</p> <p>“O” 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>“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> <p>“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</p> <p>“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</p> <p>“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</p> <p>“&amp;” 同族专利的文件</p>																										
<p>国际检索实际完成的日期</p> <p>2014年 11月 22日</p>		<p>国际检索报告邮寄日期</p> <p>2014年 12月 22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>沈敏洁</p> <p>电话号码 (86-10)62413417</p>																								

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
PX	CN 103488787 A (北京奇虎科技有限公司等) 2014年 1月 01日 (2014 - 01 - 01) 全文	1-18
PX	CN 103500214 A (北京奇虎科技有限公司等) 2014年 1月 08日 (2014 - 01 - 08) 全文	1-18
PX	CN 103491205 A (北京奇虎科技有限公司等) 2014年 1月 01日 (2014 - 01 - 01) 全文	1-18
A	US 2004064447 A1 (SIMSKE, STEVEN J. 等) 2004年 4月 01日 (2004 - 04 - 01) 全文	1-18

国际检索报告  
关于同族专利的信息

国际申请号

PCT/CN2014/086519

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	101957828	A	2011年 1月 26日	WO	2011011046	A1	2011年 1月 27日
				EP	2457151	A1	2012年 5月 30日
				HK	1149823	A1	2013年 8月 23日
				US	2011016111	A1	2011年 1月 20日
				JP	2012533818	A	2012年 12月 27日
CN	101236567	A	2008年 8月 06日	无			
CN	103164405	A	2013年 6月 19日	无			
CN	102326144	A	2012年 1月 18日	US	2009216696	A1	2009年 8月 27日
				CA	2805391	A	2009年 9月 03日
				WO	2009108726	A1	2009年 9月 03日
				CA	2716062	A	2009年 9月 03日
				WO	2010068931	A1	2010年 6月 17日
				US	2010153324	A1	2010年 6月 17日
				CA	2746675	A	2010年 6月 17日
				EP	2260373	A1	2010年 12月 15日
				CN	102016787	A	2011年 4月 13日
				JP	2011514995	A	2011年 5月 12日
				EP	2377011	A1	2011年 10月 19日
				MX	2011006340	A	2011年 10月 28日
				JP	2012512465	A	2012年 5月 31日
				US	2013066887	A1	2013年 3月 14日
				JP	2013218727	A	2013年 10月 24日
				US	2013325769	A1	2013年 12月 05日
CN	103488787	A	2014年 1月 01日	无			
CN	103500214	A	2014年 1月 08日	无			
CN	103491205	A	2014年 1月 01日	无			
US	2004064447	A1	2004年 4月 01日	GB	2393541	A	2004年 3月 31日
				DE	10328833	A1	2004年 4月 15日
				GB	2417115	A	2006年 2月 15日

表 PCT/ISA/210 (同族专利附件) (2009年7月)