



(12)发明专利

(10)授权公告号 CN 109325149 B

(45)授权公告日 2020.08.11

(21)申请号 201811161600.0

(22)申请日 2018.09.30

(65)同一申请的已公布的文献号

申请公布号 CN 109325149 A

(43)申请公布日 2019.02.12

(73)专利权人 中国银行股份有限公司

地址 100818 北京市西城区复兴门内大街1号

(72)发明人 刘继波 黄新福 张兴华 陈水清 武岳 王硕

(74)专利代理机构 北京中博世达专利商标代理有限公司 11274

代理人 申健

(51)Int.Cl.

G06F 16/81(2019.01)

(56)对比文件

CN 108427676 A,2018.08.21

CN 102650992 A,2012.08.29

CN 107256217 A,2017.10.17

CN 107256217 A,2017.10.17

US 2017011135 A1,2017.01.12

US 8381110 B2,2013.02.19

尹鹏等.一种基于XML的信息搜索技术.《计算机工程与科学》.2017,第29卷(第7期),

审查员 黄倩

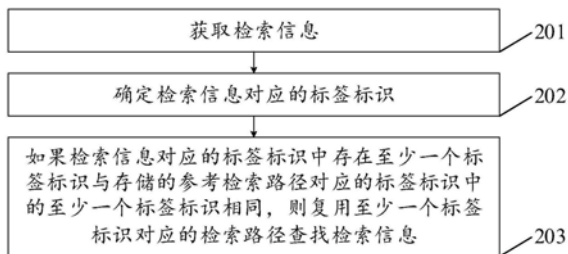
权利要求书2页 说明书8页 附图3页

(54)发明名称

XML报文检索方法及装置

(57)摘要

本申请提供一种XML报文检索方法及装置,涉及计算机领域,能够提高XML报文的检索效率。该方法包括:获取检索信息;确定所述检索信息对应的标签标识,所述标签标识包括标签名称、标签所在的层级和标签下标,所述标签名称用于反映所述检索信息中的内容,所述标签下标用于区分同一个标签对应的下层标签中属于同一个层级的相同标签;如果所述检索信息对应的标签标识中存在至少一个标签标识与存储的参考检索路径对应的标签标识中的至少一个标签标识相同,则复用所述至少一个标签标识对应的检索路径查找所述检索信息,其中,检索路径包括标签标识的节点信息。本申请适用于XML报文的检索过程。



1. 一种可扩展标记语言XML报文检索方法,其特征在于,所述方法包括:

删除存储的历史标签标识以及所述历史标签标识对应的参考检索路径,所述历史标签标识用于表示在开始当前解析XML报文的过程之前存储的标签标识;

获取检索信息;

确定所述检索信息对应的标签标识,所述标签标识包括标签名称、标签所在的层级和标签下标,所述标签名称用于反映所述检索信息中的内容,所述标签下标用于区分同一个标签对应的下层标签中属于同一个层级的相同标签;

逐层检索所述标签标识中每个标签标识;

存储所述每个标签标识及每个标签标识的节点信息,并将所述节点信息确定为参考检索路径;

如果所述检索信息对应的标签标识中存在至少一个标签标识与存储的参考检索路径对应的标签标识中的至少一个标签标识相同,则复用所述至少一个标签标识对应的检索路径查找所述检索信息,其中,检索路径包括标签标识的节点信息。

2. 根据权利要求1所述的方法,其特征在于,所述检索信息的数量为至少两条,在确定所述检索信息对应的标签标识之后,所述方法还包括:

对比每条检索信息的标签标识,将存在至少一个相同标签标识的检索信息划分为一类;

将每一类检索信息中,包括标签标识的数量最多的检索信息对应的标签标识确定为参考标签标识;

按照相似度由高到低的顺序分别对每一类检索信息进行排序,所述相似度用于表示每条检索信息对应的标签标识中,与所述参考标签标识相同的标签标识数量占所述参考标签标识总数的比例;

按照顺序检索每一类检索信息中的每条检索信息。

3. 根据权利要求1所述的方法,其特征在于,在复用所述至少一个标签标识对应的检索路径查找所述检索信息之后,所述方法还包括:

使用所述检索信息对应的检索路径替换所述参考检索路径,以更新参考检索路径。

4. 根据权利要求1所述的方法,其特征在于,在确定所述检索信息对应的标签标识之后,所述方法还包括:

如果所述检索信息对应的标签标识中不存在标签标识与存储的参考检索路径对应的标签标识相同,则逐层查找所述检索信息对应的标签标识中的每个标签标识,得到目标检索路径;

使用所述目标检索路径替换所述参考检索路径,以更新参考检索路径。

5. 一种可扩展标记语言XML报文检索装置,其特征在于,所述装置包括:

获取模块,用于获取检索信息;

确定模块,用于确定所述获取模块获取的所述检索信息对应的标签标识,所述标签标识包括标签名称、标签所在的层级和标签下标,所述标签名称用于反映所述检索信息中的内容,所述标签下标用于区分同一个标签对应的下层标签中属于同一个层级的相同标签;

检索模块,用于当所述确定模块确定的所述检索信息对应的标签标识中存在至少一个标签标识与存储的参考检索路径对应的标签标识中的至少一个标签标识相同时,复用所述

至少一个标签标识对应的检索路径查找所述检索信息,其中,检索路径包括标签标识的节点信息;

删除模块,用于删除存储的历史标签标识以及所述历史标签标识对应的参考检索路径,所述历史标签标识用于表示在开始当前解析XML报文的过程之前存储的标签标识;

所述检索模块,还用于逐层检索所述确定模块确定的所述标签标识中每个标签标识;

存储模块,用于存储所述确定模块确定的所述每个标签标识及所述检索模块检索的每个标签标识的节点信息,并将所述节点信息确定为参考检索路径。

6. 根据权利要求5所述的装置,其特征在于,所述装置还包括:

对比模块,用于对比所述确定模块确定的每条检索信息的标签标识,将存在至少一个相同标签标识的检索信息划分为一类;

所述确定模块,还用于将所述对比模块划分的每一类检索信息中,包括标签标识的数量最多的检索信息对应的标签标识确定为参考标签标识;

排序模块,用于按照相似度由高到低的顺序分别对每一类检索信息进行排序,所述相似度用于表示每条检索信息对应的标签标识中,与所述参考标签标识相同的标签标识数量占所述参考标签标识总数量的比例;

所述检索模块,还用于按照所述排序模块排的顺序检索每一类检索信息中的每条检索信息。

7. 根据权利要求5所述的装置,其特征在于,所述装置还包括:

替换模块,用于使用所述检索模块检索得到的所述检索信息对应的检索路径替换所述参考检索路径,以更新参考检索路径。

8. 根据权利要求5所述的装置,其特征在于,所述装置还包括:

所述检索模块,还用于当所述检索信息对应的标签标识中不存在标签标识与存储的参考检索路径对应的标签标识相同时,逐层查找所述检索信息对应的标签标识中的每个标签标识,得到目标检索路径;

替换模块,用于使用所述检索模块检索得到的所述目标检索路径替换所述参考检索路径,以更新参考检索路径。

XML报文检索方法及装置

技术领域

[0001] 本申请涉及计算机领域,尤其涉及一种可扩展标记语言(Extensible Markup Language,XML)报文检索方法及装置。

背景技术

[0002] XML报文采用分层存储结构,每层都包括标签,该标签对应下一层中的部分或全部内容。示例性的,图1示出了一个XML报文四层结构中的三层,该四层结构中包括根节点的第零层未示出。该XML报文的第二层包括标签1和标签2,标签1和标签2对应的标签名称相同,均为“银行”,当同一个层级中存在多个具有相同标签名称的标签时,通常为标签设置不同的标签下标,如图1中标签1“银行”设置标签下标“1”,标签2“银行”设置标签下标2。需要说明的是,图1中以括号中的数字表示标签下标。

[0003] 如果用户需要从XML报文中检索信息时,如检索信息为银行(1)中存款人(2)的电话(1),终端在接收到上述查询内容后,先在XML报文的第二层中检索银行(1)对应的标签1;在查找到标签1后进入第三层中标签1对应的存储区域,再检索存款人(2)对应的标签4,之后进入标签4对应第三层中的存储区域检索电话(1)。这样逐层进行检索的方式需要终端完成大量的工作,并且,实际中XML报文远不止四层结构,这就使得终端需要完成巨大的工作量,从而耗费大量的时间,导致XML报文的检索效率降低。

发明内容

[0004] 本申请提供一种XML报文检索方法及装置,能够提高XML报文的检索效率。

[0005] 为达到上述目的,本申请采用如下技术方案:

[0006] 第一方面,本申请提供一种XML报文检索方法,该方法可以包括:获取检索信息;确定所述检索信息对应的标签标识,所述标签标识包括标签名称、标签所在的层级和标签下标,所述标签名称用于反映所述检索信息中的内容,所述标签下标用于区分同一个标签对应的下层标签中属于同一个层级的相同标签;如果所述检索信息对应的标签标识中存在至少一个标签标识与存储的参考检索路径对应的标签标识中的至少一个标签标识相同,则复用所述至少一个标签标识对应的检索路径查找所述检索信息,其中,检索路径包括标签标识的节点信息。

[0007] 第二方面,本申请实施例提供一种XML报文检索装置,该装置包括:获取模块,用于获取检索信息;确定模块,用于确定所述获取模块获取的所述检索信息对应的标签标识,所述标签标识包括标签名称、标签所在的层级和标签下标,所述标签名称用于反映所述检索信息中的内容,所述标签下标用于区分同一个标签对应的下层标签中属于同一个层级的相同标签;检索模块,用于当所述确定模块确定的所述检索信息对应的标签标识中存在至少一个标签标识与存储的参考检索路径对应的标签标识中的至少一个标签标识相同时,复用所述至少一个标签标识对应的检索路径查找所述检索信息,其中,检索路径包括标签标识的节点信息。

[0008] 第三方面,本申请提供一种电子设备,包括存储器、通信接口和处理器,所述存储器和所述通信接口耦合至所述处理器,所述存储器用于存储计算机执行代码,所述处理器用于执行所述计算机执行代码控制电子设备执行如第一方面及其各种可能的实现方式所述的XML报文检索方法,所述通信接口用于所述电子设备与外部设备的数据传输。

[0009] 第四方面,本申请提供一种计算机可读存储介质,所述计算机可读存储介质中存储有指令,当所述指令在电子设备上运行时,使得所述电子设备执行如第一方面及其各种可能的实现方式所述的XML报文检索方法。

[0010] 第五方面,本申请提供一种包含指令的计算机程序产品,当所述计算机程序产品在电子设备上运行时,使得所述电子设备执行如第一方面及其各种可能的实现方式所述的XML报文检索方法。

[0011] 本申请提供的XML报文检索方法及装置,与现有技术提供的每次检索时,都逐层对检索信息对应的标签进行检索,由于XML报文的层级复杂,导致检索效率低相比,本申请中存储每次检索得到的标签标识及对应的节点信息作为参考检索路径,在下一次检索时,如果检索信息对应的标签标识中存在与参考检索路径中相同的标签标识,则直接复用相同标签标识的节点信息,也就能够直接查找到该相同的标签标识,省去了再次检索该相同的标签标识的过程,提高了检索效率。

附图说明

[0012] 图1为现有技术提供的XML报文的结构示意图;

[0013] 图2为本申请实施例提供的XML报文检索方法的流程图一;

[0014] 图3为本申请实施例提供的XML报文检索方法的流程图二;

[0015] 图4本申请实施例提供的XML报文检索装置的结构示意图一;

[0016] 图5本申请实施例提供的XML报文检索装置的结构示意图二;

[0017] 图6本申请实施例提供的电子设备的结构示意图一。

具体实施方式

[0018] 下面结合附图对本申请实施例提供的XML报文检索方法及装置进行详细地描述。

[0019] 本申请的描述中所提到的术语“包括”和“具有”以及它们的任何变形,意图在于覆盖不排他的包含。例如包含了一系列步骤或单元的过程、方法、系统、产品或设备没有限于已列出的步骤或单元,而是可选地还包括其他没有列出的步骤或单元,或可选地还包括对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0020] 需要说明的是,本申请实施例中,“示例性的”或者“例如”等词用于表示作例子、例证或说明。本申请实施例中被描述为“示例性的”或者“例如”的任何实施例或设计方案不应被解释为比其它实施例或设计方案更优选或更具优势。确切而言,使用“示例性的”或者“例如”等词旨在以具体方式呈现相关概念。

[0021] 在本申请的描述中,除非另有说明,“多个”的含义是指两个或两个以上。

[0022] 本申请实施例提供一种XML报文检索方法,该方法可以应用于具有XML报文解析功能的电子设备,该电子设备在本申请实施例中具体可以为电脑、服务器等。如图2所示,该方法流程包括:

[0023] 步骤201、获取检索信息。

[0024] 检索信息中包括用户所要查找的具体内容,参阅图1,检索信息具体可以为第一银行(银行的名称),或者第一银行中存款人张三的身份证号码,或者第二银行(银行的名称)中存款人李四的电话等。

[0025] 需要说明的是,检索信息由用户输入或从预存的文件中获取,对于检索信息的来源,在此不做限定。用户输入或预存的检索信息,可能存在顺序比较混乱的问题,例如,用户输入的检索信息可能为存款人张三、电话、第一银行,或者身份证号码、第二银行、存款人李四。由于检索XML报文时需要逐层进行检索,即先从第一层中检索第一银行对应的标签“银行(1)”,再进入第二层中“银行(1)”对应的存储区域检索张三对应的标签“存款人(1)”,再进入“存款人(1)”对应的存储区域检索该存款人(1)的电话(1),可以先确定检索信息中的内容所属的层级,之后按照层级被检索时的由先至后的顺序,对检索信息中包括的内容进行排序,例如,对于检索信息“存款人张三、电话、第一银行”而言,“存款人(1)”属于第二层,“电话(1)”属于第三层,“银行(1)”属于第一层,第一层先被检索,之后第二层再被检索,最后为第三层,那么对上述检索信息进行排序后得到更新后的检索信息“银行(1)、存款人(1)、电话(1)”。上述过程也可以称为对检索信息进行标准化的过程,在标准化之后,电子设备可以按照检索信息中内容的顺序检索XML报文中相应的内容。

[0026] 在检索XML报文时,需要利用应用程序编程接口(Application Programming Interface,API)查找标签。在进行检索时,电子设备绑定标签信息与用于查找标签的API,API逐个从绑定的标签信息中选取标签进行检索,如果API当前检索的层级不存在选取的标签时,则无法查找到该标签,也就意味着,API需要重新选取标签重新进行检索,这样就产生了大量不必要的调用API开销。在本申请实施例中,先将检索信息中的内容进行排序,这样与检索信息中的内容对应的标签也就按照层级被检索时由先至后的顺序进行排序,这样API在每次检索时,都能查找到与绑定的标签信息对应的标签,与现有技术相比,减少了API调用的开销。

[0027] 这样一次性将检索信息中的内容进行排序的过程,省去了电子设备逐个从检索信息中挑选下一个应该检索的内容的时间,提高了检索效率。

[0028] 可选的,由于XML报文每次解析时标签的节点信息均不相同,因此,为了避免其他XML报文的解析过程中存储的标签标识及参考检索路径对本次解析过程产生干扰,在本申请实施例中,在开始解析XML报文时,删除存储的历史标签标识以及历史标签标识对应的参考检索路径。其中,历史标签标识用于表示在开始当前解析XML报文的过程之前存储的标签标识。

[0029] 可选的,对于存储在一个特定存储空间中的历史标签标识和对应的参考检索路径,可以直接清空该特定存储空间中的所有内容;可选的,对于与其他内容共享同一个存储空间的历史标签标识和对应的参考检索路径,则电子设备需要从该存储空间所存储的所有内容中挑选出历史标签标识及对应的参考检索路径进行删除。为了便于电子设备从存储空间所存储的大量内容中确定所需删除的内容,可以为历史标签标识和对应的参考检索路径设置标识,也防止了电子设备所删除的内容错误。

[0030] 可选的,在删除存储的历史标签标识及历史标签对应的参考检索路径之后,在进行本次XML报文解析过程的第一次检索时,由于不存在能够复用的参考检索路径,需要逐层

查找检索信息对应的所有标签标识中每个标签标识。在检索到所需信息之后,存储每个标签标识及每个标签标识的节点信息,将节点信息确定为参考检索路径。这样,在下一次进行检索时,就可以复用本次检索时的检索路径,从而提高检索效率。

[0031] 步骤202、确定检索信息对应的标签标识。

[0032] 其中,标签标识包括标签名称、标签所在的层级和标签下标,标签名称用于反映检索信息中的内容,标签下标用于区分同一个标签对应的下层标签中属于同一个层级的相同标签。

[0033] XML报文中包括多个层级,除第零层外每个层级中包括多个标签,该多个标签中可能存在相同标签。对于不同层级中的相同标签,可以利用层级来区分上述相同标签;对于同一层级相同标签,如果相同标签所属的上层标签不同,则可以根据上层标签来进行区分,如果相同标签所属的上层标签相同,则可以按照标签出现的先后顺序为上述相同标签设置不同的标签下标。示例性的,如图1所示,第一层中的标签银行(1)对应的第二层标签中包括两个相同的标签“存款人”,为了区分两个存款人,分别为其设置标签下标(1)和(2)。这样利用标签名称、层级与标签下标组成的标签标识就可以唯一确定一个标签。也就避免了在检索过程中出现检索到多个相同标签的情况。

[0034] 此外,对于同一个层级中不存在其他相同标签的标签,可以为其分配默认标签下标,如标签下标(1),用以表示该标签只有一个且是该层级中的第一个。示例性的,如图1所示,由于第一层中的标签11“学校(1)”在该层级中仅存在一个且是第一个,所以该标签被分配的标签下标(1)。

[0035] 需要说明的是,电子设备获取的检索信息中的内容与XML报文中设置的标签可能并不相同。例如,电子设备获取的检索信息可能为“第一银行、张三、电话”,其中,第一银行为银行名称,张三为存款人姓名,而在XML报文中,使用诸如“银行”或“存款人”这样概括性的标签来替代具体的名称。如果直接以具体的名称进行检索,则无法查找到对应的内容,因此,需要先确定检索信息对应的标签标识。示例性的,第一银行对应的XML报文中的标签标识为“第一层银行(1)、第二层存款人(1)、第三层电话(1)”。这样电子设备只需查找标签,而不需要再判断查找到的某个标签对应的内容是否为所需内容,方便电子设备识别所要检索的内容。

[0036] 步骤203、如果检索信息对应的标签标识中存在至少一个标签标识与存储的参考检索路径对应的标签标识中的至少一个标签标识相同,则复用至少一个标签标识对应的检索路径查找检索信息。

[0037] 其中,检索路径包括标签标识的节点信息。

[0038] 需要说明的是,在XML报文中,标签相当于进入该标签所属层级的下一层的“门”,在找到“门”之后,就能够进入下一层,并从下一层的标签中继续查找所需内容。“门”所在的位置由节点信息确定,在确定节点信息之后,就可以直接找到该节点信息对应“门”,从而进入“门”对应的下一层。也就意味着,如果在多次检索过程中,所查找的标签对应的标签标识相同或查找标签标识的路径相似,则可以使用前次查找到的标签标识的节点信息,找到该标签,并继续查找本次与前次查找的不同的标签标识,这样就省去了查找相同标签标识的过程,节省了检索过程所需的时间,提高了检索效率。

[0039] 示例性的,参阅图1,如果检索信息包括两条,其一对应的标签标识为“第一层银行

(1)、第二层存款人(1)、第三层家庭地址”,其二对应的标签标识为“第一层银行(1)、第二层存款人(2)、第三层电话(1)”,则当第一次检索完成后,存储第一条检索信息对应的标签标识及其中每个标签标识对应的节点信息。在进行第二次检索时,由于第二条检索信息与第一条检索信息对应的标签标识中存在相同标签标识“第一层银行(1)”,则可以直接复用存储的标签标识“第一层银行(1)”的节点信息,之后再逐一查找与第一条检索信息对应的标签标识中不同的标签标识“第二层存款人(2)”与“第三层电话(1)”,这样就无需重复查找标签标识“第一层银行(1)”,只需要查找剩余的两个标签标识,减少了所需查找的标签标识的个数,提高了检索效率。

[0040] 可选的,电子设备如果获取到多条检索信息,在确定检索信息对应的标签标识之后,还可以将对应的标签标识中,存在至少一个相同标签标识的检索信息划分为一类。之后,将每一类检索信息中,包括标签标识的数量最多的检索信息对应的标签标识确定为参考标签标识,并按照相似度由高到低的顺序对每一类检索信息进行排序。其中,相似度用于表示每条检索信息对应的标签中,与参考标签标识相同的标签标识数量占参考标签标识总数的比例,之后按照顺序检索每一类检索信息。

[0041] 示例性的,参阅图1,如果获取了四条检索信息,第一条检索信息对应的标签标识为“第一层银行(1)、第二层存款人(1)、第三层家庭地址”,第二条检索信息对应的标签标识为“第一层银行(1)、第二层存款人(2)”,第三条检索信息对应的标签标识为“第一层银行(2)”,第四条检索信息对应的标签标识为“第一层银行(1)、第二层存款人(1)、第三层身份证号码(1)”,显然,上述第一条检索信息与第四条检索信息的检索路径最相似,如果按照上述检索信息获取的顺序进行检索的话,则在检索完“第一层银行(2)”后,使用“第一层银行(2)”的节点信息替换之前存储的参考检索路径,在检索第四条检索信息时,无法复用“第一层银行(2)”的节点信息,需要重新检索“第一层银行(1)”和“第二层存款人(1)”,降低了检索效率。

[0042] 为了尽可能保证后一条检索信息检索时能够复用存储的参考检索路径,在进行检索之前,可以先对检索信息进行排序。由于第三条检索信息与其他所有检索信息中都不存在相同标签,因此将第三条检索信息,即“第一层银行(2)”划分为一类;剩余三条检索信息对应的标签标识中包括相同的标签标识“第一层银行(1)”,因此将除第三条检索信息外的三条检索信息划分为一类。之后,将第一条检索信息或第四条检索信息对应的标签标识作为参考标签标识,以第一条检索信息对应的标签标识作为参考标签标识为例,第二条检索信息与第一条检索信息对应标签标识的相似度为 $1/3$,第四条检索信息与第一条检索信息对应标签标识的相似度为 $2/3$,因此,三条检索信息所排的顺序为:第一条检索信息,第四条检索信息,第二条检索信息。在对每一类检索信息进行排序后,可以按照所排顺序分别对每一类检索信息进行检索,这样在检索每一类检索信息,都能够至少复用存储的参考检索路径中一个标签标识的节点信息,提高了检索效率。

[0043] 可选的,由于存储多个参考检索路径会导致XML报文的逻辑变得较为复杂,增加创建XML报文的困难程度,因此,为了尽量减少本申请实施过程中的对于原有XML报文结构的改变,可以在每次检索之后,使用检索信息对应的检索路径替换参考检索路径,从而实现参考检索路径的更新。

[0044] 本申请实施例中存储每次检索得到的标签标识及对应的节点信息作为参考检索

路径,在下一次检索时,如果检索信息对应的标签标识中存在与参考检索路径中相同的标签标识,则直接复用相同标签标识的节点信息,也就能够直接查找到该相同的标签标识,省去了再次检索该相同的标签标识的过程,提高了检索效率。

[0045] 由于本申请实施例中存储了一个检索路径,而检索信息对应的检索路径可以与存储的参考检索路径中没有相同的部分,则需要逐步对检索信息对应的每个标签进行检索以获取用户所需的内容。因此,在如图2所示的实现方式的基础上,还可以实现为如图3所示的实现方式。其中,在执行步骤202之后,还可以执行步骤304和步骤305:

[0046] 步骤304、如果检索信息对应的标签标识中不存在标签标识与存储的参考检索路径对应的标签标识相同,则逐层查找检索信息对应的标签标识中的每个标签标识,得到目标检索路径。

[0047] 示例性的,参阅图1,如果当前存储的参考检索路径为标签标识“第一层银行(1)、第二层存款人(2)、第三层电话(1)”及上述各个标签标识各自对应的节点信息,而当前检索的检索信息为“第一层银行(2)”,与参考检索路径中的任何一个标签标识均不相同,那么就需要在第一层中逐步检索“银行(2)”,得到“第一层银行(2)”对应的节点信息,作为目标检索路径。

[0048] 步骤305、使用目标检索路径替换参考检索路径,以更新参考检索路径。

[0049] 在本次检索完成后,将本次检索得到的目标检索路径替换参考检索路径,以使本次得到的目标检索路径成为新的参考检索路径,保证了参考检索路径的及时更新。

[0050] 可选的,本申请实施例提供的方法也可以用在XML报文的写入过程中。例如,参阅图1,在写入标签标识“第二层存款人(1)”对应的标签后,可以复用该标签标识的节点信息,继续写入对应第三层中的“电话(1)”、“家庭地址”和“身份证号码(1)”。这样,同样能够节省XML报文写入过程中所需耗费的时间。提高XML报文的写入效率。

[0051] 本申请实施例可以根据上述方法示例对电子设备进行功能模块或者功能单元的划分,例如,可以对应各个功能划分各个功能模块或者功能单元,也可以将两个或两个以上的功能集成在一个处理模块中。上述集成的模块既可以采用硬件的形式实现,也可以采用软件功能模块或者功能单元的形式实现。其中,本申请实施例中对模块或者单元的划分是示意性的,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式。

[0052] 如图4所示,本申请实施例提供一种XML报文检索装置400,该装置400包括获取模块401、确定模块402和检索模块403。

[0053] 其中,获取模块401,用于获取检索信息。

[0054] 确定模块402,用于确定获取模块401获取的检索信息对应的标签标识,标签标识包括标签名称、标签所在的层级和标签下标,标签名称用于反映检索信息中的内容,标签下标用于区分同一个标签对应的下层标签中属于同一个层级的相同标签。

[0055] 检索模块403,用于当确定模块402确定的检索信息对应的标签标识中存在至少一个标签标识与存储的参考检索路径对应的标签标识中的至少一个标签标识相同时,复用至少一个标签标识对应的检索路径查找检索信息,其中,检索路径包括标签标识的节点信息。

[0056] 可选的,装置400还包括:

[0057] 删除模块404,用于删除存储的历史标签标识以及历史标签标识对应的参考检索路径,历史标签标识用于表示在开始当前解析XML报文的过程之前存储的标签标识。

- [0058] 检索模块403,还用于逐层检索确定模块402确定的标签标识中每个标签标识。
- [0059] 存储模块405,用于存储确定模块402确定的每个标签标识及检索模块403检索的每个标签标识的节点信息,并将节点信息确定为参考检索路径。
- [0060] 可选的,装置400还包括:
- [0061] 对比模块406,用于对比确定模块402确定的每条检索信息的标签标识,将存在至少一个相同标签标识的检索信息划分为一类。
- [0062] 确定模块402,还用于将对对比模块406划分的每一类检索信息中,包括标签标识的数量最多的检索信息对应的标签标识确定为参考标签标识。
- [0063] 排序模块407,用于按照相似度由高到低的顺序分别对每一类检索信息进行排序,相似度用于表示每条检索信息对应的标签标识中,与参考标签标识相同的标签标识数量占参考标签标识总数量的比例。
- [0064] 检索模块403,还用于按照排序模块407排的顺序检索每一类检索信息中的每条检索信息。
- [0065] 可选的,装置400还包括:
- [0066] 替换模块408,用于使用检索模块403检索得到的检索信息对应的检索路径替换参考检索路径,以更新参考检索路径。
- [0067] 可选的,装置400还包括:
- [0068] 检索模块403,还用于当检索信息对应的标签标识中不存在标签标识与存储的参考检索路径对应的标签标识相同时,逐层查找检索信息对应的标签标识中的每个标签标识,得到目标检索路径。
- [0069] 替换模块408,用于使用检索模块403检索得到的目标检索路径替换参考检索路径,以更新参考检索路径。
- [0070] 本申请提供的XML报文检索装置,与现有技术提供的每次检索时,都逐层对检索信息对应的标签进行检索,由于XML报文的层级复杂,导致检索效率低相比,本申请中存储每次检索得到的标签标识及对应的节点信息作为参考检索路径,在下次检索时,如果检索信息对应的标签标识中存在与参考检索路径中相同的标签标识,则直接复用相同标签标识的节点信息,也就能够直接查找到该相同的标签标识,省去了再次检索该相同的标签标识的过程,提高了检索效率。
- [0071] 如图5所示,本申请实施例提供了XML报文检索装置的另一种可能的结构示意图。该装置500包括:处理单元501、通信单元502和存储单元503。处理单元501用于对XML报文检索装置500的动作进行控制管理,例如,执行上述确定模块402、检索模块403执行的步骤,和/或用于执行本文所描述的技术的其它过程。通信单元502用于支持XML报文检索装置500与其他网络实体的通信,例如,执行上述获取模块401执行的步骤。存储单元503用于存储XML报文检索装置500的程序代码和数据,例如,执行上述存储模块405执行的步骤。
- [0072] 其中,参阅图6所示,上述处理单元501可以是电子设备600中的处理器601或控制器,该处理器601或控制器可以实现或执行结合本申请公开内容所描述的各种示例性的逻辑方框,模块和电路。该处理器601或控制器可以是中央处理器,通用处理器,数字信号处理器,专用集成电路,现场可编程门阵列或者其他可编程逻辑器件、晶体管逻辑器件、硬件部件或者其任意组合。其可以实现或执行结合本申请公开内容所描述的各种示例性的逻辑方

框,模块和电路。处理器601也可以是实现计算功能的组合,例如包含一个或多个微处理器组合,数字信号处理(Digital Signal Processing,DSP)和微处理器的组合等。

[0073] 通信单元502可以是电子设备600中的收发器、收发电路或通信接口602等。

[0074] 存储单元503可以是电子设备600中的存储器603等,该存储可以包括易失性存储器,例如随机存取存储器;该存储器603也可以包括非易失性存储器,例如只读存储器,快闪存储器,硬盘或固态硬盘;该存储器603还可以包括上述种类的存储器的组合。

[0075] 总线604可以是扩展工业标准结构(Extended Industry Standard Architecture,EISA)总线等。总线604可以分为地址总线、数据总线、控制总线等。为便于表示,图6中仅用一条粗线表示,但并不表示仅有一根总线或一种类型的总线。

[0076] 通过以上的实施方式的描述,所属领域的技术人员可以清楚地了解到,为描述的方便和简洁,仅以上述各功能模块的划分进行举例说明,实际应用中,可以根据需要而将上述功能分配由不同的功能模块完成,即将装置的内部结构划分成不同的功能模块,以完成以上描述的全部或者部分功能。上述描述的系统,装置和单元的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0077] 本申请实施例还提供一种计算机可读存储介质,计算机可读存储介质中存储有指令,当电子设备执行该指令时,该电子设备执行上述方法实施例所示的方法流程中电子设备执行的各个步骤。

[0078] 其中,计算机可读存储介质,例如可以是但不限于电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。计算机可读存储介质的更具体的例子(非穷举的列表)包括:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机存取存储器(Random Access Memory,RAM)、只读存储器(Read-Only Memory,ROM)、可擦式可编程只读存储器(Erasable Programmable Read Only Memory,EPR0M)、寄存器、硬盘、光纤、便携式紧凑磁盘只读存储器(Compact Disc Read-Only Memory,CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合、或者本领域熟知的任何其它形式的计算机可读存储介质。一种示例性的存储介质耦合至处理器,从而使处理器能够从该存储介质读取信息,且可向该存储介质写入信息。当然,存储介质也可以是处理器的组成部分。处理器和存储介质可以位于特定用途集成电路(Application Specific Integrated Circuit,ASIC)中。在本申请实施例中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

[0079] 以上所述,仅为本申请的具体实施方式,但本申请的保护范围并不局限于此,任何在本申请揭露的技术范围内的变化或替换,都应涵盖在本申请的保护范围之内。因此,本申请的保护范围应该以权利要求的保护范围为准。

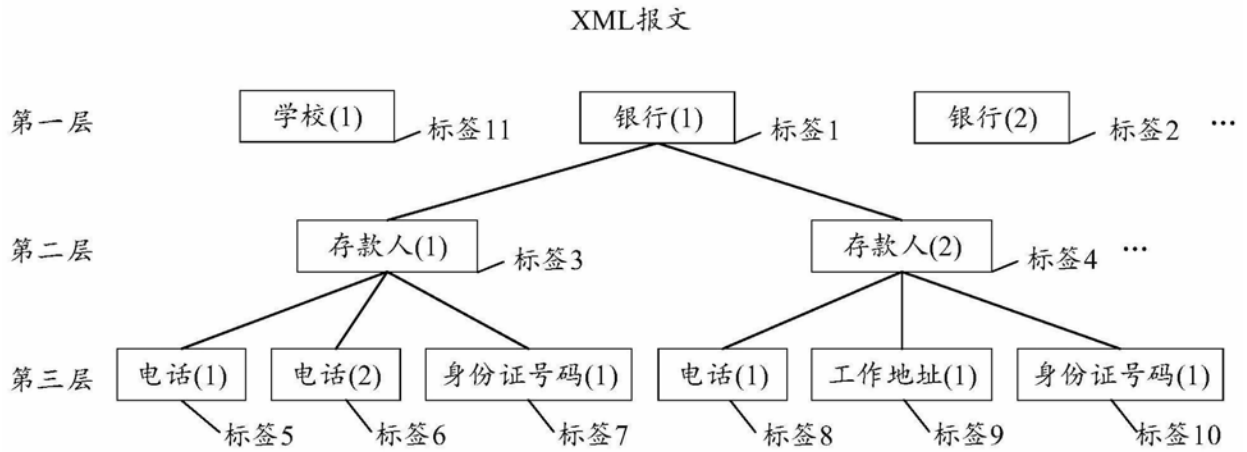


图1

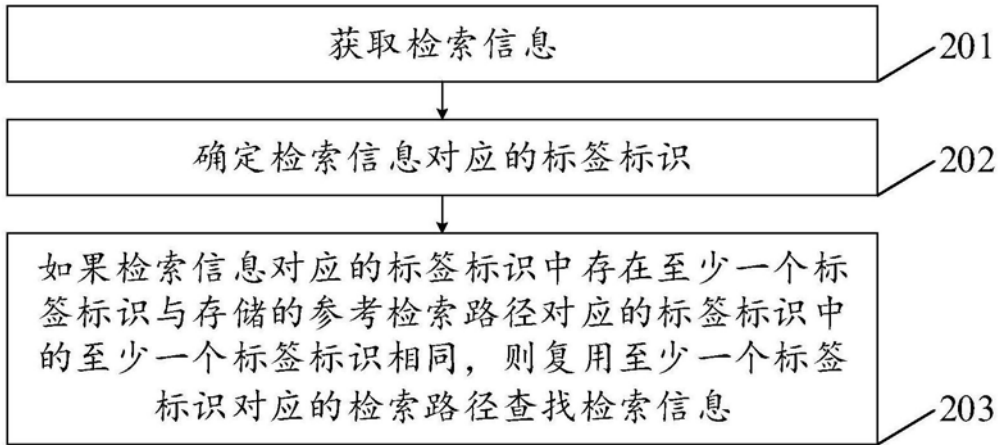


图2

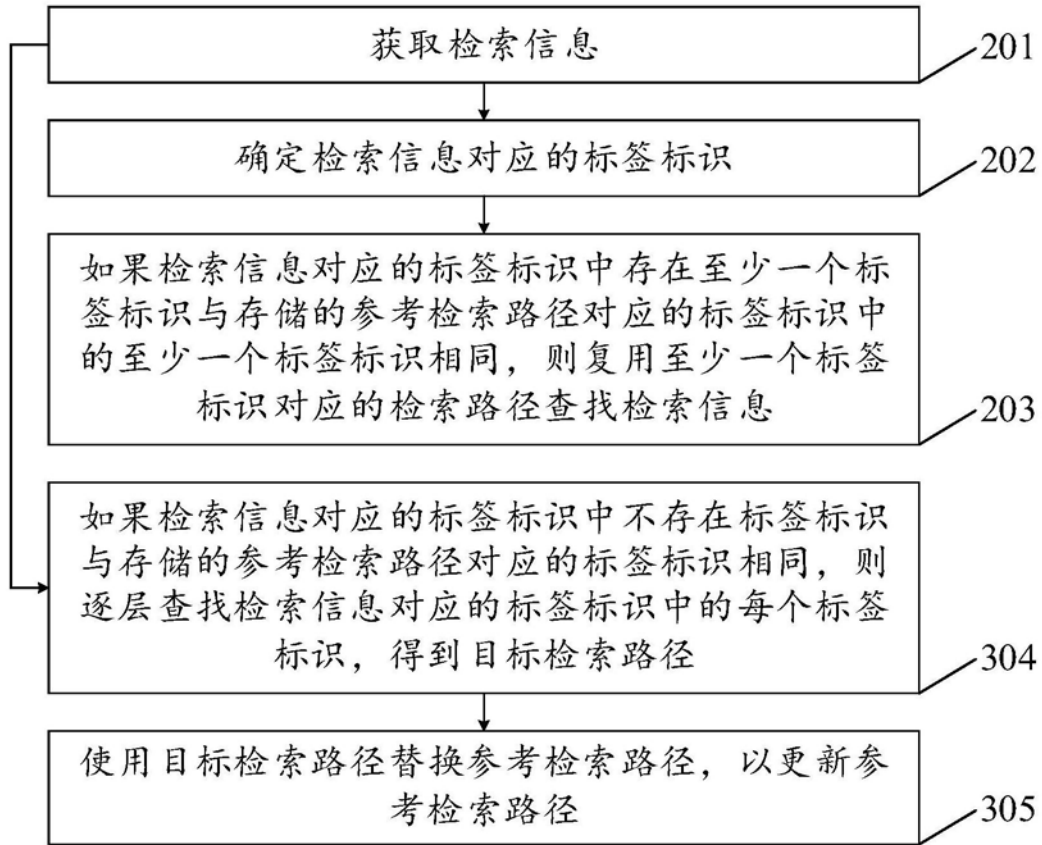


图3

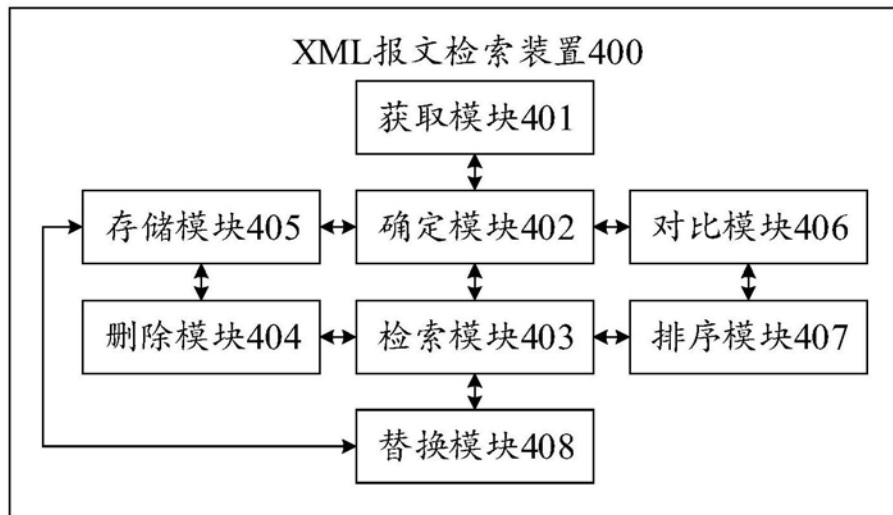


图4

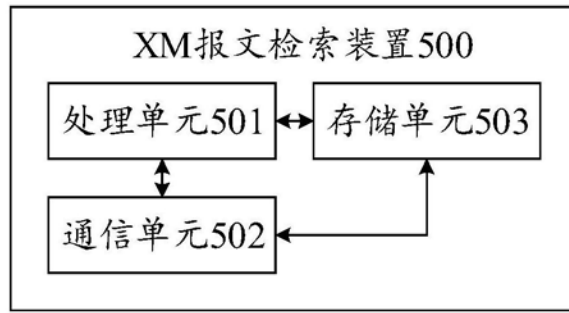


图5

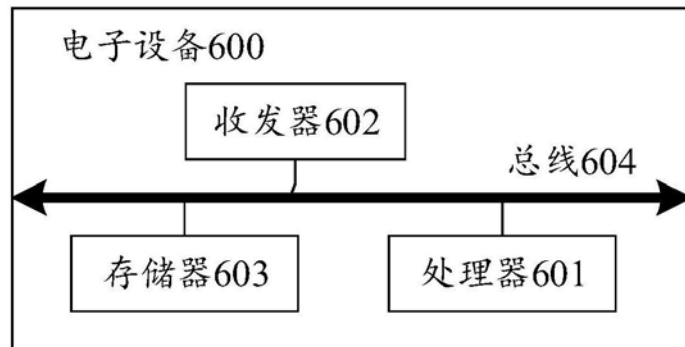


图6