



(12) 发明专利申请

(10) 申请公布号 CN 112699314 A

(43) 申请公布日 2021. 04. 23

(21) 申请号 202011569022.1

(22) 申请日 2020.12.25

(71) 申请人 百度在线网络技术(北京)有限公司  
地址 100085 北京市海淀区上地十街10号  
百度大厦三层

(72) 发明人 林泽诚 张明

(74) 专利代理机构 北京市铸成律师事务所  
11313  
代理人 阎敏 邓海鸿

(51) Int. Cl.

G06F 16/9537 (2019.01)

G06F 16/951 (2019.01)

G06F 40/30 (2020.01)

G06N 20/00 (2019.01)

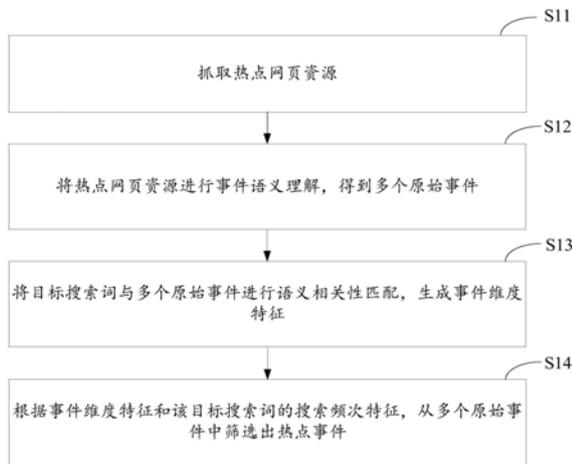
权利要求书2页 说明书8页 附图7页

(54) 发明名称

热点事件确定方法、装置、电子设备及存储介质

(57) 摘要

本申请公开了热点事件确定方法、装置、电子设备、存储介质以及计算机程序产品,涉及互联网技术领域、信息挖掘及搜索技术领域。具体实现方案为:抓取热点网页资源;将热点网页资源进行事件语义理解,得到多个原始事件;将目标搜索词与多个原始事件进行语义相关性匹配,生成事件维度特征;根据事件维度特征以及目标搜索词的搜索频次特征,从多个原始事件中筛选出热点事件。根据本申请的技术方案,能提高识别全网热点事件的效率和准确率。



1. 一种热点事件确定方法,包括:  
抓取热点网页资源;  
将所述热点网页资源进行事件语义理解,得到多个原始事件;  
将目标搜索词与所述多个原始事件进行语义相关性匹配,生成事件维度特征;  
根据所述事件维度特征以及所述目标搜索词的搜索频次特征,从所述多个原始事件中筛选出热点事件。
2. 根据权利要求1所述的方法,还包括:  
根据是否运营所述热点事件的指示信息,确定目标热点事件。
3. 根据权利要求2所述的方法,还包括:  
针对所述目标热点事件进行物料挖掘,得到所述目标热点事件对应的物料资源。
4. 根据权利要求1所述的方法,还包括:  
获取用于进行事件语义理解的目标语义模型,其中,所述目标语义模型是基于第一训练样本对预设语义模型进行训练后所得到的;  
所述将所述热点网页资源进行事件语义理解,得到多个原始事件,包括:  
将所述热点网页资源输入所述目标语义模型,得到所述目标语义模型输出的多个原始事件。
5. 根据权利要求1所述的方法,还包括:  
获取目标事件模型,其中,所述目标事件模型是基于第二训练样本对预设事件模型进行训练后所得到的,用于根据事件维度特征以及搜索频次特征筛选热点事件;  
所述根据所述事件维度特征以及所述目标搜索词的搜索频次特征,从所述多个原始事件中筛选出热点事件,包括:  
将所述事件维度特征以及所述目标搜索词的搜索频次特征输入所述目标事件模型,得到所述目标事件模型输出的从所述多个原始事件中筛选出的热点事件。
6. 根据权利要求5所述的方法,其中,基于第二训练样本对预设事件模型进行训练,包括:  
基于所述第二训练样本对应的样本事件维度特征以及样本搜索频次特征对预设事件模型进行训练。
7. 一种热点事件确定装置,包括:  
抓取单元,用于抓取热点网页资源;  
事件提取单元,用于将所述热点网页资源进行事件语义理解,得到多个原始事件;  
特征提取单元,用于将目标搜索词与所述多个原始事件进行语义相关性匹配,生成事件维度特征;  
第一确定单元,用于根据所述事件维度特征以及所述目标搜索词的搜索频次特征,从所述多个原始事件中筛选出热点事件。
8. 根据权利要求7所述的装置,还包括:  
第二确定单元,用于根据是否运营所述热点事件的指示信息,确定目标热点事件。
9. 根据权利要求8所述的装置,还包括:  
物料挖掘单元,用于针对所述目标热点事件进行物料挖掘,得到所述目标热点事件对应的物料资源。

10. 根据权利要求7所述的装置,还包括:

第一获取单元,用于获取用于进行事件语义理解的目标语义模型,其中,所述目标语义模型是基于第一训练样本对预设语义模型进行训练后所得到的;

所述事件提取单元,用于:

将所述热点网页资源输入所述目标语义模型,得到所述目标语义模型输出的多个原始事件。

11. 根据权利要求7所述的装置,还包括:

第二获取单元,用于获取目标事件模型,其中,所述目标事件模型是基于第二训练样本对预设事件模型进行训练后所得到的,用于根据事件维度特征以及搜索频次特征筛选热点事件;

其中,所述第一确定单元,用于:

将所述事件维度特征以及所述目标搜索词的搜索频次特征输入所述目标事件模型,得到所述目标事件模型输出的从所述多个原始事件中筛选出的热点事件。

12. 根据权利要求11所述的装置,其中,所述目标事件模型是基于所述第二训练样本对应的样本事件维度特征以及样本搜索频次特征对预设事件模型进行训练后所得到的。

13. 一种电子设备,其特征在于,包括:

至少一个处理器;以及

与所述至少一个处理器通信连接的存储器;其中,

所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行权利要求1-6中任一项所述的方法。

14. 一种存储有计算机指令的非瞬时计算机可读存储介质,所述计算机指令用于使计算机执行权利要求1-6中任一项所述的方法。

15. 一种计算机程序产品,包括计算机程序,该计算机程序在被处理器执行时实现根据权利要求1-6中任一项所述的方法。

## 热点事件确定方法、装置、电子设备及存储介质

### 技术领域

[0001] 本申请涉及互联网技术领域,尤其涉及信息挖掘及搜索技术领域。

### 背景技术

[0002] 热点事件运营一直都是资讯类应用重点运营的对象。全网每天会爆发众多事件,真正达到运营级别的事件并不多。人工挖掘热点事件,并根据经验判断当前事件是否是热点事件的成本很高,效率低。

### 发明内容

[0003] 本申请提供了一种热点事件确定方法、装置、电子设备、存储介质以及计算机程序产品。

[0004] 根据本申请的一方面,提供了一种热点事件确定方法,包括:

[0005] 抓取热点网页资源;

[0006] 将热点网页资源进行事件语义理解,得到多个原始事件;

[0007] 将目标搜索词与多个原始事件进行语义相关性匹配,生成事件维度特征;

[0008] 根据事件维度特征以及目标搜索词的搜索频次特征,从多个原始事件中筛选出热点事件。

[0009] 根据本申请的另一方面,提供了一种热点事件确定装置,包括:

[0010] 抓取单元,用于抓取热点网页资源;

[0011] 事件提取单元,用于将热点网页资源进行事件语义理解,得到多个原始事件;

[0012] 特征提取单元,用于将目标搜索词与多个原始事件进行语义相关性匹配,生成事件维度特征;

[0013] 第一确定单元,用于根据事件维度特征以及目标搜索词的搜索频次特征,从多个原始事件中筛选出热点事件。

[0014] 根据本公开的另一方面,提供了一种电子设备,包括:

[0015] 至少一个处理器;以及

[0016] 与该至少一个处理器通信连接的存储器;其中,

[0017] 该存储器存储有可被该至少一个处理器执行的指令,该指令被该至少一个处理器执行,以使该至少一个处理器能够执行本申请任一实施例中的方法。

[0018] 根据本公开的另一方面,提供了一种存储有计算机指令的非瞬时计算机可读存储介质,该计算机指令用于使计算机执行本申请任一实施例中的方法。

[0019] 根据本公开的另一方面,提供了一种计算机程序产品,包括计算机程序,该计算机程序被处理器执行时实现本申请任一实施例中的方法。

[0020] 根据本申请的技术方案,能提高识别全网热点事件的效率和准确率。

[0021] 应当理解,本部分所描述的内容并非旨在标识本申请的实施例的关键或重要特征,也不用于限制本申请的范围。本申请的其它特征将通过以下的说明书而变得容易理解。

## 附图说明

- [0022] 附图用于更好地理解本方案,不构成对本申请的限定。其中:
- [0023] 图1是根据本申请一个实施例的热点事件确定方法的示意图一;
- [0024] 图2是根据本申请一个实施例的热点事件确定方法的示意图二;
- [0025] 图3是根据本申请一个实施例的热点事件确定方法的示意图三;
- [0026] 图4是根据本申请一个实施例的热点事件确定方法的示意图四;
- [0027] 图5是根据本申请一个实施例的热点事件确定方法的示意图五;
- [0028] 图6是本申请实施例中运营热点事件的架构示意图;
- [0029] 图7是根据本申请一个实施例的热点事件确定装置的示意图;
- [0030] 图8是根据本申请又一个实施例的热点事件确定装置的示意图;
- [0031] 图9是用来实现本申请实施例的热点事件确定方法的电子设备的框图。

## 具体实施方式

[0032] 以下结合附图对本申请的示范性实施例做出说明,其中包括本申请实施例的各种细节以助于理解,应当将它们认为仅仅是示范性的。因此,本领域普通技术人员应当认识到,可以对这里描述的实施例做出各种改变和修改,而不会背离本申请的范围和精神。同样,为了清楚和简明,以下的描述中省略了对公知功能和结构的描述。

[0033] 图1示出了根据本申请一个实施例的热点事件确定方法的示意图,该方法可以应用于电子设备,该电子设备包括但不限于固定设备和/或移动设备,例如,固定设备包括但不限于服务器,服务器可以是云服务器或普通服务器。例如,移动设备包括但不限于:手机或平板电脑中的一项或是多项。如图1所示,该方法包括:

[0034] 步骤S11,抓取热点网页资源;

[0035] 步骤S12,将热点网页资源进行事件语义理解,得到多个原始事件;

[0036] 步骤S13,将目标搜索词与多个原始事件进行语义相关性匹配,生成事件维度特征;

[0037] 步骤S14,根据事件维度特征以及目标搜索词的搜索频次特征,从多个原始事件中筛选出热点事件。

[0038] 其中,热点网页资源是利用爬虫对全网资讯网站进行网页爬取到的,实现第一时间获取完整突发热点网页。

[0039] 其中,原始事件是待进一步筛选或确认的事件。步骤S12中的多个原始事件中,有的原始事件可能不满足后续的筛选条件,有的原始事件可能能够满足后续的筛选条件,而那些能满足后续的筛选条件的原始事件,将被判定为热点事件。

[0040] 其中,步骤S13中的目标搜索词是搜索频次超出预设阈值的搜索词。由于不同搜索词的搜索次数可能不同,例如,在预设时间段内搜索词A对应的搜索次数为*i*,搜索词B对应的搜索次数为*j*,其中,*i*的值为正整数,*j*的值为正整数,若*i*大于预设阈值,则搜索词A被视为目标搜索词;若*j*小于预设阈值,则搜索词B不被视为目标搜索词。实际应用中,该预设阈值可根据需求进行设定或调整。

[0041] 其中,步骤S13中的目标搜索词(也可称为query)是用户在终端设备上输入的信息。示例性地,该目标搜索词可以包含下述信息中的一项或多项:人物名称、物品名称、时

间、地点、关键词。举例来说,用户输入了“刘德华”,则搜索词为“刘德华”。又举例来说,用户输入了“十一国庆”,则搜索词为“十一国庆”。

[0042] 其中,事件可以有多个维度,事件维度特征是用来表征事件的维度的特征,本申请并不对事件的维度的个数进行限定,具体个数可根据设计需求进行设定或调整。比如,事件维度特征包括下述特征中的一项或多项:当前事件的核心词、主要摘要、站点权重以及全网爆发的相关站点的多少。这里,站点是指网站。

[0043] 其中,搜索频次特征是用来表征用户搜索行为的特征。比如,搜索频次特征包括下述特征中的一项或多项:用户搜索的搜索词频次,用户搜索的搜索词跟事件相关度的相关系数。

[0044] 这样,本申请方案通过将抓取到的热点网页资源进行事件语义理解,得到多个原始事件,能够快速从全网初步筛选出有可能成为热点事件的原始事件;通过将目标搜索词与多个原始事件进行语义相关性匹配,生成事件维度特征;根据事件维度特征以及目标搜索词的搜索频次特征,从多个原始事件中筛选出热点事件,能够对初步筛选出的原始事件做进一步筛选,得到热点事件,从而能够实现自动识别全网的热点事件;本申请方案相对于人工挖掘热点事件的方式而言,挖掘范围更广,挖掘速度更快;结合搜索频次特征和事件维度特征筛选热点事件,能够提高识别全网热点事件的效率和准确率。

[0045] 在本申请实施例中,在实现上述任一方法的基础上,如图2所示,该方法还可以包括:

[0046] 步骤S15,根据是否运营热点事件的指示信息,确定目标热点事件。

[0047] 其中,目标热点事件是指待运营的热点事件。

[0048] 在一些实施方式中,电子设备将步骤S15中筛选出的热点事件推送至运营人员所在终端,由运营人员对热点事件进行人工筛选,确定是否运营该热点事件;电子设备根据运营人员所在终端发送的指示信息,从多个热点事件中进一步筛选出目标热点事件。

[0049] 通过上述实施方式,能够提高热点事件的运营效率和准确率。由于运营人员无需人工挖掘热点事件,只需根据事件时效性、热度等特征从接收到的热点事件中选择将要运营的目标热点事件,能够提升工作效率,降低人工成本。

[0050] 在本申请实施例中,在实现上述任一方法的基础上,如图3所示,该方法还可以包括:

[0051] 步骤S16,针对目标热点事件进行物料挖掘,得到目标热点事件对应的物料资源。

[0052] 其中,物料是指运营一个目标热点事件所需的材料。比如,物料包括下述中的一项或多项:文章;图片;视频;音频;链接地址。

[0053] 在一些实施方式中,针对目标热点事件进行物料挖掘,包括下述中的一项或多项:

[0054] 针对目标热点事件进行全网优质资讯挖掘;

[0055] 针对目标热点事件进行全网图片挖掘;

[0056] 针对目标热点事件进行全网视频挖掘;

[0057] 针对目标热点事件进行全网音频挖掘。

[0058] 通过上述实施方式,能够自动进行在线物料筛选,实现事件物料实时挖掘,从而便于运营人员快速找到运营服务需求的资源,提升运营效率。

[0059] 在本申请实施例中,在实现上述任一方法的基础上,如图4所示,该方法还可以包

括以下步骤:

[0060] 步骤S17,获取用于进行事件语义理解的目标语义模型,其中,该目标语义模型是基于第一训练样本对预设语义模型进行训练后所得到的。

[0061] 在一些实施方式中,将热点网页资源进行事件语义理解,得到多个原始事件,包括:

[0062] 将热点网页资源输入该目标语义模型,得到该目标语义模型输出的多个原始事件。

[0063] 这里,预设语义模型可以采用各种预训练语言模型,如基于转换器的双向编码表示(Bidirectional Encoder Representations from Transformers,BERT)模型,知识增强的语义表示(Enhanced Representation from kNowledge IntEgration,ERNIE)模型等。

[0064] 需要说明的是,本申请并不对预设语义模型的训练方式和训练过程进行限定。

[0065] 通过上述实施方式,能够得到目标语义模型,进而将抓取的热点网页资源输入该目标语义模型,能够自动输出原始事件,相对于采用人工方式筛选原始事件而言,能提高筛选原始事件的效率和准确率,从而有助于提升识别热点事件的效率和准确率。

[0066] 在本申请实施例中,在实现上述任一方法的基础上,如图5所示,步骤S11之前,还包括:

[0067] 步骤S18,获取目标事件模型,其中,该目标事件模型是基于第二训练样本对预设事件模型进行训练后所得到的,用于根据事件维度特征以及搜索频次特征筛选热点事件。

[0068] 需要说明的是,该第二训练样本与上文提到的第一训练样本可以是相同的训练样本,也可以是不同的训练样本。

[0069] 在一些实施方式中,根据事件维度特征以及目标搜索词的搜索频次特征,从多个原始事件中筛选出热点事件,包括:

[0070] 将事件维度特征以及目标搜索词的搜索频次特征输入目标事件模型,得到所述目标事件模型输出的从多个原始事件中筛选出的热点事件。

[0071] 这里,预设事件模型可以采用各种机器学习算法模型,如梯度提升决策树(Gradient Boosting Decision Tree,GBDT)模型,极端梯度提升(eXtreme Gradient Boosting,XGBoost)模型等。

[0072] 实际应用中,目标事件模型筛选出的热点事件可以列表的形式输出,比如,每隔一段时间生成一个热点事件列表,供运营人员根据事件时效性、热度等特征确定是否运营当前热点事件。

[0073] 通过上述实施方式,能够得到目标事件模型,进而将搜索频次特征与原始事件输入该目标事件模型,能够自动输出热点事件,相对于采用人工方式挖掘热点事件而言,能提高挖掘热点事件的效率和准确率,从而有助于提升识别热点事件的效率和准确率。

[0074] 在本申请实施例中,在实现上述任一方法的基础上,基于第二训练样本对预设事件模型进行训练,包括:

[0075] 基于所述第二训练样本对应的样本事件维度特征以及样本搜索频次特征对预设事件模型进行训练。

[0076] 上述实施方式中,训练预设事件模型时考虑样本搜索频次特征与样本事件维度特征,能提升训练得到的目标事件模型输出热点事件的准确率。

[0077] 图6示出了运营热点事件的架构示意图,从图6可以看出,将全网资讯流输入目标语义模型,由目标语义模型进行事件生成,并输出原始事件至原始事件库,根据在线实时高频搜索词对原始事件库中的原始事件进行语义关联,得到事件维度特征和搜索频次特征;将事件维度特征和搜索频次特征输入目标事件模型,由目标事件模型对原始事件进行打分,根据打分结果确定热点事件,比如,将打分值排名前M的原始事件作为热点事件;由运营人员从热点事件中筛选出待运营的目标热点事件;基于热点事件进行物料挖掘,得到完整的待运营的目标热点事件。其中,该目标语义模型可以采用BERT模型,该目标事件模型可以采用GBDT模型。可见,通过该架构能自动识别全网热点,并挖掘热点事件对应的优质物料。运营人员通过该架构可以更快速地定位热点,并完成运营,提升热点运营效率。

[0078] 应理解,图6所示的架构图为一种可选的具体实现方式,本领域技术人员可以基于图6的例子进行各种显而易见的变化和/或替换,得到的技术方案仍属于本公开实施例的公开范围。

[0079] 本申请提供的热点事件确定方法,可以用于搜索引擎或搜索推荐等项目中。示例性地,方法的执行主体可以是电子设备,该电子设备可以是各种搜索引擎设备,如搜索引擎服务器。

[0080] 作为对上述各方法的实现,本申请还提供一种热点事件确定装置。图7示出了热点事件确定装置的示意图。如图7所示,该装置包括:

[0081] 抓取单元710,用于抓取热点网页资源;

[0082] 事件提取单元720,用于将该热点网页资源进行事件语义理解,得到多个原始事件;

[0083] 特征提取单元730,用于将目标搜索词与该多个原始事件进行语义相关性匹配,生成事件维度特征;

[0084] 第一确定单元740,用于根据该事件维度特征以及该目标搜索词的搜索频次特征,从该多个原始事件中筛选出热点事件。

[0085] 在一些实施方式中,如图8所示,该装置还可以包括:

[0086] 第二确定单元750,用于根据是否运营该热点事件的指示信息,确定目标热点事件。

[0087] 在一些实施方式中,如图8所示,该装置还可以包括:

[0088] 物料挖掘单元760,用于针对该目标热点事件进行物料挖掘,得到该目标热点事件对应的物料资源。

[0089] 在一些实施方式中,如图8所示,该装置还可以包括:

[0090] 第一获取单元770,用于获取用于进行事件语义理解的目标语义模型,其中,该目标语义模型是基于第一训练样本对预设语义模型进行训练后所得到的;

[0091] 该事件提取单元730,用于:

[0092] 将该热点网页资源输入该目标语义模型,得到该目标语义模型输出的多个原始事件。

[0093] 在一些实施方式中,如图8所示,该装置还可以包括:

[0094] 第二获取单元780,用于获取目标事件模型,其中,该目标事件模型是基于第二训练样本对预设事件模型进行训练后所得到的,用于根据事件维度特征以及搜索频次特征筛

选热点事件；

[0095] 其中,该第一确定单元740,用于:

[0096] 将该事件维度特征以及该目标搜索词的搜索频次特征输入该目标事件模型,得到该目标事件模型输出的从该多个原始事件中筛选出的热点事件。

[0097] 其中,该目标事件模型是基于该第二训练样本对应的样本事件维度特征以及样本搜索频次特征对预设事件模型进行训练后所得到的。

[0098] 本实施例的热点事件确定装置,能够自动挖掘热点事件,相对于采用人工方式挖掘热点事件而言,能提高挖掘热点事件的效率和准确率,从而有助于提升运营热点事件的效率和准确率。

[0099] 根据本申请的实施例,本申请还提供了一种电子设备、一种可读存储介质和一种计算机程序产品。

[0100] 如图9所示,是根据本申请实施例的热点事件确定方法的电子设备的框图。电子设备旨在表示各种形式的数字计算机,诸如,膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置,诸如,个人数字处理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作为示例,并且不意在限制本文中描述的和/或者要求的本申请的实现。

[0101] 如图9所示,该电子设备包括:一个或多个处理器801、存储器802,以及用于连接各部件的接口,包括高速接口和低速接口。各个部件利用不同的总线互相连接,并且可以被安装在公共主板上或者根据需要以其它方式安装。处理器可以对在电子设备内执行的指令进行处理,包括存储在存储器中或者存储器上以在外部输入/输出装置(诸如,耦合至接口的显示设备)上显示图形用户界面(Graphical User Interface,GUI)的图形信息的指令。在其它实施方式中,若需要,可以将多个处理器和/或多条总线与多个存储器和多个存储器一起使用。同样,可以连接多个电子设备,各个设备提供部分必要的操作(例如,作为服务器阵列、一组刀片式服务器、或者多处理器系统)。图9中以一个处理器801为例。

[0102] 存储器802即为本申请所提供的非瞬时计算机可读存储介质。其中,存储器存储有可由至少一个处理器执行的指令,以使至少一个处理器执行本申请所提供的热点事件确定方法。本申请的非瞬时计算机可读存储介质存储计算机指令,该计算机指令用于使计算机执行本申请所提供的热点事件确定方法。

[0103] 存储器802作为一种非瞬时计算机可读存储介质,可用于存储非瞬时软件程序、非瞬时计算机可执行程序以及模块,如本申请实施例中的热点事件确定方法对应的程序指令/模块(例如,附图7所示的抓取单元710、事件提取单元720、特征提取单元730、第一确定单元740、第二确定单元750、物料挖掘单元760、第一获取单元770和第二获取单元780)。处理器801通过运行存储在存储器802中的非瞬时软件程序、指令以及模块,从而执行服务器的各种功能应用以及数据处理,即实现上述方法实施例中的热点事件确定方法。

[0104] 存储器802可以包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需要的应用程序;存储数据区可存储根据热点事件确定方法的电子设备的使用所创建的数据等。此外,存储器802可以包括高速随机存取存储器,还可以包括非瞬时存储器,例如至少一个磁盘存储器件、闪存器件、或其他非瞬时固态存储器件。在一些

实施例中,存储器802可选包括相对于处理器801远程设置的存储器,这些远程存储器可以通过网络连接至热点事件确定方法的电子设备。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

[0105] 热点事件确定方法的电子设备还可以包括:输入装置803和输出装置804。处理器801、存储器802、输入装置803和输出装置804可以通过总线或者其他方式连接,图9中以通过总线连接为例。

[0106] 输入装置803可接收输入的数字或字符信息,以及产生与热点事件确定方法的电子设备的用户设置以及功能控制有关的键信号输入,例如触摸屏、小键盘、鼠标、轨迹板、触摸板、指示杆、一个或者多个鼠标按钮、轨迹球、操纵杆等输入装置。输出装置804可以包括显示设备、辅助照明装置(例如,LED)和触觉反馈装置(例如,振动电机)等。该显示设备可以包括但不限于,液晶显示器(Liquid Crystal Display,LCD)、发光二极管(Light Emitting Diode,LED)显示器和等离子体显示器。在一些实施方式中,显示设备可以是触摸屏。

[0107] 根据本申请的实施例,本申请还提供了一种电子设备。该设备可以包括:

[0108] 一个或多个处理器;以及

[0109] 存储装置,用于存储一个或多个程序,当一个或多个程序被一个或多个处理器执行,使得一个或多个处理器实现上述方法实施例中的搜索词分推荐方法。

[0110] 其中,电子设备的处理器和存储装置的功能以及实现方式,可以参考上述电子设备实施例中的关于处理器和存储器的描述。

[0111] 此处描述的系统和技术各种实施方式可以在数字电子电路系统、集成电路系统、专用集成电路(Application Specific Integrated Circuit,ASIC)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括:实施在一个或者多个计算机程序中,该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释,该可编程处理器可以是专用或者通用可编程处理器,可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令,并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

[0112] 这些计算程序(也称作程序、软件、软件应用、或者代码)包括可编程处理器的机器指令,并且可以利用高级过程和/或面向对象的编程语言、和/或汇编/机器语言来实施这些计算程序。如本文使用的,术语“机器可读介质”和“计算机可读介质”指的是用于将机器指令和/或数据提供给可编程处理器的任何计算机程序产品、设备、和/或装置(例如,磁盘、光盘、存储器、可编程逻辑装置(Programmable Logic Devices,PLD)),包括,接收作为机器可读信号的机器指令的机器可读介质。术语“机器可读信号”指的是用于将机器指令和/或数据提供给可编程处理器的任何信号。

[0113] 为了提供与用户的交互,可以在计算机上实施此处描述的系统和技术,该计算机具有:用于向用户显示信息的显示装置(例如,阴极射线管(Cathode Ray Tube,CRT)或者LCD监视器);以及键盘和指向装置(例如,鼠标或者轨迹球),用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以用于提供与用户的交互;例如,提供给用户的反馈可以是任何形式的传感反馈(例如,视觉反馈、听觉反馈、或者触觉反馈);并且可以用任何形式(包括声输入、语音输入、或者触觉输入)来接收来自用户的输入。

[0114] 可以将此处描述的系统和技术实施在包括后台部件的计算系统(例如,作为数据

服务器)、或者包括中间件部件的计算系统(例如,应用服务器)、或者包括前端部件的计算系统(例如,具有图形用户界面或者网络浏览器的用户计算机,用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术的实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信(例如,通信网络)来将系统的部件相互连接。通信网络的示例包括:局域网(Local Area Network,LAN)、广域网(Wide Area Network,WAN)和互联网。

[0115] 计算机系统可以包括客户端和服务端。客户端和服务端一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务端的关系。服务器可以是云服务器,又称为云计算服务器或云主机,是云计算服务体系中的一项主机产品,以解决传统物理主机与虚拟专用服务器(Virtual private server,VPS)服务中,存在的管理难度大,业务扩展性弱的缺陷。服务器也可以为分布式系统的服务器,或者是结合了区块链的服务器。

[0116] 根据本申请实施例的技术方案,将抓取到的热点网页资源进行事件语义理解,得到多个原始事件;基于搜索频次特征与多个原始事件进行语义相关性匹配,生成事件维度特征;根据事件维度特征以及目标搜索词的搜索频次特征,从多个原始事件中筛选出热点事件;如此,能够实现自动识别全网的热点事件,本申请方案相对于人工筛选热点事件的方式而言,筛选范围更广,筛选速度更快;结合搜索频次特征和事件维度特征筛选热点事件,能够提高识别全网热点事件的效率和准确率。

[0117] 应该理解,可以使用上面所示的各种形式的流程,重新排序、增加或删除步骤。例如,本申请中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行,只要能够实现本申请公开的技术方案所期望的结果,本文在此不进行限制。

[0118] 上述具体实施方式,并不构成对本申请保护范围的限制。本领域技术人员应该明白的是,根据设计要求和因素,可以进行各种修改、组合、子组合和替代。任何在本申请的精神和原则之内所作的修改、等同替换和改进等,均应包含在本申请保护范围之内。

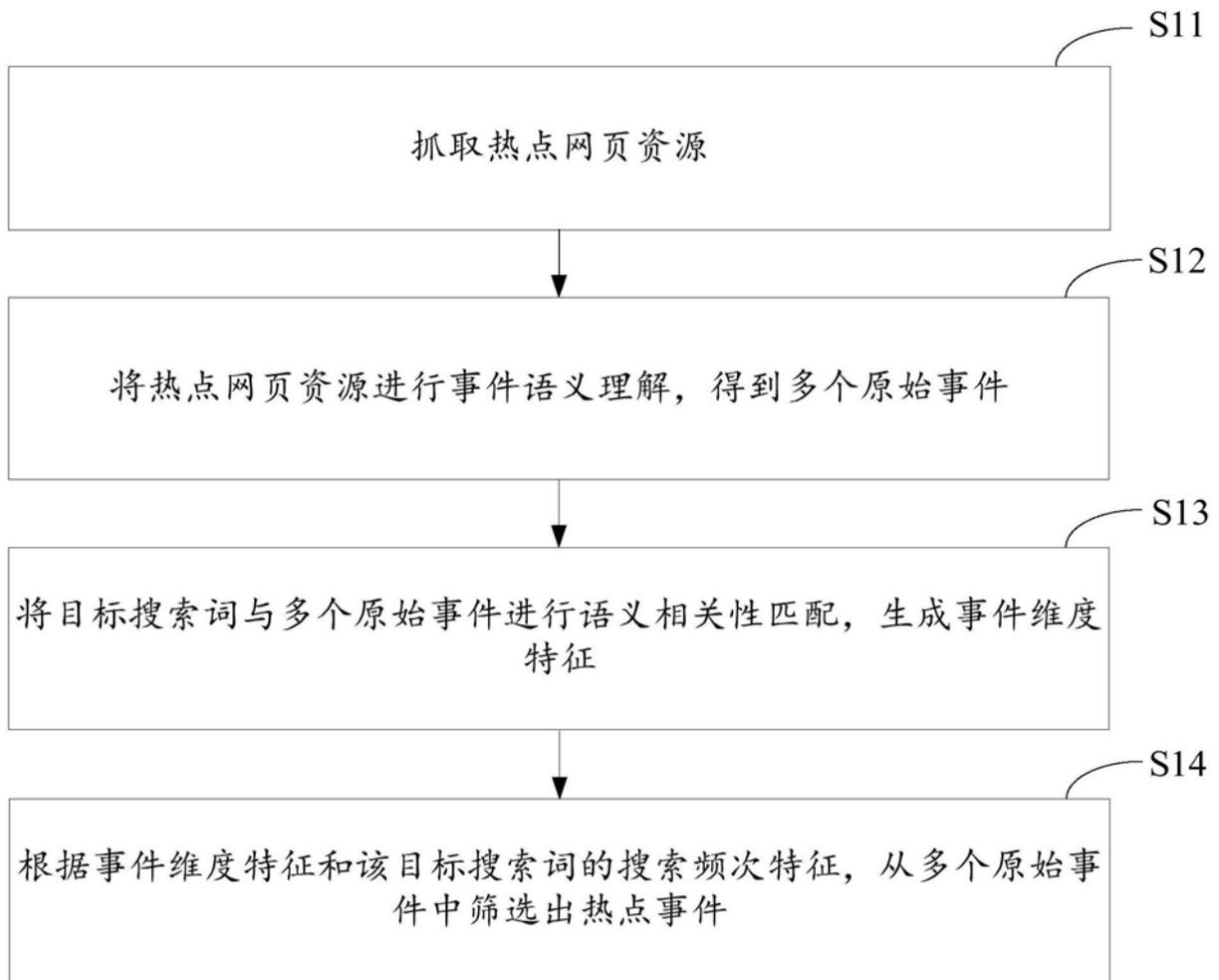


图1

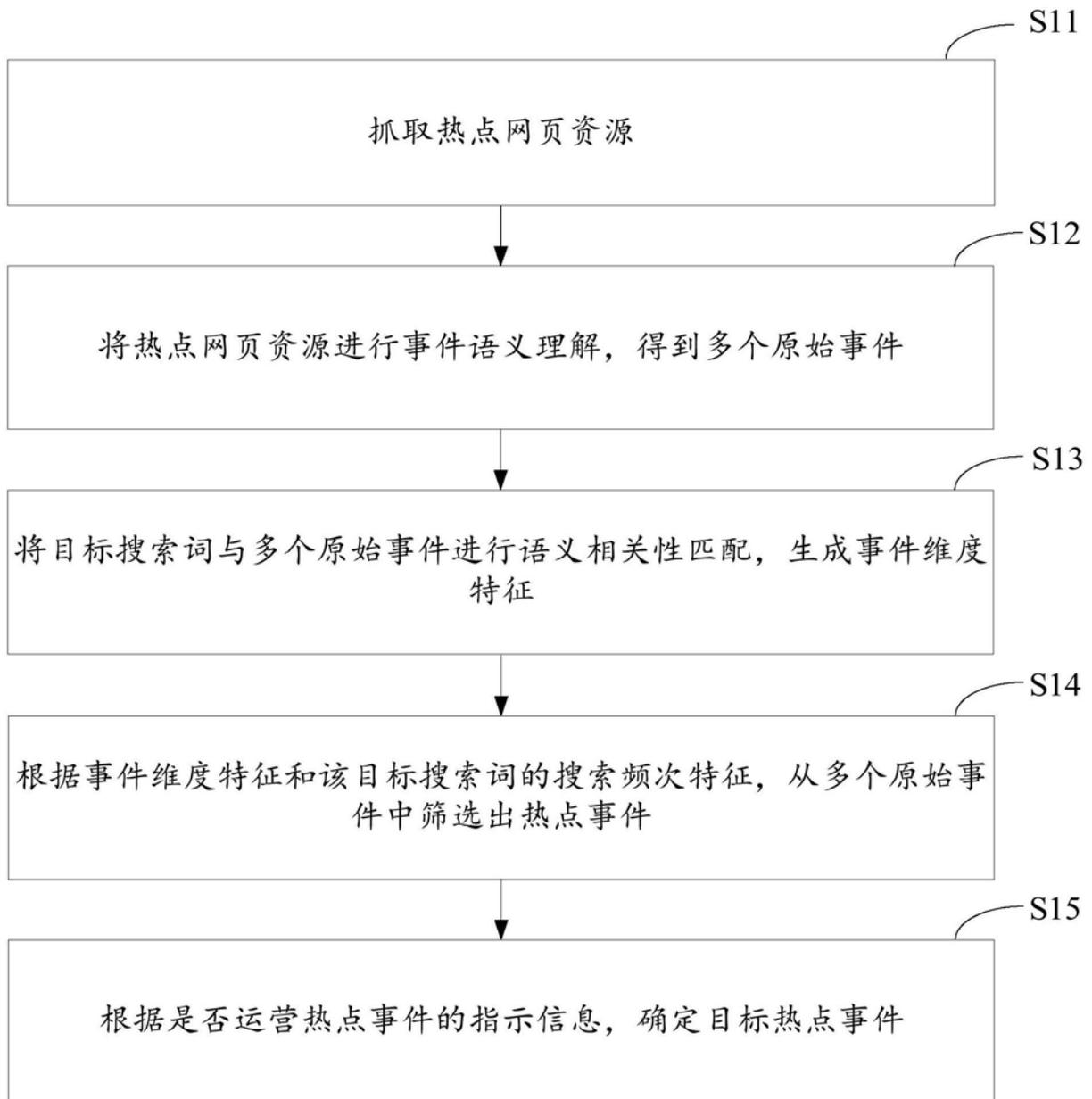


图2

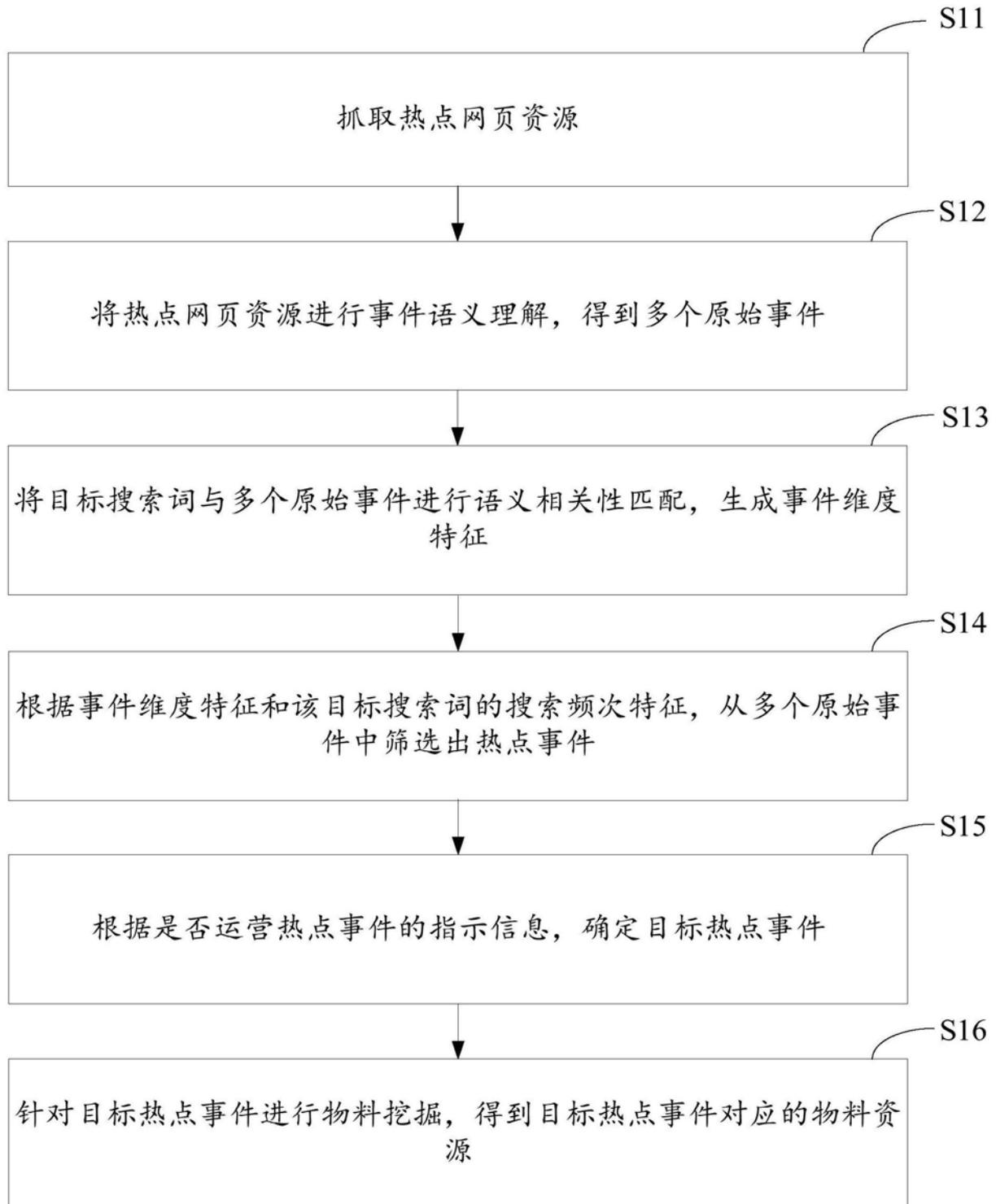


图3

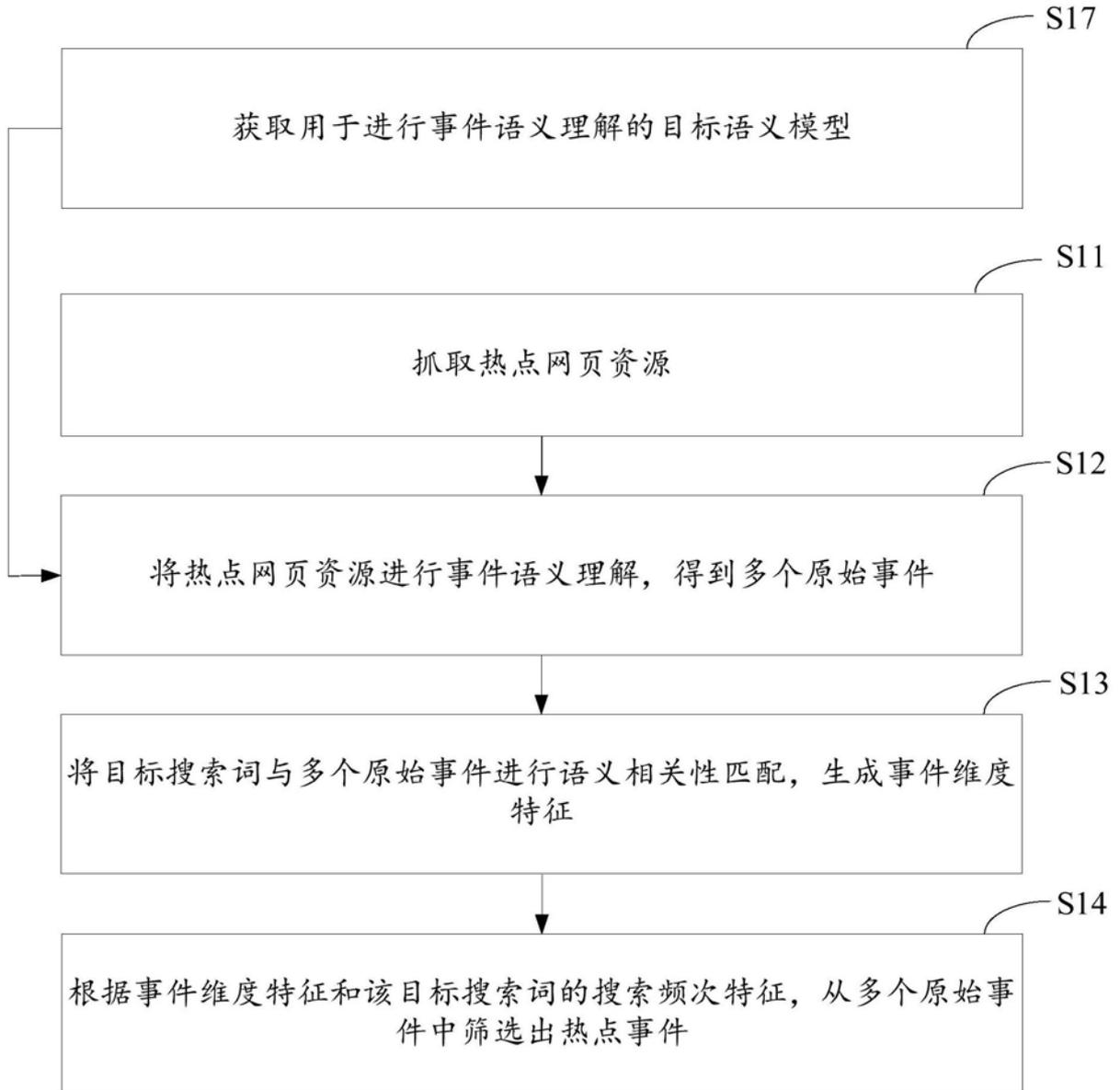


图4

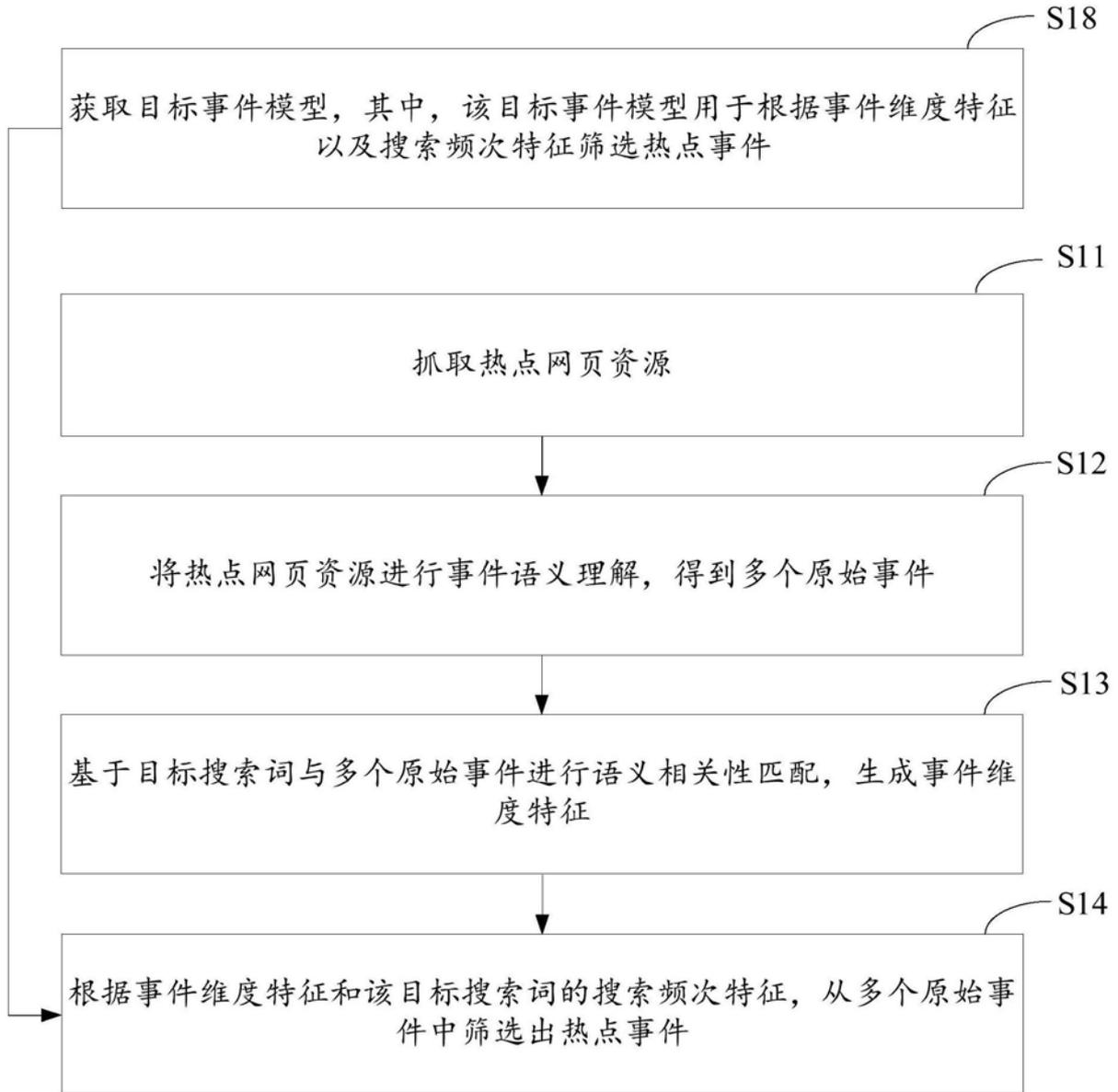


图5

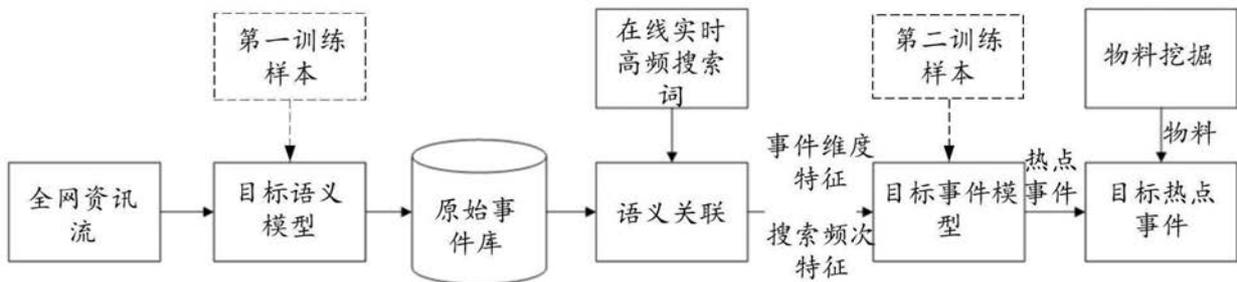


图6

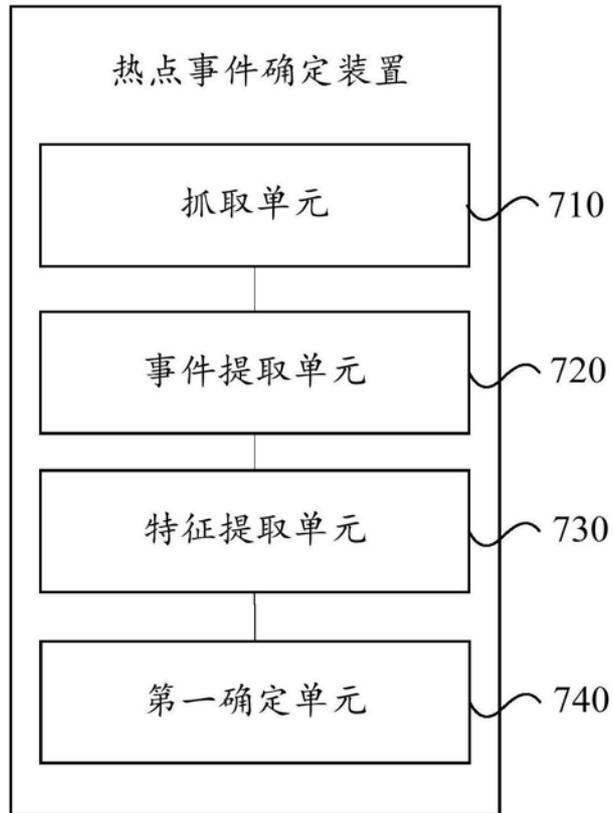


图7

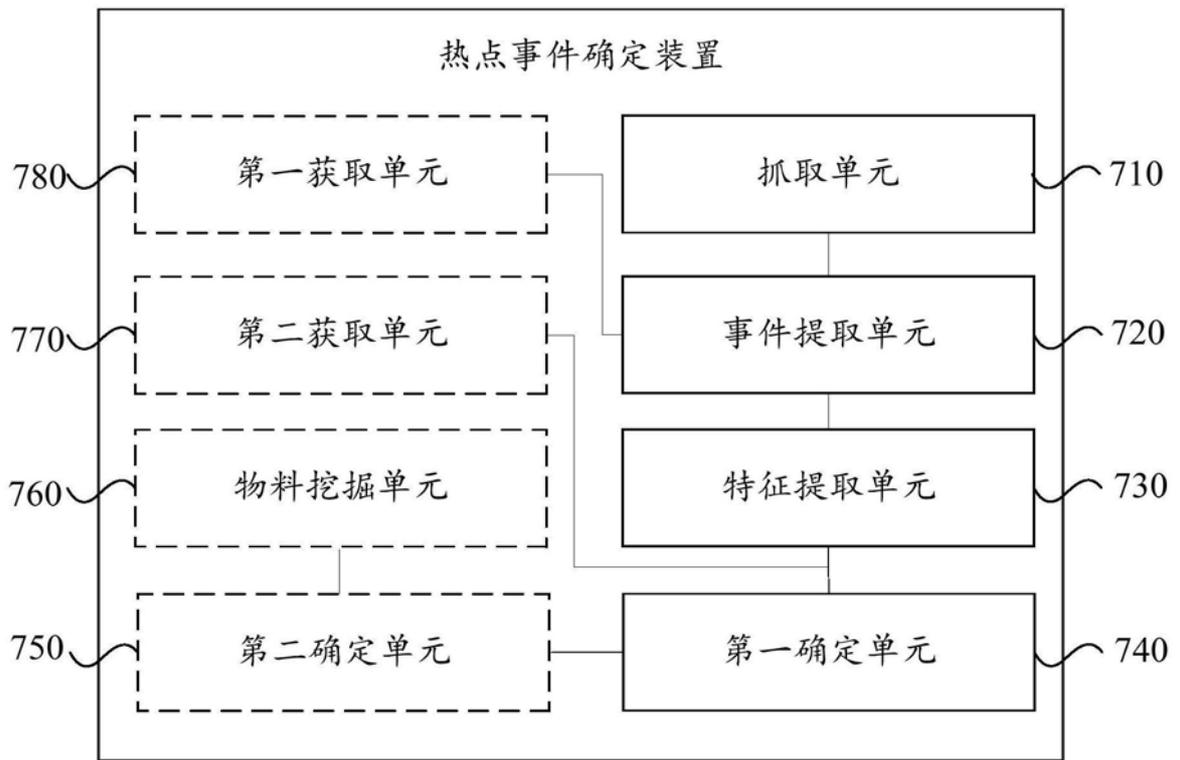


图8

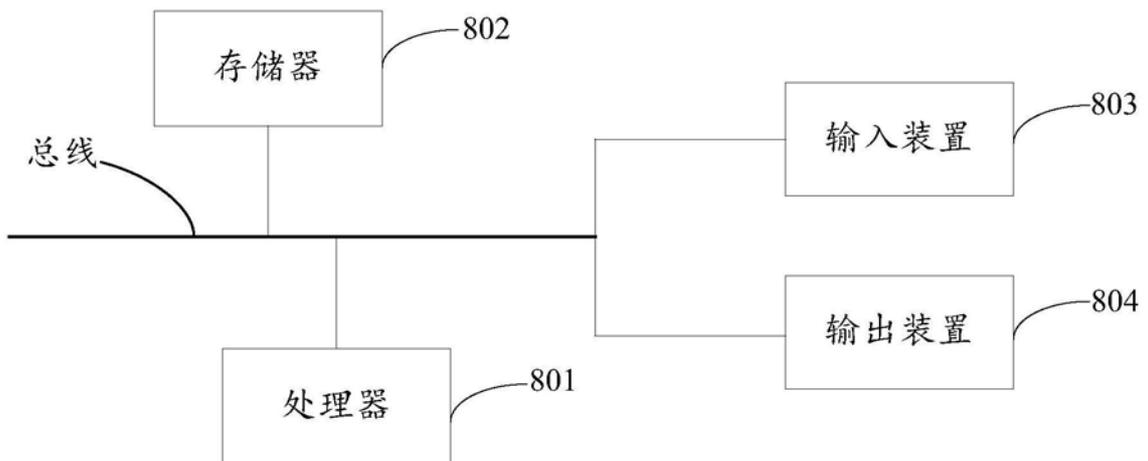


图9