



(12) 发明专利

(10) 授权公告号 CN 109635203 B

(45) 授权公告日 2020.12.25

(21) 申请号 201811566906.4

(22) 申请日 2018.12.19

(65) 同一申请的已公布的文献号
申请公布号 CN 109635203 A

(43) 申请公布日 2019.04.16

(73) 专利权人 北京达佳互联信息技术有限公司
地址 100084 北京市海淀区中关村东路1号
院8号楼20层B2201

(72) 发明人 刘明

(74) 专利代理机构 北京三高永信知识产权代理
有限责任公司 11138
代理人 唐述灿

(51) Int. Cl.
G06F 16/9535 (2019.01)
G06F 16/95 (2019.01)

(56) 对比文件

CN 103020239 A, 2013.04.03

CN 102982174 A, 2013.03.20

CN 102651022 A, 2012.08.29

CN 101127043 A, 2008.02.20

CN 101055587 A, 2007.10.17

US 7136851 B2, 2006.11.14

审查员 李骁

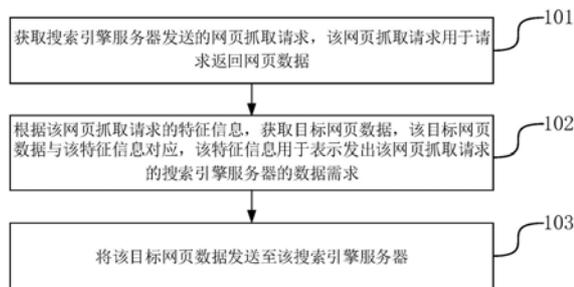
权利要求书3页 说明书14页 附图6页

(54) 发明名称

网页抓取请求处理方法、装置、服务器及存储介质

(57) 摘要

本公开是关于一种网页抓取请求处理方法、装置、服务器及存储介质,属于网络技术领域。该方法包括:获取搜索引擎服务器发送的网页抓取请求;根据所述网页抓取请求的特征信息,获取目标网页数据;将所述目标网页数据发送至所述搜索引擎服务器。该方法通过网页抓取请求的特征信息,来有针对性的为搜索引擎服务器获取目标网页数据,由于特征信息能够区分各个搜索引擎服务器,因此,基于该特征信息的获取,能够满足不同搜索引擎服务器的不同数据需求,也即能够实现对不同的搜索引擎服务器返回定制化的网页数据。



1. 一种网页抓取请求处理方法,其特征在于,应用于搜索引擎优化服务器,所述方法包括:

接收代理服务器发送的目标网页抓取请求,所述目标网页抓取请求为搜索引擎服务器发送的网页抓取请求中,符合预设白名单的网页抓取请求,所述网页抓取请求用于请求返回网页数据;

根据所述目标网页抓取请求的特征信息,获取目标网页数据,所述目标网页数据与所述特征信息对应,所述特征信息用于表示发出所述目标网页抓取请求的所述搜索引擎服务器的数据需求,所述特征信息为从所述目标网页抓取请求的目标字段中提取到的搜索引擎标识信息或者关键字字符串,所述关键字字符串包括资源类型信息和语言类型信息,所述资源类型信息用于指示所述搜索引擎服务器请求获取的资源类型,所述语言类型信息用于指示所述搜索引擎服务器所支持的语言类型;

通过所述代理服务器,将所述目标网页数据发送至所述搜索引擎服务器。

2. 根据权利要求1所述的网页抓取请求处理方法,其特征在于,所述根据所述目标网页抓取请求的特征信息,获取目标网页数据包括:

根据所述特征信息,查询与所述特征信息对应的缓存数据;

当查询到所述缓存数据时,将所述缓存数据获取为所述目标网页数据。

3. 根据权利要求1所述的网页抓取请求处理方法,其特征在于,所述根据所述目标网页抓取请求的特征信息,获取目标网页数据包括:

根据所述特征信息,查询与所述特征信息对应的缓存数据;

当查询不到所述缓存数据时,通过配置服务器获取所述目标网页数据,所述配置服务器用于为各个特征信息配置对应的网页数据。

4. 根据权利要求3所述的网页抓取请求处理方法,其特征在于,所述当查询不到所述缓存数据时,通过配置服务器获取所述目标网页数据之后,所述方法还包括:

将所述特征信息缓存为对象键名,所述对象键名用于指示下一次获取所述目标网页数据的索引;

将所述目标网页数据缓存为对象键值,所述对象键值用于指示下一次根据所述对象键名查询到的索引内容。

5. 根据权利要求3所述的网页抓取请求处理方法,其特征在于,所述当查询不到所述缓存数据时,通过配置服务器获取所述目标网页数据包括:

每间隔第一预设时长,根据所述特征信息,从所述配置服务器获取与所述特征信息对应的网页数据。

6. 根据权利要求1所述的网页抓取请求处理方法,其特征在于,所述根据所述目标网页抓取请求的特征信息,获取目标网页数据包括:

根据所述特征信息,查询与所述特征信息对应的缓存数据;

当查询不到所述缓存数据,且接收到配置服务器发送的配置失败信息时,通过网站服务器获取所述目标网页数据,所述配置服务器用于为各个特征信息配置对应的网页数据,所述配置失败信息用于指示所述配置服务器没有为所述特征信息配置对应网页数据。

7. 根据权利要求2至6中任一项权利要求所述的网页抓取请求处理方法,其特征在于,所述方法还包括:

每间隔第二预设时长,清空各个特征信息所对应的缓存数据。

8. 一种网页抓取请求处理装置,其特征在于,应用于搜索引擎优化服务器,所述装置包括:

第一获取单元,被配置为执行接收代理服务器发送的目标网页抓取请求,所述目标网页抓取请求为搜索引擎服务器发送的网页抓取请求中,符合预设白名单的网页抓取请求,所述网页抓取请求用于请求返回网页数据;

第二获取单元,被配置为执行根据所述目标网页抓取请求的特征信息,获取目标网页数据,所述目标网页数据与所述特征信息对应,所述特征信息用于表示发出所述目标网页抓取请求的所述搜索引擎服务器的数据需求,所述特征信息为从所述目标网页抓取请求的目标字段中提取到的搜索引擎标识信息或者关键字字符串,所述关键字字符串包括资源类型信息和语言类型信息,所述资源类型信息用于指示所述搜索引擎服务器请求获取的资源类型,所述语言类型信息用于指示所述搜索引擎服务器所支持的语言类型;

发送单元,被配置为执行通过所述代理服务器,将所述目标网页数据发送至所述搜索引擎服务器。

9. 根据权利要求8所述的网页抓取请求处理装置,其特征在于,所述第二获取单元包括:

查询子单元,被配置为执行根据所述特征信息,查询与所述特征信息对应的缓存数据;

获取子单元,被配置为执行当查询到所述缓存数据时,将所述缓存数据获取为所述目标网页数据。

10. 根据权利要求8所述的网页抓取请求处理装置,其特征在于,所述第二获取单元包括:

查询子单元,被配置为执行根据所述特征信息,查询与所述特征信息对应的缓存数据;

获取子单元,被配置为执行当查询不到所述缓存数据时,通过配置服务器获取所述目标网页数据,所述配置服务器用于为各个特征信息配置对应的网页数据。

11. 根据权利要求10所述的网页抓取请求处理装置,其特征在于,所述装置还包括:

缓存单元,被配置为执行将所述特征信息缓存为对象键名,所述对象键名用于指示下一次获取所述目标网页数据的索引;

所述缓存单元,还被配置为执行将所述目标网页数据缓存为对象键值,所述对象键值用于指示下一次根据所述对象键名查询到的索引内容。

12. 根据权利要求10所述的网页抓取请求处理装置,其特征在于,所述获取子单元还被配置为执行:

每间隔第一预设时长,根据所述特征信息,从所述配置服务器获取与所述特征信息对应的网页数据。

13. 根据权利要求8所述的网页抓取请求处理装置,其特征在于,所述第二获取单元包括:

查询子单元,被配置为执行根据所述特征信息,查询与所述特征信息对应的缓存数据;

获取子单元,被配置为执行当查询不到所述缓存数据,且接收到配置服务器发送的配置失败信息时,通过网站服务器获取所述目标网页数据,所述配置服务器用于为各个特征信息配置对应的网页数据,所述配置失败信息用于指示所述配置服务器没有为所述特征信

息配置对应网页数据。

14. 根据权利要求9至13中任一项权利要求所述的网页抓取请求处理装置,其特征在
于,所述装置还包括:

清空单元,被配置为执行每间隔第二预设时长,清空各个特征信息所对应的缓存数据。

15. 一种服务器,其特征在在于,包括:

处理器;

用于存储处理器可执行指令的存储器;

其中,所述处理器被配置为执行以实现如权利要求1至权利要求7任一项所述的网页抓
取请求处理方法所执行的操作。

16. 一种非临时性计算机可读存储介质,当所述存储介质中的指令由服务器的处理器
执行时,使得服务器能够执行一种以实现如权利要求1至权利要求7任一项所述的网页抓取
请求处理方法所执行的操作。

网页抓取请求处理方法、装置、服务器及存储介质

技术领域

[0001] 本公开涉及网络技术领域,尤其涉及一种网页抓取请求处理方法、装置、服务器及存储介质。

背景技术

[0002] 相关技术中,用户采用搜索引擎进行搜索时,搜索引擎服务器可以基于用户的搜索字,返回相应的网页数据的跳转链接。

[0003] 在上述过程中,搜索引擎服务器会不定时向各个网站服务器发出网页抓取请求,以获取各个网站当前的网页数据,从而存储到搜索引擎服务器上以备调用,其中,该网页抓取请求用于抓取网页数据。而对各个网站服务器而言,对于各个搜索引擎服务器发送的网页抓取请求,均会返回相同的网页数据,不能够满足不同搜索引擎服务器的不同数据需求,无法实现搜索引擎服务器的定制化搜索。

发明内容

[0004] 本公开提供一种网页抓取请求处理方法、装置、服务器及存储介质,能够满足不同搜索引擎服务器的不同数据需求,实现搜索引擎服务器的定制化搜索。

[0005] 根据本公开实施例的第一方面,提供一种网页抓取请求处理方法,该方法包括:

[0006] 获取搜索引擎服务器发送的网页抓取请求,该网页抓取请求用于请求返回网页数据;

[0007] 根据该网页抓取请求的特征信息,获取目标网页数据,该目标网页数据与该特征信息对应,该特征信息用于表示发出该网页抓取请求的搜索引擎服务器的数据需求;

[0008] 将该目标网页数据发送至该搜索引擎服务器。

[0009] 在一种可能实施方式中,根据该网页抓取请求的特征信息,获取目标网页数据包括:

[0010] 根据该特征信息,查询与该特征信息对应的缓存数据;

[0011] 当查询到该缓存数据时,将该缓存数据获取为该目标网页数据。

[0012] 在一种可能实施方式中,根据该网页抓取请求的特征信息,获取目标网页数据包括:

[0013] 根据该特征信息,查询与该特征信息对应的缓存数据;

[0014] 当查询不到该缓存数据时,通过配置服务器获取该目标网页数据,该配置服务器用于为各个特征信息配置对应的网页数据。

[0015] 在一种可能实施方式中,当查询不到该缓存数据时,通过配置服务器获取该目标网页数据之后,该方法还包括:

[0016] 将该特征信息缓存为对象键名,该对象键名用于指示下一次获取该目标网页数据的索引;

[0017] 将该目标网页数据缓存为对象键值,该对象键值用于指示下一次根据该对象键名

查询到的索引内容。

[0018] 在一种可能实施方式中,当查询不到该缓存数据时,通过配置服务器获取该目标网页数据包括:

[0019] 每间隔第一预设时长,根据该特征信息,从该配置服务器获取与该特征信息对应的网页数据。

[0020] 在一种可能实施方式中,根据该网页抓取请求的特征信息,获取目标网页数据包括:

[0021] 根据该特征信息,查询与该特征信息对应的缓存数据;

[0022] 当查询不到该缓存数据,且接收到配置服务器发送的配置失败信息时,通过网站服务器获取该目标网页数据,该配置服务器用于为各个特征信息配置对应的网页数据,该配置失败信息用于指示该配置服务器没有为该特征信息配置对应网页数据。

[0023] 在一种可能实施方式中,该方法还包括:

[0024] 每间隔第二预设时长,清空各个特征信息所对应的缓存数据。

[0025] 在一种可能实施方式中,该特征信息为从该网页抓取请求的目标字段中提取到的搜索引擎标识信息;或,

[0026] 该特征信息为从该网页抓取请求的目标字段中提取到的关键字字符串。

[0027] 在一种可能实施方式中,获取搜索引擎服务器发送的网页抓取请求包括:

[0028] 接收代理服务器发送的目标网页抓取请求,该目标网页抓取请求为该搜索引擎服务器发送的网页抓取请求中,符合预设白名单的网页抓取请求;

[0029] 将该目标网页数据发送至该搜索引擎服务器包括:

[0030] 通过该代理服务器,将该目标网页数据发送至该搜索引擎服务器。

[0031] 根据本公开实施例的第二方面,提供一种网页抓取请求处理装置,该装置包括:

[0032] 第一获取单元,被配置为执行获取搜索引擎服务器发送的网页抓取请求,该网页抓取请求用于请求返回网页数据;

[0033] 第二获取单元,被配置为执行根据该网页抓取请求的特征信息,获取目标网页数据,该目标网页数据与该特征信息对应,该特征信息用于表示发出该网页抓取请求的搜索引擎服务器的数据需求;

[0034] 发送单元,被配置为执行将该目标网页数据发送至该搜索引擎服务器。

[0035] 在一种可能实施方式中,该第二获取单元包括:

[0036] 查询子单元,被配置为执行根据该特征信息,查询与该特征信息对应的缓存数据;

[0037] 获取子单元,被配置为执行当查询到该缓存数据时,将该缓存数据获取为该目标网页数据。

[0038] 在一种可能实施方式中,该第二获取单元包括:

[0039] 查询子单元,被配置为执行根据该特征信息,查询与该特征信息对应的缓存数据;

[0040] 获取子单元,被配置为执行当查询不到该缓存数据时,通过配置服务器获取该目标网页数据,该配置服务器用于为各个特征信息配置对应的网页数据。

[0041] 在一种可能实施方式中,该装置还包括:

[0042] 缓存单元,被配置为执行将该特征信息缓存为对象键名,该对象键名用于指示下一次获取该目标网页数据的索引;

[0043] 该缓存单元,还被配置为执行将该目标网页数据缓存为对象键值,该对象键值用于指示下一次根据该对象键名查询到的索引内容。

[0044] 在一种可能实施方式中,该获取子单元还被配置为执行:

[0045] 每间隔第一预设时长,根据该特征信息,从该配置服务器获取与该特征信息对应的网页数据。

[0046] 在一种可能实施方式中,该第二获取单元包括:

[0047] 查询子单元,被配置为执行根据该特征信息,查询与该特征信息对应的缓存数据;

[0048] 获取子单元,被配置为执行当查询不到该缓存数据,且接收到配置服务器发送的配置失败信息时,通过网站服务器获取该目标网页数据,该配置服务器用于为各个特征信息配置对应的网页数据,该配置失败信息用于指示该配置服务器没有为该特征信息配置对应网页数据。

[0049] 在一种可能实施方式中,其特征信息,该装置还包括:

[0050] 清空单元,被配置为执行每间隔第二预设时长,清空各个特征信息所对应的缓存数据。

[0051] 在一种可能实施方式中,该特征信息为从该网页抓取请求的目标字段中提取到的搜索引擎标识信息;或,

[0052] 该特征信息为从该网页抓取请求的目标字段中提取到的关键字字符串。

[0053] 在一种可能实施方式中,该第一获取单元还被配置为执行:

[0054] 接收代理服务器发送的目标网页抓取请求,该目标网页抓取请求为该搜索引擎服务器发送的网页抓取请求中,符合预设白名单的网页抓取请求;

[0055] 该发送单元还被配置为执行:

[0056] 通过该代理服务器,将该目标网页数据发送至该搜索引擎服务器。

[0057] 根据本公开实施例的第三方面,提供一种服务器,该服务器包括:

[0058] 处理器;

[0059] 用于存储该处理器可执行指令的存储器;

[0060] 其中,该处理器被配置为:

[0061] 获取搜索引擎服务器发送的网页抓取请求,该网页抓取请求用于请求返回网页数据;

[0062] 根据该网页抓取请求的特征信息,获取目标网页数据,该目标网页数据与该特征信息对应,该特征信息用于表示发出该网页抓取请求的搜索引擎服务器的数据需求;

[0063] 将该目标网页数据发送至该搜索引擎服务器。

[0064] 根据本公开实施例的第四方面,提供一种非临时性计算机可读存储介质,当该存储介质中的指令由服务器的处理器执行时,使得服务器能够执行一种网页抓取请求处理方法,该方法包括:

[0065] 获取搜索引擎服务器发送的网页抓取请求,该网页抓取请求用于请求返回网页数据;

[0066] 根据该网页抓取请求的特征信息,获取目标网页数据,该目标网页数据与该特征信息对应,该特征信息用于表示发出该网页抓取请求的搜索引擎服务器的数据需求;

[0067] 将该目标网页数据发送至该搜索引擎服务器。

[0068] 根据本公开实施例的第五方面,提供一种应用程序,包括一条或多条指令,该一条或多条指令可以由服务器的处理器执行时,使得服务器能够执行一种网页抓取请求处理方法,该方法包括:

[0069] 获取搜索引擎服务器发送的网页抓取请求,该网页抓取请求用于请求返回网页数据;

[0070] 根据该网页抓取请求的特征信息,获取目标网页数据,该目标网页数据与该特征信息对应,该特征信息用于表示发出该网页抓取请求的搜索引擎服务器的数据需求;

[0071] 将该目标网页数据发送至该搜索引擎服务器。

[0072] 本公开的实施例提供的技术方案可以包括以下有益效果:

[0073] 通过获取网页抓取请求,从而根据该网页抓取请求的特征信息,来有针对性的为搜索引擎服务器获取目标网页数据,并将该目标网页数据发送至搜索引擎服务器,由于特征信息能够区分各个搜索引擎服务器,因此,基于该特征信息的获取,能够满足不同搜索引擎服务器的不同数据需求,也即能够实现对不同的搜索引擎服务器返回定制化的网页数据。

[0074] 应当理解的是,以上的一般描述和后文的细节描述仅是示例性和解释性的,并不能限制本公开。

附图说明

[0075] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本公开的实施例,并与说明书一起用于解释本公开的原理。

[0076] 图1是根据一示例性实施例示出的一种网页抓取请求处理方法的流程图。

[0077] 图2是根据一示例性实施例示出的一种网页抓取请求处理方法的实施环境示意图。

[0078] 图3是根据一示例性实施例示出的一种网页抓取请求处理方法的交互流程图。

[0079] 图4是根据一示例性实施例示出的一种网页抓取请求处理方法的示意图。

[0080] 图5是根据一示例性实施例示出的一种网页抓取请求处理方法的交互流程图。

[0081] 图6是根据一示例性实施例示出的一种网页抓取请求处理方法的交互流程图。

[0082] 图7是根据一示例性实施例示出的一种网页抓取请求处理装置的逻辑结构框图。

[0083] 图8是根据一示例性实施例示出的一种服务器的逻辑结构框图。

具体实施方式

[0084] 这里将详细地对示例性实施例进行说明,其示例表示在附图中。下面的描述涉及附图时,除非另有表示,不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本公开相一致的所有实施方式。相反,它们仅是与如所附权利要求书中所详述的、本公开的一些方面相一致的装置和方法的例子。

[0085] 图1是根据一示例性实施例示出的一种网页抓取请求处理方法的流程图,如图1所示,网页抓取请求处理方法用于服务器中,包括以下步骤:

[0086] 在步骤101中,获取搜索引擎服务器发送的网页抓取请求,该网页抓取请求用于请求返回网页数据。

[0087] 在步骤102中,根据该网页抓取请求的特征信息,获取目标网页数据,该目标网页数据与该特征信息对应,该特征信息用于表示发出该网页抓取请求的搜索引擎服务器的数据需求。

[0088] 在步骤103中,将该目标网页数据发送至该搜索引擎服务器。

[0089] 本公开实施例提供的方法,通过获取网页抓取请求,从而根据该网页抓取请求的特征信息,来有针对性的为搜索引擎服务器获取目标网页数据,并将该目标网页数据发送至搜索引擎服务器,由于特征信息能够区分各个搜索引擎服务器,因此,基于该特征信息的获取,能够满足不同搜索引擎服务器的不同数据需求,也即能够实现对不同的搜索引擎服务器返回定制化的网页数据。

[0090] 在一种可能实施方式中,根据该网页抓取请求的特征信息,获取目标网页数据包括:

[0091] 根据该特征信息,查询与该特征信息对应的缓存数据;

[0092] 当查询到该缓存数据时,将该缓存数据获取为该目标网页数据。

[0093] 在一种可能实施方式中,根据该网页抓取请求的特征信息,获取目标网页数据包括:

[0094] 根据该特征信息,查询与该特征信息对应的缓存数据;

[0095] 当查询不到该缓存数据时,通过配置服务器获取该目标网页数据,该配置服务器用于为各个特征信息配置对应的网页数据。

[0096] 在一种可能实施方式中,当查询不到该缓存数据时,通过配置服务器获取该目标网页数据之后,该方法还包括:

[0097] 将该特征信息缓存为对象键名,该对象键名用于指示下一次获取该目标网页数据的索引;

[0098] 将该目标网页数据缓存为对象键值,该对象键值用于指示下一次根据该对象键名查询到的索引内容。

[0099] 在一种可能实施方式中,当查询不到该缓存数据时,通过配置服务器获取该目标网页数据包括:

[0100] 每间隔第一预设时长,根据该特征信息,从该配置服务器获取与该特征信息对应的网页数据。

[0101] 在一种可能实施方式中,根据该网页抓取请求的特征信息,获取目标网页数据包括:

[0102] 根据该特征信息,查询与该特征信息对应的缓存数据;

[0103] 当查询不到该缓存数据,且接收到配置服务器发送的配置失败信息时,通过网站服务器获取该目标网页数据,该配置服务器用于为各个特征信息配置对应的网页数据,该配置失败信息用于指示该配置服务器没有为该特征信息配置对应网页数据。

[0104] 在一种可能实施方式中,该方法还包括:

[0105] 每间隔第二预设时长,清空各个特征信息所对应的缓存数据。

[0106] 在一种可能实施方式中,该特征信息为从该网页抓取请求的目标字段中提取到的搜索引擎标识信息;或,

[0107] 该特征信息为从该网页抓取请求的目标字段中提取到的关键字字符串。

[0108] 在一种可能实施方式中,获取搜索引擎服务器发送的网页抓取请求包括:

[0109] 接收代理服务器发送的目标网页抓取请求,该目标网页抓取请求为该搜索引擎服务器发送的网页抓取请求中,符合预设白名单的网页抓取请求;

[0110] 将该目标网页数据发送至该搜索引擎服务器包括:

[0111] 通过该代理服务器,将该目标网页数据发送至该搜索引擎服务器。

[0112] 上述所有可选技术方案,可以采用任意结合形成本公开的可选实施例,在此不再一一赘述。

[0113] 图2是根据一示例性实施例示出的一种网页抓取请求处理方法的实施环境示意图,参见图2,本公开实施例应用于搜索引擎和网站之间的交互场景,搜索引擎服务器201向各个网站发送网页抓取请求,并存储各个网站返回的网页数据,从而可以基于存储的各个网页数据为访问者提供搜索功能。而用于维护网站的服务器包括代理服务器202、SEO (Search Engine Optimization,搜索引擎优化)服务器203、配置服务器204和网站服务器205。

[0114] 其中,该代理服务器202用于对所接收到的数据进行筛选,将目标网页抓取请求发送至SEO服务器203,将除了目标网页抓取请求之外的数据发送至网站服务器205,其中,该目标网页抓取请求为符合预设白名单的网页抓取请求,除了目标网页抓取请求之外的数据可以是不符合预设白名单的网页抓取请求,或者为终端发送的访问请求等,例如,该代理服务器202可以是nginx代理服务器,从而减少了占用的内存,增加了代理服务器202的并发能力。

[0115] 其中,该SEO服务器203用于对网页抓取请求进行处理,分担网站服务器205的数据处理压力;该配置服务器204用于为各个搜索引擎配置对应的网页数据,可选地,还配置服务器204还可以用于对SEO服务器203以及代理服务器202进行管理,该配置服务器204上可以安装有客户端,使得用户能够基于该客户端实现对各个搜索引擎进行配置网页数据的操作,当然,也可以基于该客户端对代理服务器202和SEO服务器203的参数设置进行管理;该网站服务器205用于维护网站的正常运营,为访问该网站的终端提供网页资讯展示服务等。

[0116] 可选地,该代理服务器202、SEO服务器203、配置服务器204以及网站服务器205中,任意两个或以上的服务器也可以位于同一物理实体,本公开实施例不对此进行具体限定。

[0117] 基于上述应用场景,图3是根据一示例性实施例示出的一种网页抓取请求处理方法的交互流程图,图4是根据一示例性实施例示出的一种网页抓取请求处理方法的示意图,如图3和图4所示,该网页抓取请求处理方法用于上述各个服务器的交互过程,包括以下步骤:

[0118] 在步骤301中,当代理服务器202接收到任一数据时,读取该数据的目标字段。

[0119] 在上述步骤301中,该数据可以是各个搜索引擎服务器201发出的网页抓取请求,当然,该数据也可以是终端发出的访问请求,本公开实施例不对该数据的类型进行具体限定,例如,该数据可以是http协议请求的形式,该网页抓取请求可以是爬虫数据。

[0120] 可选地,该目标字段可以是该数据的请求头字段,当该数据为网页抓取请求时,该请求头字段可以携带搜索引擎标识信息、资源类型信息以及语言类型信息等,当该数据为访问请求时,该请求头字段则不携带该搜索引擎标识信息,而是可以携带用于指示终端状

况的终端标识信息等。在上述过程中,以网页抓取请求为http协议请求为例,将该请求头字段的UA (User-Agent, 用户代理) 类型信息获取为该搜索引擎标识信息,该UA类型信息可以是Baiduspider或googlebot等,对于不同的搜索引擎服务器201,可以具有不同的UA类型信息,本公开实施例不对该UA类型信息的内容进行具体限定。

[0121] 在步骤302中,当代理服务器202从该目标字段中读取到该搜索引擎标识信息后,将该数据确定为网页抓取请求,从该目标字段提取资源类型信息和语言类型信息。

[0122] 其中,该资源类型信息用于指示搜索引擎服务器201请求获取的资源类型,例如,当该资源类型信息指示网页资源时,请求头字段中的Accept类型信息可以为text/html,可选地,该资源类型信息还可以指示图片信息、视频信息等,本公开实施例不对该资源类型信息的内容进行具体限定。

[0123] 其中,该语言类型信息用于指示搜索引擎服务器201所支持的语言类型,例如,当该语言类型信息指示简体中文时,请求头字段中的Accept-Language类型信息可以为zh-cn,可选地,该语言类型信息还可以指示繁体中文、英文、日文等,本公开实施例不对该语言类型信息的内容进行具体限定。

[0124] 上述步骤302是一种对网页抓取请求的进行进一步解析的处理方法,而在一些实施例中,当该数据不是网页抓取请求时,还执行采用下述步骤:当代理服务器202没有提取到该搜索引擎标识信息后,将该数据发送至网站服务器205,也即是,将除了网页抓取请求之外的数据发送至网站服务器205,从而通过代理服务器202实现了对海量的数据的初步过滤,减小了网站服务器205的处理压力。

[0125] 在步骤303中,当该搜索引擎标识信息、该资源类型信息和该语言类型信息符合预设白名单时,代理服务器202将该网页抓取请求发送至SEO服务器203。

[0126] 其中,该预设白名单可以是该代理服务器202默认的与网站匹配的信息,也可以是技术人员在该代理服务器202上重新设置的信息,本公开实施例不对该预设白名单的来源进行具体限定,例如,当该网站为中文网站时,技术人员可以在代理服务器202上将该预设白名单设置为:搜索引擎标识信息指示为任一搜索引擎,资源类型信息指示为网页资源,且该语言类型信息指示为简体中文或繁体中文,也即是,不对搜索引擎进行限定,当任一搜索引擎发出的网页抓取请求中,请求返回网页资源,并且该搜索引擎支持的搜索语言为简体中文或繁体中文时,视为该网页抓取请求符合该预设白名单,需要说明的是,只有当网页抓取请求的搜索引擎标识信息、资源类型信息以及语言类型信息均符合该预设白名单时,才将该网页抓取请求确定为符合预设白名单,本公开实施例不对该预设白名单的内容进行具体限定。

[0127] 在一些实施例中,该预设白名单上可以维护有多个搜索引擎以及相应的搜索引擎标识信息、资源类型信息和语言类型信息,在一种可能实施方式中,可以按照预设规则将多个搜索引擎收录到该预设白名单上,例如,该预设规则可以是收录访问流量高的搜索引擎,可以是收录信誉评价好的搜索引擎,还可以是收录与网站服务器205之间存在合作关系的搜索引擎等,该预设规则还可以定期更新,本公开实施例不对该预设规则的内容进行具体限定。

[0128] 上述步骤303是一种代理服务器202对网页抓取请求的处理方式,也即是将目标网页抓取请求发送至服务器,其中,该目标网页抓取请求为各个搜索引擎服务器所发出的网

页抓取请求中,符合预设白名单的网页抓取请求。而可选地,当该搜索引擎标识信息、该语言类型信息和该资源类型信息不符合预设白名单时,还可以执行下述步骤:代理服务器202将不符合预设白名单的网页抓取请求发送至网站服务器205。

[0129] 在一种可能实施方式中,例如,当该预设白名单收录了访问量前10的搜索引擎,代理服务器202接收该预设白名单内的任一搜索引擎的搜索引擎服务器201发出的网页抓取请求时,将该网页抓取请求发送至SEO服务器203,当接收预设白名单外的任一搜索引擎的搜索引擎服务器201发出的网页抓取请求时,将该网页抓取请求发送至网站服务器205。

[0130] 上述步骤301-303中,为代理服务器202对接收到的任一数据的处理方法,将目标网页抓取请求发送至SEO服务器203,将除了目标网页抓取请求之外的数据发送至网站服务器205,通过预设白名单,对所有数据进行筛选,以得到目标网页抓取请求,从而降低了网站服务器205的处理压力,减小了网站服务器205的维护成本,从而保证了网站服务器205对各个终端的访问请求的响应速度。当然,在一些实施例中,还可以不设置预设白名单,也即是,代理服务器202将所有的网页抓取请求发送至SEO服务器203,将除了网页抓取请求之外的数据发送至网站服务器205,从而加快了代理服务器202对数据进行筛选处理的速度,且使得网页数据可以在更多的搜索引擎中被检索到。

[0131] 在步骤304中,当SEO服务器203接收该网页抓取请求后,从该网页抓取请求的目标字段中提取搜索引擎标识信息,将该搜索引擎标识信息获取为特征信息。

[0132] 在上述步骤304中,SEO服务器203可以将该搜索引擎标识信息作为特征信息,使得该特征信息可以指示发出该网页抓取请求的搜索引擎的数据需求,从而针对不同的搜索引擎服务器201,能够返回不同的网页数据,实现定制化的返回网页数据。

[0133] 在一些实施例中,还可以将该目标字段中提取到的关键字字符串作为该特征信息,例如,该关键字字符串可以是该资源类型信息,基于上述示例,将该资源类型信息获取为特征信息,从而可以实现针对请求视频信息的搜索引擎,和请求图片信息的搜索引擎返回不同的网页数据,实现定制化的返回网页数据。可选地,该关键字字符串还可以是该语言类型信息等,从而针对提供不同语言搜索服务的搜索引擎,能够返回不同的网页数据,本公开实施例不对该关键字字符串的类型进行具体限定。

[0134] 在步骤305中,SEO服务器203查询与该特征信息对应的缓存数据。

[0135] 在上述过程中,SEO服务器203可以维护有多个特征信息对应的多个缓存数据,在一些实施例中,SEO服务器203对每一个响应过的目标网页抓取请求,将该目标网页抓取请求的特征信息存储为索引,将发送的网页数据存储为该索引下的索引内容,从而实现了对该多个缓存数据的维护和更新。

[0136] 可选地,可以采用键值的方式来存储该缓存数据,也即是,将该索引作为对象键名,将该发送的网页数据作为对象键值,使得不同特征信息对应于不同的对象键名,从而通过注册表中的key值实现缓存数据的维护,本公开实施例不对该缓存数据的储存形式进行具体限定。例如,当以键值的方式存储该缓存数据时,SEO服务器203通过检索是否存在于该特征信息相匹配的对象键名,当存在匹配的对象键名时,执行下述步骤306,从而实现对该缓存数据的查询操作。

[0137] 在步骤306中,当SEO服务器203查询到该缓存数据时,将该缓存数据获取为该目标网页数据,将该目标网页数据发送至代理服务器202。

[0138] 上述过程中,该目标网页数据与该特征信息对应,SEO服务器203通过将缓存数据获取为目标网页数据,从而避免了将该网页抓取请求发送至网站服务器205,减轻了网站服务器205的数据处理压力,进一步地,由于每个缓存数据可以对应于不同的特征信息,也即是SEO服务器203可以向不同的搜索引擎服务器201发送不同的目标网页数据,例如,对日文搜索引擎的网站,该目标网页数据可以是已有的网页数据加上网站关键词的日文翻译,从而增加了网站在日文搜索引擎上的曝光率,优化了对网页抓取请求的处理方式。

[0139] 在上述步骤307中,SEO服务器203通过该代理服务器202,对不同的搜索引擎服务器201返回不同的目标网页数据,从而形成了流程上的闭环,在一些实施例中,SEO服务器203还可以不通过该代理服务器202,从而直接向该搜索引擎服务器201返回该目标网页数据,简化了网页抓取请求处理流程,节约了网页抓取请求处理的时间。

[0140] 在步骤307中,当代理服务器202接收该目标网页数据,将该目标网页数据发送至搜索引擎服务器201。

[0141] 在步骤308中,SEO服务器203每间隔第二预设时长,清空各个特征信息所对应的缓存数据。

[0142] 其中,该第二预设时长可以是默认的,也可以是用户进行个性化设置的,本公开实施例不对该第二预设时长的获取方式进行具体限定,例如,该第二预设时长可以是1天。在上述过程中,清空缓存数据时,可以仅从各个索引下删除对应的索引内容,而不删除该索引,从而避免了下一次加载缓存数据时,还需要重新添加索引,也即是,基于上述示例,仅将各个对象键名所对应的对象键值置为空,而不删除各个对象键名。

[0143] 本公开实施例提供的方法,通过获取网页抓取请求,从而根据该网页抓取请求的特征信息,来有针对性的为搜索引擎服务器201获取目标网页数据,并将该目标网页数据发送至搜索引擎服务器201,由于特征信息能够区分各个搜索引擎服务器201,因此,基于该特征信息的获取,能够满足不同搜索引擎服务器201的不同数据需求,也即能够实现为不同的搜索引擎服务器201返回定制化的网页数据,进一步地,由于引入了SEO服务器203对网页抓取请求进行处理,避免了海量的频繁发送的网页抓取请求对网站服务器205造成压力,分担了网页抓取请求的处理压力,从而减小了网站服务器205的维护成本,也就能够保证了对各个终端的访问请求的相应速度;此外,在SEO服务器203中维护多个缓存数据,可以快速查找到与特征信息对应的缓存数据,从而将该缓存数据获取为目标网页数据,进一步降低网站服务器205的处理压力;此外,每间隔第二预设时长清空缓存数据,从而能够及时对网页数据进行更新,避免向搜索引擎服务器201返回失效的网页数据。

[0144] 在上述实施例中,SEO服务器203能够查询到与该特征信息对应的缓存数据,从而将该缓存数据获取为目标网页数据,并发送至搜索引擎服务器201,而可选地,当SEO服务器203不能查询到与该特征信息对应的缓存数据时,可以执行下述实施例所执行的步骤,图5是根据一示例性实施例示出的一种网页抓取请求处理方法的交互流程图,参见图5,该实施例包括:

[0145] 在步骤501中,当代理服务器202接收到任一数据时,读取该数据的目标字段。

[0146] 在步骤502中,当代理服务器202从该目标字段中读取到该搜索引擎标识信息后,将该数据确定为网页抓取请求,从该目标字段提取资源类型信息和语言类型信息。

[0147] 在步骤503中,当该搜索引擎标识信息、该资源类型信息和该语言类型信息符合预

设白名单时,代理服务器202将该网页抓取请求发送至SEO服务器203。

[0148] 在步骤504中,当SEO服务器203接收该网页抓取请求后,从该网页抓取请求的目标字段中提取搜索引擎标识信息,将该搜索引擎标识信息获取为特征信息。

[0149] 在步骤505中,SEO服务器203查询与该特征信息对应的缓存数据。

[0150] 在步骤506中,当SEO服务器203查询不到该缓存数据时,向配置服务器204发送网页数据获取请求,该网页数据获取请求至少携带该特征信息,该配置服务器204用于为各个特征信息配置对应的网页数据。

[0151] 上述步骤506是一种SEO服务器203从配置服务器204获取目标网页数据的方法,在一种可能实施方式中,该步骤506还可以采用下述步骤进行替换:SEO服务器203每间隔第一预设时长,向配置服务器204发送该网页数据获取请求。也即是,当SEO服务器203查询不到该缓存数据时,不会立刻向配置服务器204发送该网页数据获取请求,而是等到下一个间隔的第一预设时长时,再向配置服务器204发送该网页数据获取请求,从而能够对该第一预设时长内的网页数据获取请求进行汇总,避免了频繁的数据传输给配置服务器204带来处理压力。其中,该第一预设时长可以是默认的,也可以是用户进行个性化设置的,本公开实施例不对该第一预设时长的获取方式进行具体限定,例如,该第一预设时长可以是1分钟。

[0152] 在步骤507中,当配置服务器204接收该网页数据获取请求,查询与该特征信息匹配的网页数据,当查询到该网页数据,将该网页数据发送至SEO服务器203。

[0153] 在上述过程中,可以是配置服务器204在已经配置好的多个网页数据中搜索该特征信息,从而获取与该特征信息匹配的网页数据,将该网页数据发送至SEO服务器203。

[0154] 在步骤508中,当SEO服务器203接收该网页数据,将该网页数据获取为目标网页数据,将该目标网页数据发送至代理服务器202。

[0155] 在步骤509中,当代理服务器202接收该目标网页数据,将该目标网页数据发送至搜索引擎服务器201。

[0156] 在步骤510中,SEO服务器203将该特征信息缓存为对象键名,该对象键名用于指示下一次获取该目标网页数据的索引。

[0157] 其中,该对象键名为一种缓存数据的索引的可能实施方式,也即是注册表中的key名。

[0158] 在步骤511中,SEO服务器203将该目标网页数据缓存为对象键值,该对象键值用于指示下一次根据该对象键名查询到的索引内容。

[0159] 其中,该对象键值是一种储存索引内容的可能实现方式,也即是注册表中的key值,在上述步骤510-511中,提供了一种添加缓存数据的方法,也即是,当SEO服务器203查询不到目标特征信息对应的缓存数据时,通过配置服务器204获取到与该目标特征信息对应的网页数据,从而将该目标特征信息以及对应的网页数据添加到SEO服务器203的缓存数据中,使得当下一次接收到携带该目标特征信息的网页抓取请求时,可以直接执行上一个实施例所执行的操作,从而优化了网页抓取请求处理的流程。

[0160] 在步骤512中,SEO服务器203每间隔第二预设时长,清空各个特征信息所对应的缓存数据。

[0161] 本公开实施例提供的方法,通过获取网页抓取请求,从而根据该网页抓取请求的特征信息,来有针对性的为搜索引擎服务器201获取目标网页数据,并将该目标网页数据发

送至搜索引擎服务器201,由于特征信息能够区分各个搜索引擎服务器201,因此,基于该特征信息的获取,能够满足不同搜索引擎服务器201的不同数据需求,也即能够实现为不同的搜索引擎服务器201返回定制化的网页数据,进一步地,由于引入了SEO服务器203对网页抓取请求进行处理,避免了海量的频繁发送的网页抓取请求对网站服务器205造成压力,分担了网页抓取请求的处理压力,从而减小了网站服务器205的维护成本,也就能够保证了对各个终端的访问请求的相应速度;在SEO服务器203中维护多个缓存数据,可以快速查找到与特征信息对应的缓存数据,并将该缓存数据作为目标网页数据发出,另外,当查询不到与特征信息对应的缓存数据时,从配置服务器204中获取目标网页数据,从而通过配置服务器204分担网站服务器205的处理压力;进一步地,将从配置服务器204获取到的该目标网页数据添加到缓存数据中,实现了缓存数据的更新,简化了下一次处理该网页抓取请求的流程;此外,每间隔第二预设时长清空缓存数据,从而能够及时对网页数据进行更新,避免向搜索引擎服务器201返回失效的网页数据。

[0162] 在上述实施例中,SEO服务器203能够从配置服务器204中获取到与特征信息对应的目标网页数据,从而通过代理服务器202发送该目标网页数据,从而分担了网站服务器205的数据处理压力,而可选地,如果配置服务器204没有为某个特征信息配置对应的目标网页数据,那么可以执行下述实施例所执行的步骤,图6是根据一示例性实施例示出的一种网页抓取请求处理方法的交互流程图,参见图6,该实施例包括:

[0163] 在步骤601中,当代理服务器202接收到任一数据时,读取该数据的目标字段。

[0164] 在步骤602中,当代理服务器202从该目标字段中读取到该搜索引擎标识信息后,将该数据确定为网页抓取请求,从该目标字段提取资源类型信息和语言类型信息。

[0165] 在步骤603中,当该搜索引擎标识信息、该资源类型信息和该语言类型信息符合预设白名单时,代理服务器202将该网页抓取请求发送至SEO服务器203。

[0166] 在步骤604中,当SEO服务器203接收该网页抓取请求后,从该网页抓取请求的目标字段中提取搜索引擎标识信息,将该搜索引擎标识信息获取为特征信息。

[0167] 在步骤605中,SEO服务器203查询与该特征信息对应的缓存数据。

[0168] 在步骤606中,当SEO服务器203查询不到该缓存数据时,向配置服务器204发送网页数据获取请求,该网页数据获取请求至少携带该特征信息,该配置服务器204用于为各个特征信息配置对应的网页数据。

[0169] 在步骤607中,当配置服务器204接收该网页数据获取请求,查询与该特征信息匹配的网页数据,当查询不到该网页数据,向SEO服务器203发送配置失败信息。

[0170] 在上述步骤607中,如果配置服务器204没有为某个特征信息配置对应的特征信息,那么SEO服务器203还可以执行下述步骤608。

[0171] 在步骤608中,当SEO服务器203接收该配置失败信息,向网站服务器205发送该网页抓取请求。

[0172] 在上述过程中,SEO服务器203通过向网站服务器205发送网页抓取请求,从而将该网站服务器205当前的网页数据获取为该目标网页数据。

[0173] 在步骤609中,当网站服务器205接收该网页抓取请求,向SEO服务器203发送当前的网页数据。

[0174] 在步骤610中,当SEO服务器203接收该当前的网页数据,将该当前的网页数据获取

为目标网页数据,将该目标网页数据发送至代理服务器202。

[0175] 在步骤611中,当代理服务器202接收该目标网页数据,将该目标网页数据发送至搜索引擎服务器201。

[0176] 在步骤612中,SEO服务器203每间隔第二预设时长,清空各个特征信息所对应的缓存数据。

[0177] 本公开实施例提供的方法,通过获取网页抓取请求,从而根据该网页抓取请求的特征信息,来有针对性的为搜索引擎服务器201获取目标网页数据,并将该目标网页数据发送至搜索引擎服务器201,由于特征信息能够区分各个搜索引擎服务器201,因此,基于该特征信息的获取,能够满足不同搜索引擎服务器201的不同数据需求,也即能够实现为不同的搜索引擎服务器201返回定制化的网页数据,进一步地,由于引入了SEO服务器203对网页抓取请求进行处理,避免了海量的频繁发送的网页抓取请求对网站服务器205造成压力,分担了网页抓取请求的处理压力,从而减小了网站服务器205的维护成本,也就能够保证了对各个终端的访问请求的相应速度;在SEO服务器203中维护多个缓存数据,可以快速查找到与特征信息对应的缓存数据,并将该缓存数据作为目标网页数据发出,另外,当查询不到与特征信息对应的缓存数据时,从配置服务器204中获取目标网页数据,从而通过配置服务器204分担网站服务器205的处理压力;进一步地,当配置服务器204不能提供与该特征信息对应的网页数据时,向网站服务器205获取当前的网页数据作为该目标网页数据;此外,每间隔第二预设时长清空缓存数据,从而能够及时对网页数据进行更新,避免向搜索引擎服务器201返回失效的网页数据。

[0178] 图7是根据一示例性实施例示出的一种网页抓取请求处理装置的逻辑结构框图。参照图7,该装置包括第一获取单元701,第二获取单元702和发送单元703。

[0179] 第一获取单元701,被配置为执行获取搜索引擎服务器发送的网页抓取请求,该网页抓取请求用于请求返回网页数据;

[0180] 第二获取单元702,被配置为执行根据该网页抓取请求的特征信息,获取目标网页数据,该目标网页数据与该特征信息对应,该特征信息用于表示发出该网页抓取请求的搜索引擎服务器的数据需求;

[0181] 发送单元703,被配置为执行将该目标网页数据发送至该搜索引擎服务器。

[0182] 本公开实施例提供的装置,通过获取网页抓取请求,再根据该网页抓取请求的特征信息来获取目标网页数据,再向搜索引擎服务器发送该目标网页数据,使得对网页抓取请求进行处理,避免了海量的频繁发送的网页抓取请求对网站服务器造成压力,分担了网页抓取请求的处理压力,从而减小了网站服务器的维护成本,也就能够保证了对各个终端的访问请求的相应速度。

[0183] 在一种可能实施方式中,基于图7的装置组成,该第二获取单元702包括:

[0184] 查询子单元,被配置为执行根据该特征信息,查询与该特征信息对应的缓存数据;

[0185] 获取子单元,被配置为执行当查询到该缓存数据时,将该缓存数据获取为该目标网页数据。

[0186] 在一种可能实施方式中,基于图7的装置组成,该第二获取单元702包括:

[0187] 查询子单元,被配置为执行根据该特征信息,查询与该特征信息对应的缓存数据;

[0188] 获取子单元,被配置为执行当查询不到该缓存数据时,通过配置服务器获取该目

标网页数据,该配置服务器用于为各个特征信息配置对应的网页数据。

[0189] 在一种可能实施方式中,基于图7的装置组成,该装置还包括:

[0190] 缓存单元,被配置为执行将该特征信息缓存为对象键名,该对象键名用于指示下一次获取该目标网页数据的索引;

[0191] 该缓存单元,还被配置为执行将该目标网页数据缓存为对象键值,该对象键值用于指示下一次根据该对象键名查询到的索引内容。

[0192] 在一种可能实施方式中,该获取子单元还被配置为执行:

[0193] 每间隔第一预设时长,根据该特征信息,从该配置服务器获取与该特征信息对应的网页数据。

[0194] 在一种可能实施方式中,基于图7的装置组成,该第二获取单元702包括:

[0195] 查询子单元,被配置为执行根据该特征信息,查询与该特征信息对应的缓存数据;

[0196] 获取子单元,被配置为执行当查询不到该缓存数据,且接收到配置服务器发送的配置失败信息时,通过网站服务器获取该目标网页数据,该配置服务器用于为各个特征信息配置对应的网页数据,该配置失败信息用于指示该配置服务器没有为该特征信息配置对应网页数据。

[0197] 在一种可能实施方式中,其特征在于,基于图7的装置组成,该装置还包括:

[0198] 清空单元,被配置为执行每间隔第二预设时长,清空各个特征信息所对应的缓存数据。

[0199] 在一种可能实施方式中,该特征信息为从该网页抓取请求的目标字段中提取到的搜索引擎标识信息;或,

[0200] 该特征信息为从该网页抓取请求的目标字段中提取到的关键字串。

[0201] 在一种可能实施方式中,该第一获取单元701还被配置为执行:

[0202] 接收代理服务器发送的目标网页抓取请求,该目标网页抓取请求为该搜索引擎服务器发送的网页抓取请求中,符合预设白名单的网页抓取请求;

[0203] 该发送单元703还被配置为执行:

[0204] 通过该代理服务器,将该目标网页数据发送至该搜索引擎服务器。

[0205] 上述所有可选技术方案,可以采用任意结合形成本公开的可选实施例,在此不再一一赘述。

[0206] 关于上述实施例中的装置,其中各个单元执行操作的具体方式已经在有关该网页抓取请求处理方法的实施例中进行了详细描述,此处将不做详细阐述说明。

[0207] 图8是根据一示例性实施例示出的一种服务器的逻辑结构框图,该服务器800可因配置或性能不同而产生比较大的差异,可以包括一个或一个以上处理器(central processing units,CPU)801和一个或一个以上的存储器802,其中,该存储器802中存储有至少一条指令,该至少一条指令由该处理器801加载并执行以实现上述各个示例性实施例提供的网页抓取请求处理方法。当然,该服务器800还可以具有有线或无线网络接口、键盘以及输入输出接口等部件,以便进行输入输出,该服务器800还可以包括其他用于实现设备功能的部件,在此不做赘述。

[0208] 在示例性实施例中,还提供了一种包括指令的非临时性计算机可读存储介质,例如包括指令的存储器802,上述指令可由服务器800的处理器801执行以完成上述网页抓取

请求处理方法,该方法包括:获取搜索引擎服务器发送的网页抓取请求,该网页抓取请求用于请求返回网页数据;根据该网页抓取请求的特征信息,获取目标网页数据,该目标网页数据与该特征信息对应,该特征信息用于表示发出该网页抓取请求的搜索引擎服务器的数据需求;将该目标网页数据发送至该搜索引擎服务器。可选地,上述指令还可以由服务器800的处理器801执行以完成上述示例性实施例中所涉及的其他步骤。例如,所述非临时性计算机可读存储介质可以是ROM、随机存取存储器(RAM)、CD-ROM、磁带、软盘和光数据存储设备等。

[0209] 在示例性实施例中,还提供了一种应用程序,包括一条或多条指令,该一条或多条指令可以由服务器800的处理器801执行,以完成上述网页抓取请求处理方法,该方法包括:获取搜索引擎服务器发送的网页抓取请求,该网页抓取请求用于请求返回网页数据;根据该网页抓取请求的特征信息,获取目标网页数据,该目标网页数据与该特征信息对应,该特征信息用于表示发出该网页抓取请求的搜索引擎服务器的数据需求;将该目标网页数据发送至该搜索引擎服务器。可选地,上述指令还可以由服务器800的处理器801执行以完成上述示例性实施例中所涉及的其他步骤。

[0210] 本领域技术人员在考虑说明书及实践这里公开的发明后,将容易想到本公开的其他实施方案。本申请旨在涵盖本公开的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本公开的一般性原理并包括本公开未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的,本公开的真正范围和精神由下面的权利要求指出。

[0211] 应当理解的是,本公开并不局限于上面已经描述并在附图中示出的精确结构,并且可以在不脱离其范围进行各种修改和改变。本公开的范围仅由所附的权利要求来限制。

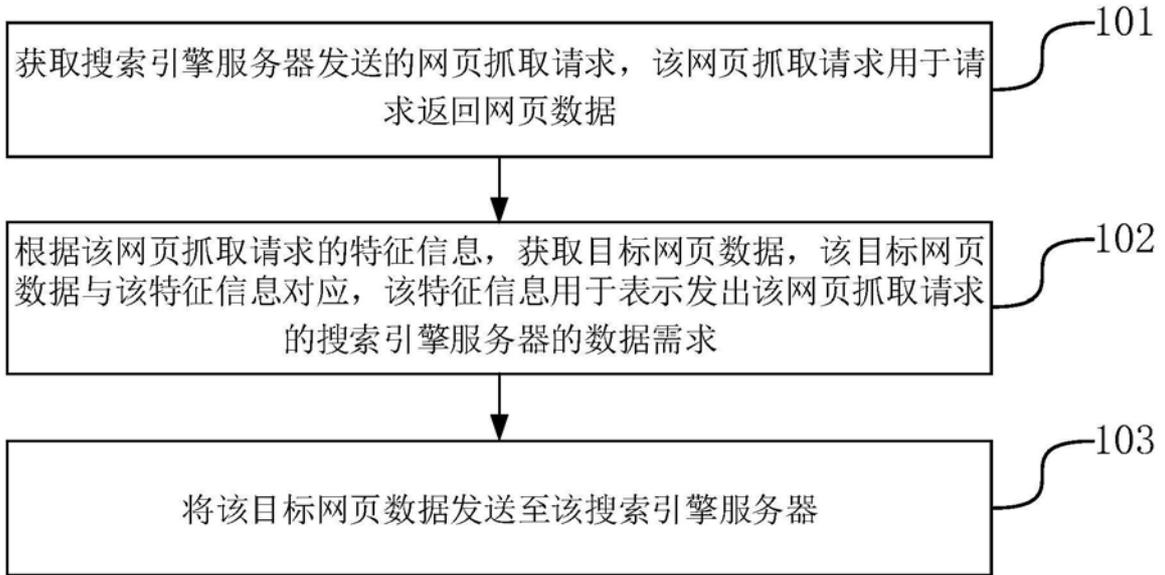


图1

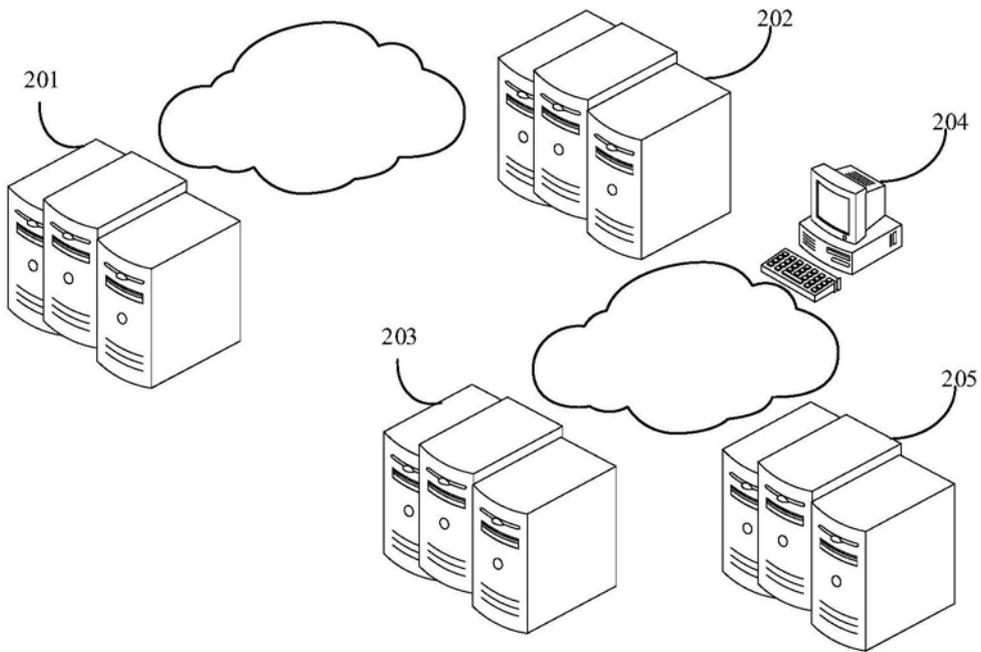


图2

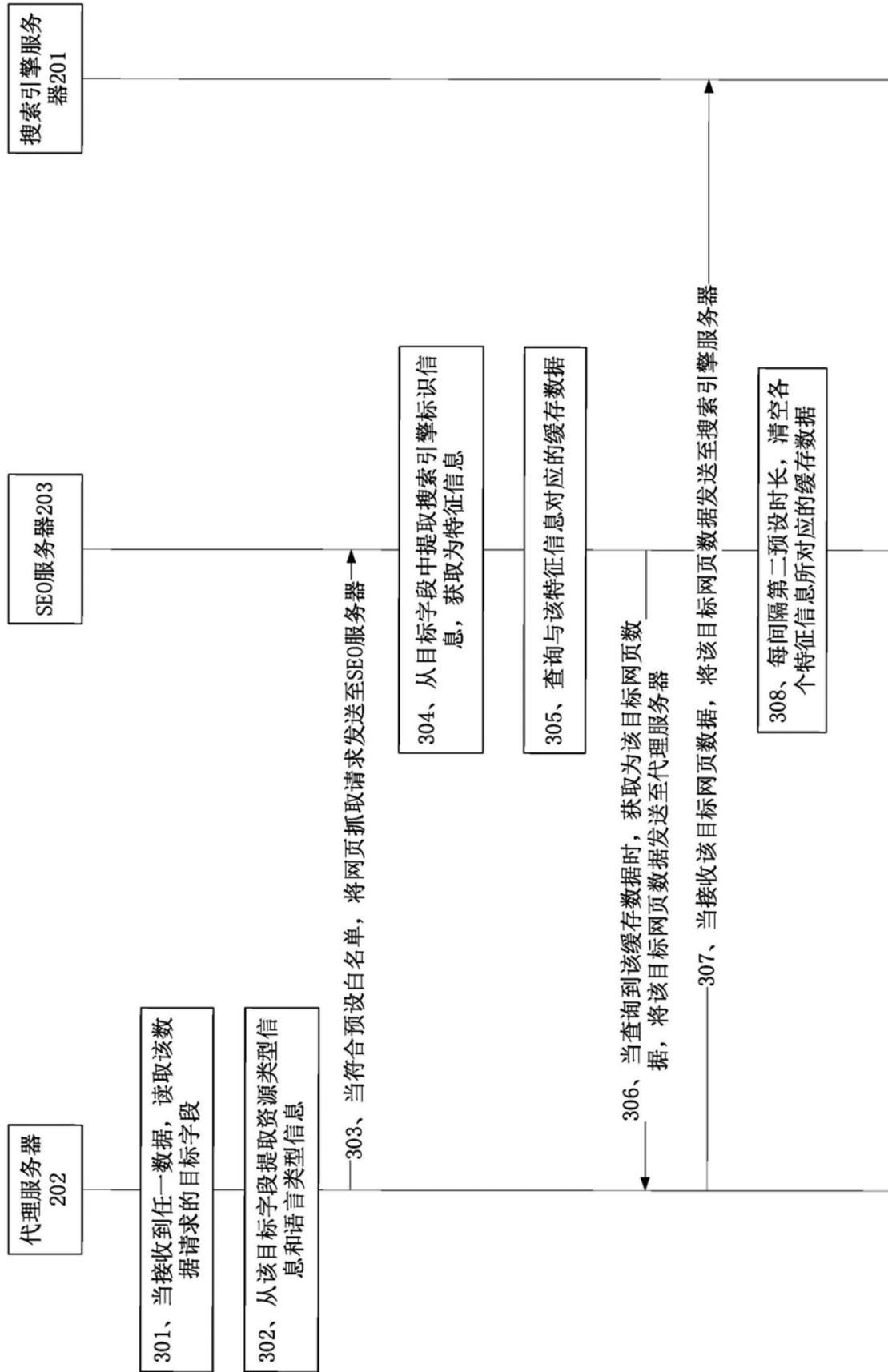


图3

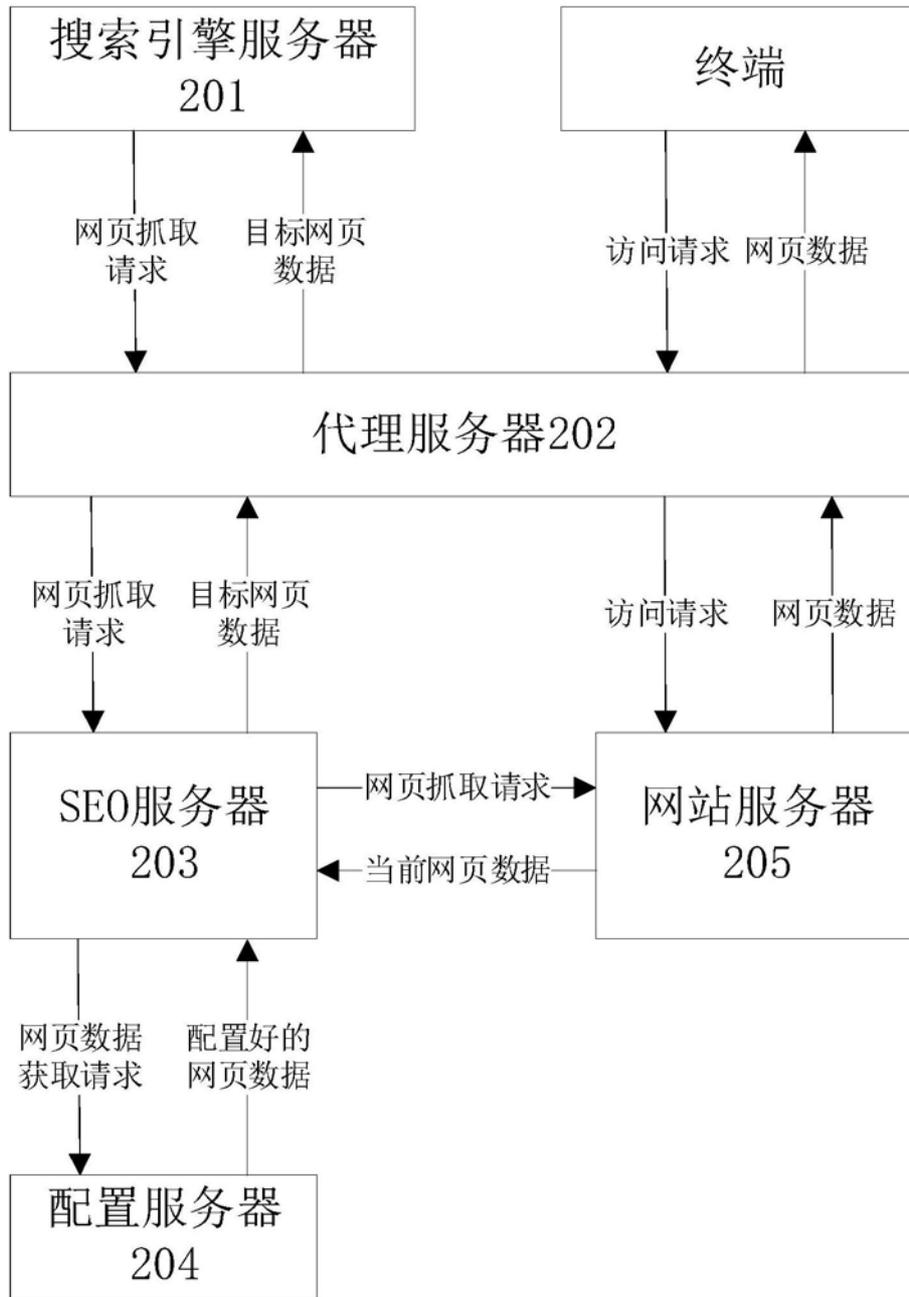


图4

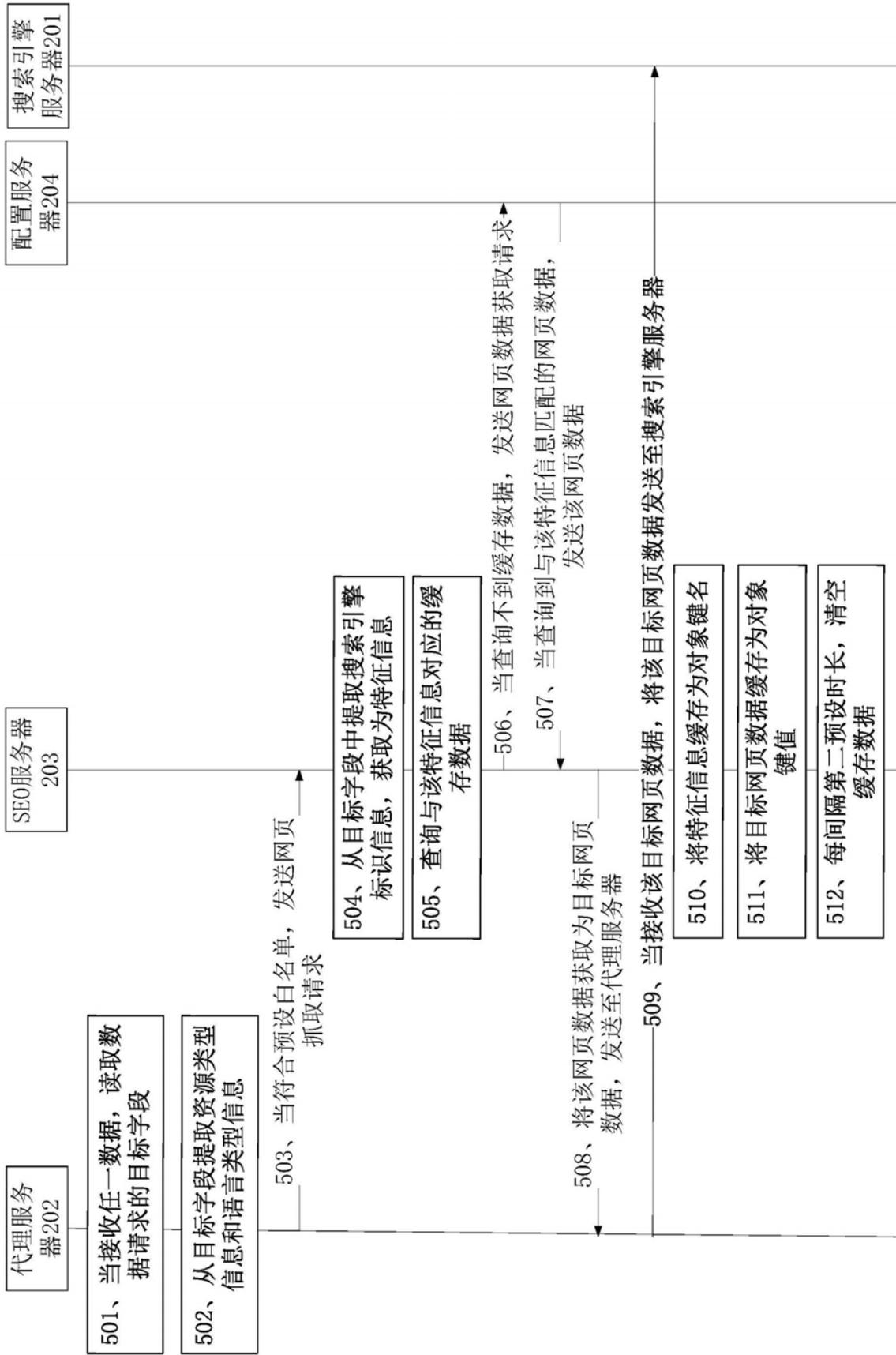


图5

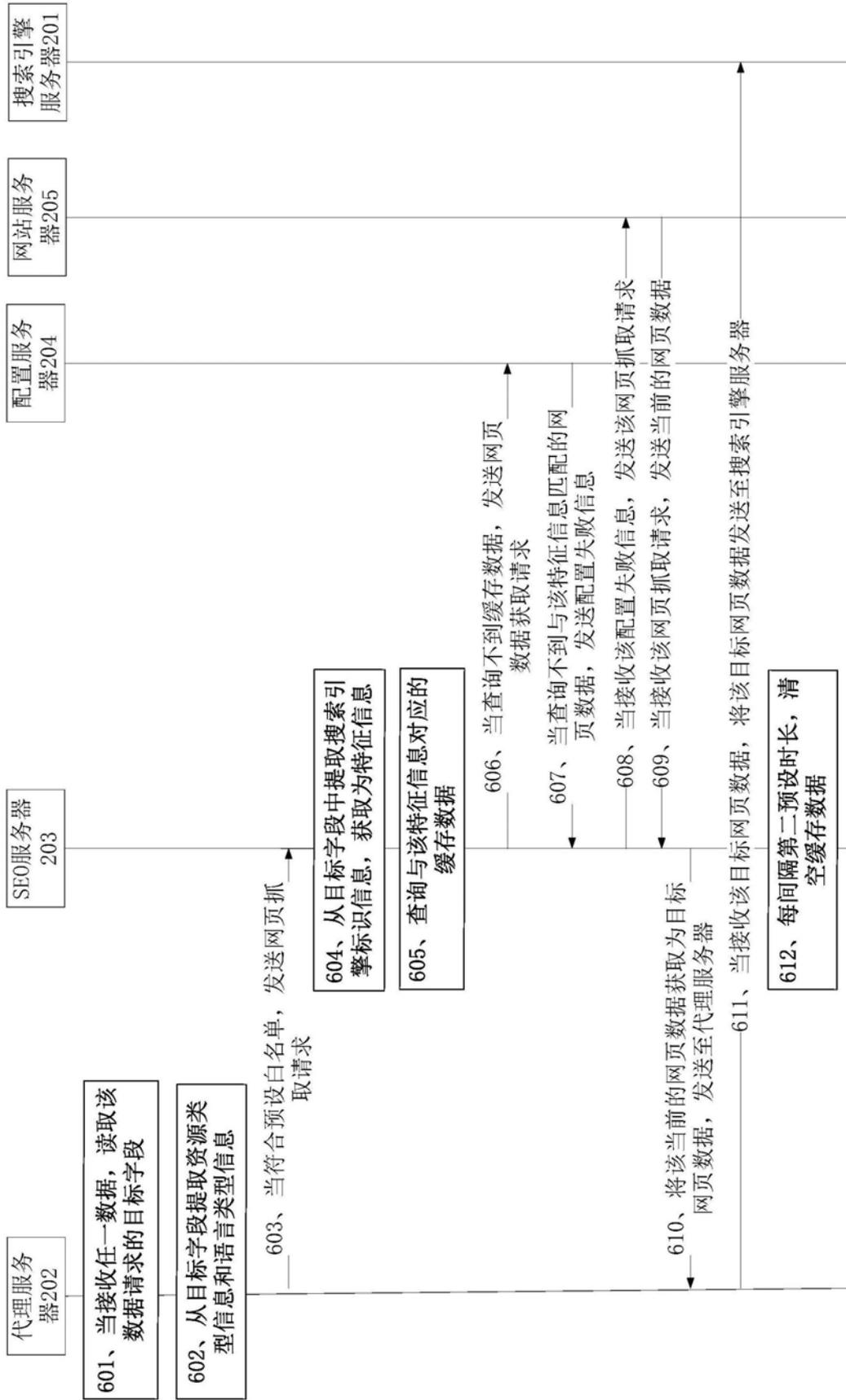


图6

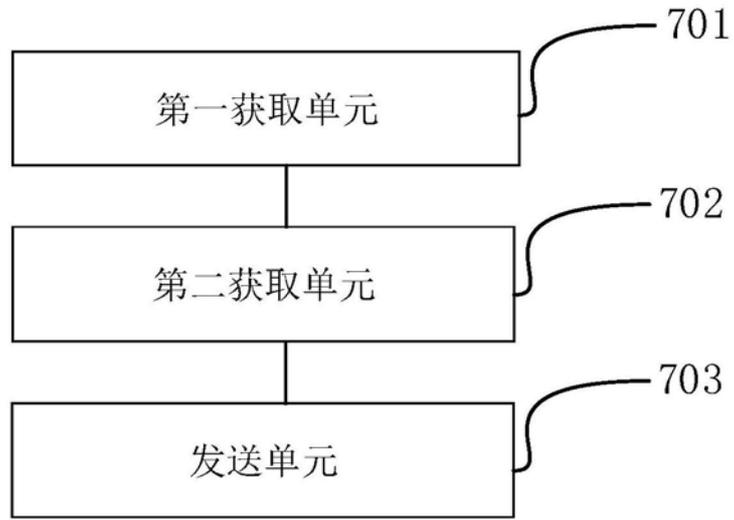


图7

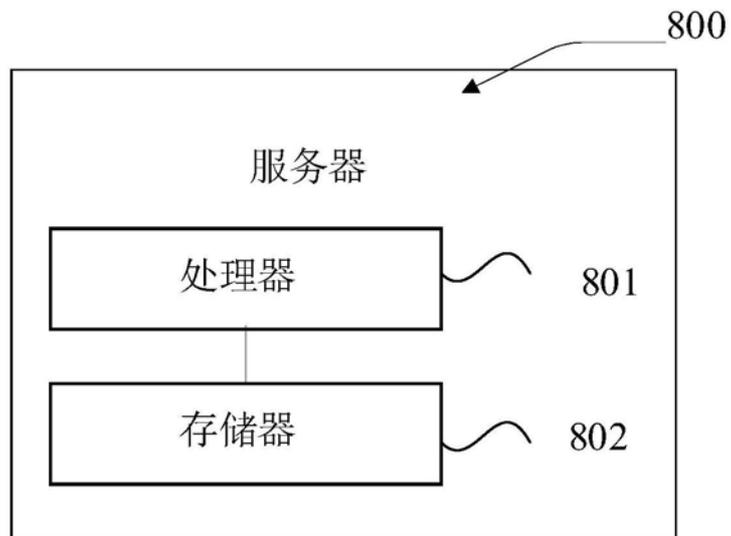


图8