



(12)发明专利

(10)授权公告号 CN 104509041 B

(45)授权公告日 2017.11.07

(21)申请号 201380035862.1

(72)发明人 N·V·卡尔德维尔 D·克劳克斯

(22)申请日 2013.07.03

B·E·查尔兹 S·阿扎姆
F·卡拉贝

(65)同一申请的已公布的文献号

申请公布号 CN 104509041 A

(74)专利代理机构 永新专利商标代理有限公司
72002

(43)申请公布日 2015.04.08

代理人 王英

(30)优先权数据

13/542,102 2012.07.05 US

(51)Int.Cl.

H04L 12/58(2006.01)

(85)PCT国际申请进入国家阶段日

G06Q 10/10(2006.01)

2015.01.05

(86)PCT国际申请的申请数据

PCT/US2013/049393 2013.07.03

(56)对比文件

(87)PCT国际申请的公布数据

W02014/008439 EN 2014.01.09

US 2007233459 A1, 2007.10.04,
CN 101877644 A, 2010.11.03,
US 7016937 B1, 2006.03.21,
CA 2318595 A1, 2001.03.23,
US 2008244750 A1, 2008.10.02,

(73)专利权人 微软技术许可有限责任公司

审查员 蔡红

地址 美国华盛顿州

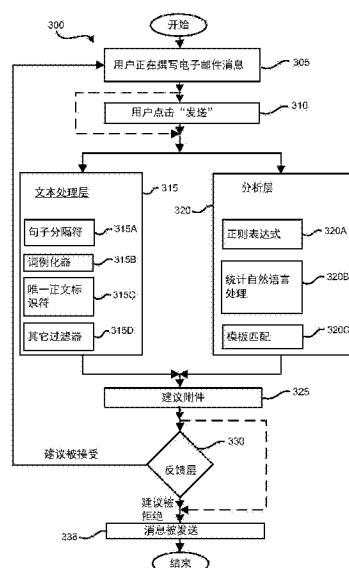
权利要求书3页 说明书11页 附图8页

(54)发明名称

被遗忘的附件的检测方法及装置

(57)摘要

用户正在撰写电子邮件或其它消息。文本处理层准备消息以供分析，并且执行某种初始分析和准备，例如句子分隔符、“词例化器”、“唯一正文”标识和/或可能的其它过滤或分析例程。分析层使用“正则表达式”分析、统计自然语言处理分析和/或模板匹配分析来分析该消息。如果附件被指示，向用户提供添加附件的建议。当用户开始创建消息时，指示该消息应该被发送时，或在它们之间的某个点处时，所述分析可以开始。用户可以提供接受建议或拒绝建议的反馈。



1. 在其中电子消息可以与或可以不与附件一起发送的环境中,一种用于确定电子消息是否建议应该有附件的方法,所述消息具有主题行和正文,所述方法包括:

检查消息以确定所述消息是否已经具有附件;

如果所述消息还没有具有附件,那么检查所述消息的主题行或所述消息的正文以确定所述消息应该具有附件的可能性;以及

如果所述消息的主题行或所述消息的正文指示所述消息应该具有附件,则向用户提供所述消息可能需要附件的建议;

其中,检查所述消息的主题行包括:

搜索所述消息的主题行以查找至少一个预定关键字的存在,

如果至少一个预定关键字存在,那么提供所述建议,以及

如果没有预定的关键字存在,那么检查所述消息的正文,以及

其中,检查所述消息的正文包括:

解析所述消息以标识所述消息内的至少一个句子,

检查所述至少一个句子以查找至少一个预定关键字的存在、至少一个统计上显著的短语或至少一个模板匹配的存在中的至少一个以生成指示所述消息应该具有附件的可能性的分数,以及

如果所述分数大于预定分数,则提供所述建议,以及

其中,还包括使用否定过滤器,所述否定过滤器依赖于其它规则、关键字和/或短语以否决来自不同过滤器的已经确定所述消息应该具有附件的指示。

2. 如权利要求1所述的方法,其特征在于,解析所述消息包括:

搜索所述消息以查找至少一个预定字符的存在;以及

如果至少一个预定字符存在,则将所述消息中的在所述至少一个预定字符之前的所述至少一个句子标识为所述消息中的文本。

3. 如权利要求1所述的方法,其特征在于,所述消息的正文包括当前文本和先前消息的文本,并且还包括在检查所述消息的正文以确定所述消息应该具有附件的可能性之前,将所述当前文本与先前消息的文本区分开来,但仅有所述消息的正文中的所述当前文本被检查以确定所述消息应该具有附件的可能性。

4. 如权利要求1所述的方法,其特征在于,还包括:

查阅预定关键字和预定短语的数据库;以及

在检查所述至少一个句子以查找至少一个预定关键字或至少一个预定短语的存在之后,基于来自用户关于所述建议是被接受还是拒绝的指示来更新所述数据库。

5. 在其中电子消息可以与或可以不与附件一起发送的环境中,所述消息具有主题行和正文,一种计算设备包括:

包含用于操作系统和至少一个文档处理程序的指令的存储器设备;

网络接口单元;

用户输入和显示设备;

通信上耦合到所述存储器设备、所述用户输入和显示设备以及所述网络接口单元的处理器,所述处理器执行所述指令以:

检查消息以确定所述消息是否已经具有附件;

如果所述消息还没有具有附件,那么检查所述消息的主题行或所述消息的正文以确定所述消息应该具有附件的可能性;以及

如果所述消息的主题行或所述消息的正文指示所述消息应该具有附件,则通过所述用户输入和显示设备向用户提供所述消息可能需要附件的建议;

其中,检查所述消息的主题行包括:

搜索所述消息的主题行以查找至少一个预定关键字的存在;

如果至少一个预定关键字存在,那么提供所述建议;

如果没有预定的关键字存在,那么检查所述消息的正文;以及

其中,检查所述消息的正文包括:

解析所述消息以标识所述消息内的至少一个句子;

检查所述至少一个句子以查找至少一个预定关键字的存在、至少一个统计上显著的短语或至少一个模板匹配的存在中的至少一个以生成指示所述消息应该具有附件的可能性的分数;以及

如果所述分数大于预定分数,则提供所述建议;以及

其中,还包括使用否定过滤器,所述否定过滤器依赖于其它规则、关键字和/或短语以否决来自不同过滤器的已经确定所述消息应该具有附件的指示。

6. 如权利要求5所述的计算设备,其特征在于,所述处理器通过进一步执行指令来完成以下操作来解析所述消息:

搜索所述消息以查找至少一个预定字符的存在;

如果至少一个预定字符存在,则将所述消息中的在所述至少一个预定字符之前的所述至少一个句子标识为所述消息中的文本。

7. 如权利要求5所述的计算设备,其特征在于,所述消息的正文包括当前文本和先前消息的文本,并且在检查所述消息的正文以确定所述消息应该具有附件的可能性之前,所述处理器执行所述指令以将所述当前文本与先前消息的文本区分开来,且仅检查所述消息的正文中的所述当前文本以确定所述消息应该具有附件的可能性。

8. 在其中电子消息可以与或可以不与附件一起发送的环境中,一种用于确定电子消息是否建议应该有附件的系统,所述消息具有主题行和正文,所述系统包括:

用于检查消息以确定所述消息是否已经具有附件的装置;

用于如果所述消息还没有具有附件,那么检查所述消息的主题行或所述消息的正文以确定所述消息应该具有附件的可能性的装置;以及

用于如果所述消息的主题行或所述消息的正文指示所述消息应该具有附件,则向用户提供所述消息可能需要附件的建议的装置;

其中,用于检查所述消息的主题行的装置包括:

用于搜索所述消息的主题行以查找至少一个预定关键字的存在装置;

用于如果至少一个预定关键字存在,那么提供所述建议的装置;

用于如果没有预定的关键字存在,那么检查所述消息的正文的装置;以及

其中,用于检查所述消息的正文的装置包括:

用于解析所述消息以标识所述消息内的至少一个句子的装置;

用于检查所述至少一个句子以查找至少一个预定关键字的存在、至少一个统计上显著

的短语或至少一个模板匹配的存在中的至少一个以生成指示所述消息应该具有附件的可能性的分数的装置；以及

用于如果所述分数大于预定分数，则提供所述建议的装置，以及

其中，还包括用于使用否定过滤器的装置，所述否定过滤器依赖于其它规则、关键字和/或短语以否决来自不同过滤器的已经确定所述消息应该具有附件的指示。

9. 如权利要求8所述的系统，其特征在于，还包括：

用于搜索所述消息以查找至少一个预定字符的存在的装置；

用于如果至少一个预定字符存在，则将所述消息中的在所述至少一个预定字符之前的所述至少一个句子标识为所述消息中的文本的装置。

10. 如权利要求8所述的系统，其特征在于，所述消息的正文包括当前文本和先前消息的文本，并且所述系统还包括用于在检查所述消息的正文以确定所述消息应该具有附件的可能性之前将所述当前文本与先前消息的文本区分开来，且仅检查所述消息的正文中的所述当前文本以确定所述消息应该具有附件的可能性的装置。

被遗忘的附件的检测方法及装置

背景技术

[0001] 电子邮件已经成为个人和商业设置两者中常用的通信模式。在许多情况下，用户想要将某些东西附加到电子邮件中，例如文本文档、图片、电影、声音剪辑、演示、扫描等。不幸的是，用户可能太过匆忙或可能太过分心，或可能太过专注于消息，使得用户遗忘了该附件并仅仅点击了“发送”。有时，用户立刻注意到该差错，并重新发送具有附件的电子邮件。其它时候，用户从一个或多个接收者处获得没有附件的通知，并且用户随后重新发送具有附件的电子邮件。原始消息的这种双重发送（一次没有附件，而一次具有附件）以及来自接收者的没有附件的通知浪费了用户和接收者的时间，并且浪费了被用于发送该电子邮件的通信介质上的带宽。

发明内容

[0002] 当用户开始创建消息，指示应当发送消息，或者在这两者之间的某一点时，检查消息（例如但不局限于电子邮件消息）以确定该消息是否应该很可能包括附件，例如文字处理文档、电子表格、照片、扫描、音乐文件、视频文件等。如果该消息已经具有附件，那么可以在没有进一步检查的情况下发送该消息。如果该消息不具有附件，那么检查该消息的主题行和/或消息的正文以确定消息应该具有附件的可能性。

[0003] 如果主题行具有至少一个预定关键字，那么将该消息可能需要附件的一条建议呈现给该用户。如果该用户肯定地回复，那么消息不被发送，并且该用户被允许添加附件和/或编辑该消息。如果该用户否定地回复，则该消息按该用户最初撰写地那样被发送。即使该用户没有响应所述建议，也可以发送该消息。如果建议先前已经被发送给该用户，则在没有用户响应的情况下还是可以发送消息。

[0004] 如果主题行不具有任何预定的关键字，那么，通过解析消息以标识该消息内的至少一个句子来检查该消息的正文。随后对每个句子检查以查找至少一个 预定单词、至少一个统计上显著的短语或至少一个模板匹配。基于这种检查，生成指示该消息应该具有附件的可能性的分数。

[0005] 如果分数大于某个预定值，则将该消息可能需要附件的一条建议呈现给该用户。如果该用户肯定地回复，那么消息不被发送，并且该用户被允许添加附件和/或编辑该消息。如果该用户否定地回复，则该消息按该用户最初撰写地那样被发送。

[0006] 包含预定关键字、预定短语和模板的列表的数据库可以被用于标识消息中指示消息应该具有附件的单词。每次发送消息时，可以更新该数据库以添加、移除、修改或改变关键字、短语和模板的分数。

[0007] 在此所述的主题可被实现为计算机控制装置、计算机进程、计算系统或诸如计算机可读存储介质之类的制品。通过阅读下面的详细描述并审阅相关联的附图，这些及各种其他特征将变得显而易见。

[0008] 本发明内容是对以下详细描述中描述的一些概念的简要和/或简化的介绍。本发明内容并不旨在标识所要求保护的主题的关键特征或必要特征，也不旨在将本发明内容用

来限制所要求保护的主题的范围。此外,所要求保护的主题不限于解决在本公开的任一部分中所提及的任何或所有缺点的实现。

附图说明

- [0009] 图1示出了在其中通过网络服务器、因特网服务供应商 (ISP) 和通信介质中的一个或多个连接多个通信设备的示例性环境。
- [0010] 图2A示出了订阅者通信设备的示例性配置。
- [0011] 图2B示出了网络服务器或ISP的示例性配置。
- [0012] 图3示出了说明用于被遗忘的附件检测的通常的一系列事件的示例性的、简要的、高级流程图。
- [0013] 图4A-4B是示出说明附件检测模块的操作的示例性过程的流程图。
- [0014] 图5示出了用于分析消息的文本以确定该消息是否可能应该具有附连于其的附件的示例性过程。
- [0015] 图6示出能如在此描述执行的设备的示例性计算机体系结构。

具体实施方式

[0016] 下面的详细描述针对用于标识应该具有附加到其的附件的消息或其它文 档的概念和技术。尽管在结合计算机系统上的操作系统和应用程序的执行而执行的程序模块的一般上下文中提出了本文描述的主题,但是本领域技术人员将认识到,其他实现可以结合其他类型的程序模块来执行。一般而言,程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、组件、数据结构和其他类型的结构。此外,本领域技术人员将明白,可以利用其他计算机系统配置来实施本文描述的主题,这些计算机系统配置包括手持式设备、多处理器系统、基于微处理器的或可编程消费电子产品、小型计算机、大型计算机等等。

[0017] 在以下详细描述中,参考了构成详细描述的一部分并作为说明示出了各具体实施方式或示例的附图。现在参考附图(贯穿若干附图中,相同的附图标记表示相同的元素),将提出用于被遗忘的附件检测的计算系统、计算机可读存储介质和计算机实现的方法的各方面。

[0018] 图1示出了在其中通过网络服务器115A、115B、因特网服务供应商 (ISP) 120A、120B 和通信介质125中的一个或多个连接多个订阅者通信设备110A-110H的示例性环境10。网络邮件服务器的一个示例是运行微软Exchange Server®程序的服务器。订阅者通信设备110可以是,例如,个人计算机、智能电话、商业计算机、平板或另一类型的执行一个或多个程序的计算设备。网络服务器115或ISP 120也是一种类型的计算设备。通信介质125可以是,例如,局域网、广域网、因特网、Wi-Fi®连接、或能够将数据(包括电子邮件消息和附件)从一点传递到另一点的一些其它通信介质。

[0019] 两个订阅者设备110可以以各种方式彼此通信,这取决于每个订阅者设备正在其中工作的环境。例如,如果它们(例如,110A、110B)都被连接到网络邮件服务器115A,则它们可以通过网络邮件服务器115A通信,如果它们(例如,110C、110D)都被连接ISP 120A,则它们可以通过ISP 120A通信,它们(例如110E、110G)可以通过网络邮件服务器115B和ISP 120B通信,并且它们(例如110A、110E)可以通过通信介质125(通常通过网络邮件服务器115

和ISP 120)通信。通常有4种方法(或协议)被用于处理消息：“交换-启用(Exchange-enabled)”(微软Exchange Server®程序)、IMAP(因特网消息访问协议)、POP(邮局协议)和WebMail。虽然在这些协议中存在显著的差异,但出于被遗忘的附件检测目的,这些协议之间的主要差异在于如何创建电子邮件消息。在交换-启用或POP系统中,在用户计算机上存在电子邮件客户端程序,藉由该程序用户即使离线也可以创建电子邮件消息。然而,在IMAP和WebMail系统中,用户藉由其可以创建消息的程序在服务器上,因此,为了创建电子邮件消息,用户通常通过使用web浏览器程序来连接到邮件服务器115。这些通信方法仅仅是示例性的,并非限制性的。

[0020] 图2A示出了订阅者通信设备110(例如订阅者通信设备110A)的示例性配置。设备110具有存储在其上的一个或多个程序或模块,例如操作系统205、藉由其用户可以创建消息的文档处理程序210(例如但不局限于电子邮件程序或因特网浏览器程序,例如在网络服务器115是IMAP或webmail服务器的情况下),用于执行在此所述的操作(包括检测规则数据库215A和/或包括到检测规则数据库215A的访问)的附件检测模块215,以及一个或多个例如可以被用于创建附件等的其它程序220,例如但不局限于,因特网浏览器、病毒/恶意软件检测和防范程序、文字处理程序或电子表格程序。

[0021] 图2B示出了网络服务器115的示例性配置。ISP 120还可以使用图2所示的配置。在一个实施例中,服务器115具有在其上的一个或多个程序,例如操作系统230、诸如电子邮件服务器的消息处理程序235、通过其用户可以创建消息的程序、或其它消息-创建和/或处理主机程序、用于执行在此所述的操作(包括检测规则数据库240A和/或包括到检测规则数据库240A的访问)的附件检测模块240,以及一个或多个例如但不局限于网络安全、服务器负载分派等的其它程序245。

[0022] 简要地回顾图1,在其中也示出中央数据收集系统130。较佳地,但并非必要地,运行附件检测模块215、240的设备110、115、120将关于用户对附件可能被指示的建议的响应和关于当附件已经存在在消息中时已经被使用的文本的信息发送给系统130。这提供了对在此所述的检测数据库和规则的更新和微调。在系统130和设备之间的连接(作为示例而非限制)可以通过因特网或任意其它便利的通信介质。

[0023] 图3示出了说明用于被遗忘的附件检测的通常的一系列事件300的示例性的、简要的、高级流程图。示出了几个实施例。在一个实施例中,用户编写305消息,例如但不局限于,电子邮件,并且用户点击“发送”310。随后,可使用文本处理层315、分析层320或这两者来分析消息。这两个层被并行示出,但在替换实施例中,它们可以是以任意一个为先的串联形式。在又一替换实施例中,仅使用这些层中的一个。

[0024] 可以具有几个模块的文本处理层315准备用于分析的消息并执行某种有限初始分析,例如:句子分隔符315A、“词例化器(tokenizer)”315B、“唯一正文”标识符315C以及可能的其它过滤或分析例程315D。

[0025] 分析层320使用“正则表达式”分析320A、统计自然语言处理分析320B、模板匹配分析320C中的任一者、上述任意两个或上述的全部来分析消息。

[0026] 如果文本处理层315、分析层320或这两者指示电子邮件应该具有附加其上的附件,但却没有附加,则向用户提供325添加附件的建议。如果来自用户的反馈330是建议被接受,那么将用户返回到电子邮件处理以便该用户可以将期望的附件附加到电子邮件消息。

如果来自用户的反馈330是添加附件的建议被拒绝,那么,将消息如用户最初撰写地那样被发送335。

[0027] 在一个替换实施例中,由从操作305到操作315和320的虚线所示(所述虚线绕过了操作310),当用户指示想创建新的消息的期望时、当用户指示对消息回复的期望时、当用户开始编写305消息时,当用户已经编写了预定数目的字符、段落或行时、当用户第一次保存消息时和/或当用户打开所保存的消息的草稿时等等,文本处理层315和/或分析层320开始。文本处理层315和/或分析层320还可以当用户将文档附加到消息时开始,并且该信息可以被用于更新和/或微调检测规则数据库。

[0028] 如果层315或层320中的任一者或这两者指示电子邮件应该具有附加其上的附件,但却没有附加,则向用户提供325添加附件的建议。该建议可以是一种要求来自用户的响应330的建议,如前所述。然而,在替换实施例中,建议325可以仅仅是通过含信息的消息传递给用户,诸如但不限于用户可以只是关闭它或让其保持打开以作为提醒的弹出框,或者在1或2分钟后就将淡出的弹出框。类似地,相对于图4A和4B,分析可以在406或甚至在412处开始并且建议416和用户响应418将如在325和330中那样被处理。这个替换实施例也由图4A和4B中的虚线所指示。类似地,相对于图5,分析可以在502或甚至504、506、508、510、511或512处开始,并且建议416将如在325和330中 那样被处理。

[0029] 如果用户在对建议325的响应330中指示了“发送”,那么,由于在先前已经提供给用户可指示附件的建议,该消息可以简单地在没有进一步通知用户的情况下被发送。或者,即使已经提供给用户可指示附件的建议,当用户按下“发送”时如果分析还是指示应该有附件,那么可以将任意弹出框带入前台,或可以再次将建议发送给用户,该建议要求或者不要求来自用户的响应。

[0030] 在一个实施例中,消息被保持直到用户提供了关于对附件的建议是被接受还是被拒绝的反馈330。在替换实施例中,如果用户没有在预定的时间量内对应该有附件的建议做出响应,那么如在从操作325到操作335(绕过了操作330)的虚线指示的那样,即使没有用户的反馈,该消息也可被发送。

[0031] 图4A-4B是说明附件检测模块的操作的示例性过程400的流程图。不一定按任何特定次序来呈现本文公开的过程400和其它过程的操作,并且以替代的次序来执行某些或全部操作是可能的并被构想。为了易于描述和说明,按所示次序来呈现各操作。可以添加操作、省略操作、同时执行操作或以另一种次序执行操作,而不脱离所附权利要求书的范围。

[0032] 还有,所示出的过程可以在任意时间结束,并且不需要被完整执行。所述过程的部分或全部操作和/或基本上等价的操作可以通过执行计算机存储介质上所包括的计算机可读指令来执行,如本文所定义的。如在说明书和权利要求书中使用的术语“计算机可读指令”及其变型,在本文是用来广泛地包括例程、应用、应用模块、程序模块、程序、组件、数据结构、算法等等。计算机可读指令可以在各种系统配置上实现,包括单处理器或多处理器系统、小型计算机、大型计算机、个人计算机、手持式计算设备、基于微处理器的可编程消费电子产品、其组合等等。

[0033] 这样,此处所述的逻辑操作被实现为(1)一系列计算机实现的动作或运行于计算系统上的程序模块,和/或(2)计算系统内的互连的机器逻辑电路或电路模块。该实现是取决于计算系统的性能及其他要求的选择问题。因此,此处描述的逻辑操作被不同地称为状

态、操作、结构设备、动作或模块。这些操作、结构设备、动作和模块可以用软件、固件、专用数字逻辑及其任何组合来实现。

[0034] 附件检测模块215、240可以在通信设备110中,可以在网络服务器115 中,可以在ISP 120中,或其任意组合。例如,设备110可以将模块215置于其内。或者,如果多个设备110被连接到(例如公司中的)网络服务器115,那么可能期望除了在设备110上之外或者作为其替代,仅将附件检测模块240置于服务器115上。类似地,可以期望ISP 120具有附件检测模块240以减少由没有包括所指示的附件的第一电子邮件消息、一个或多个指示没有附件的回复电子邮件以及包括附件的另外的电子邮件所引起的话务和对ISP资源的使用。为了讨论简单明了,将讨论附件检测模块215的操作,应该理解附件检测模块240执行相同的功能。附件检测模块215、240较佳地是另一程序(例如电子邮件程序210、235)内的模块或插件,但是,附件检测模块也可以被作为单独的程序实现。

[0035] 过程400在操作402处开始,其中用户使用文档处理程序210或使用与网络服务器或ISP中的消息处理程序235通信的web浏览器程序来撰写文档,例如但不局限于电子邮件消息。一旦用户完成,用户将例如通过点击按键指示该用户希望该文档被发送,例如但不局限于,点击“发送”按键、点击下拉菜单上的“发送”命令、触摸触摸屏上的“发送”图标或某种其它动作。将注意到,在一些情况下,用户通过意外地点击、选择或触摸“发送”按键、命令或图标(例如当用户实际上意指“保存”消息的草稿时)而无意地发出“发送”指令的信号。

[0036] 在操作404,由文档处理程序240或由附件检测模块215来检测“发送”指令信号。作为对其的响应,附件检测模块215检查文档以确定在操作406是否已经有东西被附加到文档。例如,如果该文档是电子邮件,那么,可以附加图片、视频、声音剪辑、文本文档、电子表格文档、剪贴画、“v-卡(v-Card)”等。在替换实施例中,如由操作402和406之间(绕过了操作404)的虚线所指示那样,并如先前所讨论的,对于文档处理程序210开始分析406来说并不要求“发送”指令。为了方便下面的讨论,例如以区别由文档处理程序210或消息处理程序235产生的文档与可能附加于其的文档,由文档处理程序210或消息处理程序235所产生的文档将被称为“消息”,而(例如可能由文字处理、电子表格或其它程序220产生的)任意附加的文档将被称为“文档”。

[0037] 如果有附加到消息的文档,那么在操作408处附件检测模块215将较佳地(但非必要地)更新其检测规则数据库,并且在操作410处将允许消息和附加的文档如同用户最初撰写地那样被发送,或者将允许消息和附加的文档被传递到另一个程序上,例如拼写检查器程序、元数据移除程序或病毒检查程序,以供在被发送之前进行进一步处理。

[0038] 如果需要,即使当已有附件,可以通过运行如下所述的分析操作412、424和/或426来更新检测规则数据库。这提供了当前检测规则数据库是令人满意的附加保证,和/或对数据库微调以标识即使当当前的检测规则数据库可能没有建议应该有附件时也可能期望附件的情形。

[0039] 如果文档没有被附加到消息,那么在操作412处附件检测模块215将检查消息的“主题”行(有时称为“标题”)以确定是否应该有附加的文档。这可以通过搜索主题行以查找一个或多个关键字的存在来完成。一些关键字的示例是,但不局限于,附加、被附加、附寄、被附寄、附件、被附属、附录、文档、演示、文章、图形等。也可利用其他关键字。

[0040] 如果,在操作414,确定主题行指示了应该有附件,那么,在操作416处模块215向用

户显示对话框或其它类型的用户界面元素,由其向用户询问是否应该有对正在撰写的消息的附件。对话框可以是例如“应该有附件吗?”,或“该消息建议附件应该被包括。”,或其它这样的对话。对话框可以具有按键、图标、下拉菜单或其它类型的用户界面元素,这允许用户进行响应,诸如“是的,有附件。”、“是的,我想要添加附件。”、“不,没有附件。”、“总之发送。”、“不,并且不要再提问。”。

[0041] 如果在操作418处的用户响应是否定的(例如,没有附件、总之发送、不再提问),那么在操作408处模块215将较佳地(但非必要地)更新其检测规则数据库,并将允许消息如同用户最初撰写地那样被发送,或将允许消息被传递到另一个程序上,例如拼写检查器程序、元数据移除程序或病毒检查程序,以供在被发送之前进行进一步处理。这在操作410处发生。

[0042] 如果在操作418处的用户响应是肯定的(例如该用户想要将东西附加到消息),那么,在操作420处模块215将较佳地(但非必要地)更新其检测规则数据库,并将用户返回到文档准备程序210,在那用户可以附加文档、编辑消息等。

[0043] 在操作418,模块215还可以呈现对话框或其它类型的用户界面元素以允许该用户直接选择在步骤422处要附加的文档,而不是将用户返回到文档准备程序210。一旦文档被附加,那么在操作408处模块215将较佳地(但非必要地)更新其检测规则数据库,并且将允许消息和附加的文档被发送,或者将允许消息和附加的文档被传递到另一个程序上,例如拼写检查器程序、元数据移除程序或病毒检查程序,以供在被发送之前进行进一步处理。

[0044] 在替换实施例中,如先前所讨论的,不要求用户响应418,如在操作416和408之间的虚线所示,且如上所讨论的。

[0045] 如果在判定操作414处,主题行不指示文档应该被附加,那么模块215将开始对消息的更加详细的分析。将理解,消息,特别是电子邮件消息,可以不是单独创建的且完全新的消息,而是可以包含来自一个或多个先前的电子邮件消息中的文本。例如,正在由用户创建的消息可以是在消息线程中的最新消息。在该线程中的较早的消息可以提及一附件,并且用户可能想要“回复”或“回复全部”以确认该消息的接收,以提供关于较早消息的内容或关于较早消息的附件的意见,以指示较早的消息不具有所指示的附件等,但可能并不想要附加任何东西到回复消息。

[0046] 因此,较佳地(但并非必要地),模块215检查424消息以标识并区分(图3中的唯一正文标识符315C)该消息中的唯一正文;就是说以从消息线程中将当前文本与旧的文本标识并区分出来。这可以通过例如(但并非限制)搜索消息中的文本以查找诸如“主题”、“发件人”、“收件人”、“发送”、“主题:回复:”、“主题:转发:”等词。在一种类型的电子邮件消息程序中,新的文本被添加到线程中的先前文本之上,这样,某人按时间的前后顺序从底部向顶部阅读线程,但某人是从顶部向底部阅读线程内的个别消息。在这种情况下,这样的词通常指示在这些词之下(跟随这些词)的文本来自先前的消息,并且,在该情况中,在这样的一个或多个词之后的文本不被考虑。

[0047] 在另一种类型的电子邮件消息程序中,新的文本被添加在所述线程中的先前文本之下,这样,某人从顶部向底部阅读整个线程。在这种情况下,这样的词通常指示在最低的“收件人”或“发件人”或“主题”行之上(之前)的文本是来自先前的消息,并且在该情况中,这样的文本不被考虑。或者,当用户点击“回复”或“回复全部”时文档处理程序210可以将

唯一的字符或代码插入到消息中，并且模块215在检测到该字符或代码时将知道在该字符之上(或之下)的文本是新的文本。

[0048] 一旦已经标识了消息的当前文本部分，那么在操作426处模块215分析该部分以确定是否建议附件。下文参考图5更详细地讨论该确定。如果在操作426处执行的分析指示428建议附件，则如上所讨论的，在操作416处模块215向用户显示对话框或其它类型的用户界面元素，由其向用户询问是否应该有对电子邮件消息的附件。

[0049] 如果在操作426处执行的分析不指示428建议附件，那么，在操作408处模块215将较佳地(但非必要地)更新其检测规则数据库，并在操作410处将允许消息如同用户最初撰写地那样被发送，或将允许消息被传递到另一个程序上，例如拼写检查器程序、元数据移除程序或病毒检查程序，以供在被发送之前进行进一步处理。

[0050] 其它过滤器(图3的组件315D)还可以也存在，并且依赖于其它规则、关键字和/或短语以否决来自不同过滤器的已经确定应该包括附件的指示。例如，短语“附加的车库”使用关键字“附加的”并将指示应该包括附件。然而，否定过滤器“附加的车库”将否决该指示。其它这样的否定过滤器短语包括，例如，“附加到我的房屋”、“附加的栅栏”、“封闭的庭院”、“封闭的花园”等。

[0051] 图5示出了用于分析消息中的当前文本以确定该消息是否可能应该具有附加于其的附件的示例性过程500。过程500在操作502处开始，其中，该文本通过将该文本的每个部分标识为要么词要么标点来“词例化”。随后在操作504处分析标点字符以通过定位句子分隔符来标识句子片段。例如，句号(“.”)经常被用作句子的分隔符。然而，句号还可以是数字中的小数点、电子邮件的部分或网站URL的部分。然而，如果句号正被用作句子的分隔符，通常其后会跟随有一个或多个空格，或段落标记(有时原先被称为“回车符”)。一些用户不使用句号或不始终如一地使用句号以指示句子的结束，并且因此段落标记也能被用于标识句子的结束，特别是当存在两个或更多段落标记而没有插入字符或在段落标记之间仅有一个或多个插入空格字符时。

[0052] 一旦已经将词分隔成句子，那么，在操作506处短句子较佳地被“丢弃”，即它们不再被分析。例如，“嗨，鲍勃。”和“谢谢！”将被丢弃。还有，在操作508处至少相对于图5的正则表达式分析510不包含一个或多个关键字的句子也可以被丢弃。一些关键字的示例是，但不局限于，附加、被附加、附寄、被附寄、附件、被附属、附录、文档、演示、文章、图形等。例如，“我们正在书写关于2012.5.1的我们的订单状态的询问。”。并且“如你所知，燃料的价格直接影响运输成本。”将被丢弃。剩余的句子随后是用于进一步分析的候选。所述进一步分析较佳地(但并非必须地)具有两个组件，这两个组件可以被并行地、以两种顺序中的任一种串行地、或仅使用两个组件之一来实施。

[0053] 一个进一步分析的组件是“正则表达式”分析(图3的操作320A)，它在操作510生成“置信”分数。该置信分数是指示句子有多少可能指示应该包括附件的评分。置信分数可以是任意期望且便利的比例，例如从1(极其不可能)到5(极其可能)或A到E等。还可以使用其它比例，例如1到10或1到100。

[0054] “正则表达式”分析是基于规则的分析，并且较佳地(但并非必须地)查找仅生成高置信度的表达式。例如，“附加的文件…”，“被附加的是…”、“附加的是”、“一个副本被附加。”、“多个副本被附加。”、“附寄的文档…”，等。

[0055] 另一个进一步分析的组件是在操作511处的统计分析(图3的操作320B),它也生成“置信”分数。例如,根词语“附加”或“附寄”在词“文件”或词“文档”的一个、两个或甚至N个词内,例如在“如你将从我所附加的文件中注意到的……”中。作为另一个示例,“如可以从我们的咨询通知中所见的,你的2012.5.1的订单已经发货”。在第二示例中,“附加”、“附寄”、甚至“文档”中的任意一个都没有被使用,因此,“正则表达式”分析将不会指示附件是可能的,但统计分析将基于短语“如可所见的”指示附件是可能的。在操作508处被丢弃的非关键字句子上执行操作511也是较佳的。

[0056] 统计分析可以基于通过例如图1的系统130监视来自特定用户的反馈、监视来自多个用户的反馈,或这两者的组合。例如,如果统计分析查找特定的短语并确定很可能应该有附件并关于附件询问用户,但用户始终(或至少大多数时间)回复并指示没有附件,那么,当该短语出现时的置信等级将被降低和/或至少对于该用户来说该短语可以从统计分析中移除。注意,短语可以具有对于大多数用户来说应该有附件的高置信度,但是基于来自特定用户的反馈可以指示相对于该特定用户的附件的低置信。将会理解,某些文本串可以同时满足正则表达式分析和统计分析。

[0057] 另外,用户可以,如果需要,添加当附件被指示时该用户通常使用的词或短语。例如,用户可以经常使用“纸”或“证据”或某些速记词语取代“附件”或“附寄”。较佳地(但并非必须地),该信息将被用于更新用于其它用户的统计分析。

[0058] 另一个进一步分析的组件是在操作512处的模板匹配分析(图3的操作320C),它也能生成“置信”分数。模板匹配类似于(但不等同于)短语检查。模板的一个示例是“请<词#1>附加的<词#2>”,其中如果<词#1>匹配模板则它将是动词,并且如果<词#2>匹配模板则它将是名词。较佳地(但并非必须地),数据库会包含用于该模板的可接受的动词和可接受的名词的列表。其它模板可以使用不同的动词和名词。这样,例如,如果在消息中的文本是“请查阅附加的文档”,那么,是动词的“查阅”将满足<词#1>的要求,而是名词的“文档”将满足<词#2>的要求。因此,这将指示用户想要消息具有附件的置信。然而,相对地,如果消息中的文本是,例如“请忽略附加的残破的车库”,那么,“忽略”将是动词,但较佳地不是在数据库中的动词列表中的一个动词,而“残破的”是形容词,不是名词,因此,这将指示该短语不是指示用户想要消息具有附件。其它模板、动词列表和名词列表是可能的且被构思的。模板较佳地(但并非必要地)基于用户对消息应该具有附件的建议的响应被更新。

[0059] 在操作418处的对在操作416处所显示的问题(例如“该消息有附件吗?”)的来自用户的响应(例如“是,我想要添加附件”或“不,我不想添加附件”)较佳地(但并非必要地)由模块215发送到图1的中央数据收集系统130,该系统查看多个用户的响应以更新并微调正则表达式数据库和统计分析数据库这两者。随后,这些经更新的数据库可以被发送给运行模块215、240的设备。

[0060] 在操作514,做出关于来自正则表达式分析或统计分析之一的置信分数是否高于阈值的判定。例如,在1到5的比例上高于“4”提供了应该有附件的置信。该阈值不是关键的,并且不需要为整数值,但将会理解地,当阈值增加了则错误的附件建议的可能性将降低,但需要附件的电子邮件将在没有附件的情况下被发送的可能性将增加。相反,将会理解,当阈值减少则错误的附件建议的可能性将增加,但需要附件的电子邮件将在没有附件的情况下被发送的可能性将降低。

[0061] 如果在操作514,任一分析或这两个分析的置信分数高于阈值,那么在操作416向用户显示询问是否有附件的对话框或其它类型的用户界面元素,并且所述数据库较佳地(但并非必须地)取决于用户的响应在操作408或420处被更新。如果在操作514处,这两个分析的置信分数都低于阈值,那么,所述数据库较佳地(但并非必须地)在操作408处被更新,并且随后消息被如同用户最初撰写地那样在操作410处被发送。

[0062] 可能发生文档处理程序210在计算机设备110上,但附件检测模块在服务器115或ISP 120上运行的情况。在这个环境中,一旦检测到建议附件,设备115、120将保持消息直到从用户接收到响应,并且如果用户指示该用户希望将文档附加到消息,设备115、120将消息从用户的“发件箱”文件夹或“已发送”文件夹移动到或使得其被移动到用户的“草稿”文件夹以供由用户进一步处理。

[0063] 虽然,在此描述的操作、模块和组件是参照电子邮件程序作出的,但它们并不局限于此种使用。它们可以在少量或没有修改的情况下被用于可以被配置为包括附件的任意通信上。一种类型的这样的通信是,例如,在允许附件的SMS(短消息系统)系统中的SMS消息。另一个示例是用户可以在公司的网页上提供个人或订单信息,并且该网页允许(但并非要求)用户附加文档的情形。那么,当用户点击,例如“提交”按键时,完成的网页和任意指定的附件被保存和/或发送给公司内的指定的电子地址。在此所述的操作、模块和组件可以被用于查找在由用户所完成的文本字段中的关键字和/或短语以便确定用户是否可能忘记包括附件。

[0064] 还有,虽然在此的讨论提及的是以点击按键或按下图标等形式的用户输入,但其它形式的用户输入也是被期望和构思的。用户还可以通过语音或姿势来指示指令。例如,用户可以说“发送”或“没有附件地发送”或“没有附件”或“添加附件”或“需要附件”,并且语音识别软件将解释并实现该语音命令。作为另一个示例,用户可以向右摆姿势以指示“发送”或“没有附件地发送”或“没有附件”,并且向左摆姿势以指示“添加附件”或“需要附件”,并且运动检测软件将解释并实现该姿势。

[0065] 图6示出了能够如在此处所述那样执行的设备110、115、120、130的示例性计算机体系结构600。因此,计算机体系结构600示出了用于服务器计算机、移动电话、PDA、智能电话、台式计算机、上网本计算机、平板计算机和/或膝上型计算机的体系结构。计算机体系结构600可用于执行本文所呈现的软件操作的任何方面。

[0066] 示例的计算机体系结构600包括中央处理单元602(“CPU”)、包括随机存取存储器606(“RAM”)和只读存储器(“ROM”)608的系统存储器604、以及将存储器604耦合至CPU 602的系统总线610。基本输入/输出系统被存储在ROM 608中,该系统包含帮助诸如在启动期间计算机体系结构600中的元件之间传递信息的基本例程。计算机体系结构600还包括用于存储操作系统614以及一个或多个程序或模块620(例如如上相对图1-5所述的模块)的大容量存储设备612。

[0067] 大容量存储设备612通过连接至总线610的大容量存储控制器(未示出)连接至CPU 602。大容量存储设备612及其相关联的计算机可读介质为计算机体系结构600提供非易失性存储。虽然对此处包含的计算机可读介质的描述引用了诸如硬盘或CD-ROM驱动器之类的大容量存储设备,但是本领域的技术人员将理解,计算机可读介质可以是由计算机体系结构600访问的任何可用计算机存储介质或通信介质。

[0068] 尽管存储器604和大容量存储设备612优选是分开的组件,存储器604也可以被包括在大容量存储设备612中。存储器604和大容量存储器612可一起被看成是并被称为存储设备。

[0069] 也可存在其它硬件组件626。例如,可以存在传感器组件,诸如磁力计、环境光传感器、邻近度传感器、加速计、陀螺仪、全球定位系统(GPS)传感器、麦克风或其它音频输入设备、相机等。

[0070] 作为示例而非限制,计算机存储介质可包括以用于存储诸如计算机可读指令、数据结构、程序模块或其它数据等信息的任何方法或技术实现的易失性和非易失性、可移动和不可移动介质。例如,计算机存储介质包括,但不仅限于, RAM、ROM、EPROM、EEPROM、闪存或其他固态存储器技术, CD-ROM、数字多功能盘(“DVD”)、HD-DVD、蓝光或其他光学存储,磁带盒、磁带、磁盘存储器或其他磁存储设备,或可以用来存储所需信息并可由计算机体系结构600访问的任何其他介质。为了声明的目的,短语“计算机存储介质”及其变型本质上不包括波、信号和/或其他瞬态和/或无形通信介质。

[0071] 通信介质包括诸如载波或其它传输机制等已调制数据信号中的计算机可读指令、数据结构、程序模块或其它数据,且包含任何传递介质。术语“已调制数据信号”指的是其一个或多个特征以在信号中编码信息的方式被更改或设定的信号。作为示例而非限制,通信介质包括諸如有线网络或直接线连接之类的有线介质,以及诸如声学、RF、红外及其他无线介质之类的无线介质。任何上述各项的组合也包括在计算机可读介质的范围内。

[0072] 根据各实施例,计算机体系结构600可以使用通过诸如网络125之类的网络到远程计算机和/或服务器的逻辑连接来在联网环境中操作。计算机体系结构600可以通过连接至总线610的网络接口单元616来连接到网络。网络接口单元616也可用于连接到其他类型的网络和远程计算机系统。计算机体系结构600还可以包括用于接收并处理来自用户输入设备630(例如但不限于,键盘、鼠标、触摸屏、触摸板、小键盘或电子指示笔,在图6中以虚线示出)的输入的输入/输出控制器618。类似地,输入/输出控制器618可以向一个或多个用户显示设备632(例如但不限于显示屏、打印机或其它类型的输出设备,在图6中也以虚线示出)提供输出。用户输入设备630和用户输出设备632可以在相同组件(例如触敏屏)中被实现。用户输入设备630和用户输出设备632可以与一个设备集成,例如在手持设备的情况下,或可以是分开的组件,例如与许多桌面系统一起使用的键盘、鼠标和显示器。

[0073] 在此所描述的软件组件在被加载到CPU 602中并被执行时将CPU 602和总体计算机体系结构600从通用计算系统变换成为被定制为促进本文提出的功能的专用计算系统。CPU 602可以用任意数量的晶体管或其他分立的电路元件(它们可以分别地或共同地呈现任意数量的状态)构建。更具体地,CPU 602可以响应于包含在本文所公开的软件模块中的可执行指令来作为有限状态机进行操作。CPU 602可以是单处理器,或可以是多处理器。这些计算机可执行指令可以通过指定CPU 602如何在各状态之间转换来变换CPU 602,由此变换了构成CPU 602的晶体管或其它分立硬件元件。

[0074] 对本文所提出的软件模块的编码也可变换本文所提出的计算机可读介质的物理结构。在本说明书的不同实现中,物理结构的具体转换可取决于各种因素。这样的因素的示例可以包括,但不仅限于:用于实现计算机可读介质的技术、计算机可读介质被表征为主存储器还是辅存储器等等。例如,如果计算机可读介质被实现为基于半导体的存储器,则本文

所公开的软件可以通过变换半导体存储器的物理状态而在计算机可读介质上编码。例如，软件可以变换构成半导体存储器的晶体管、电容器或其它分立电路元件的状态。软件还可变换这些组件的物理状态以在其上存储数据。

[0075] 作为另一示例，本文所公开的计算机可读介质可以使用磁或光技术来实现。在这些实现中，本文所提出的软件可以在磁或光介质中编码了软件时变换所述磁或光介质的物理状态。这些变换可包括更改给定磁性介质内的特定位置的磁性特征。这些变换还可以包括改变给定光学介质内的特定位置的物理特征或特性，以改变这些位置的光学特性。在没有偏离本说明书的范围和精神的情况下，物理介质的其他变换也是可以的，前面提供的示例只是为了便于此讨论。

[0076] 鉴于以上内容，在计算机体系结构中发生许多类型的物理变换以便存储并执行本文所提出的软件组件。还有，计算机体系结构可以包括其它类型的计算设备，包括：手持式计算机、嵌入式计算机系统、个人数字助理、以及本领域技术人员已知的其它类型的计算设备。还可以构想的是，计算机体系结构可以不包括在此所示的全部组件、可以包括未在此明确定出的其他组件、或者可利用完全不同于在此所示的体系结构。

[0077] 以上所述的主题仅作为说明提供，并且不应被解释为限制。可对在此描述的主题作出各种修改或改变，无需跟随在此示出和描述的各示例实施例和各应用。虽然用计算机结构特征、方法和变换动作、特定计算机器、以及计算机可读介质专用的语言描述了本文中所描述的主题，但是应当理解，所附权利要求 不必限于本文中所描述的具体特征、动作、或介质。相反，这些具体特征、动作以及介质是作为实现权利要求的示例形式而公开的。

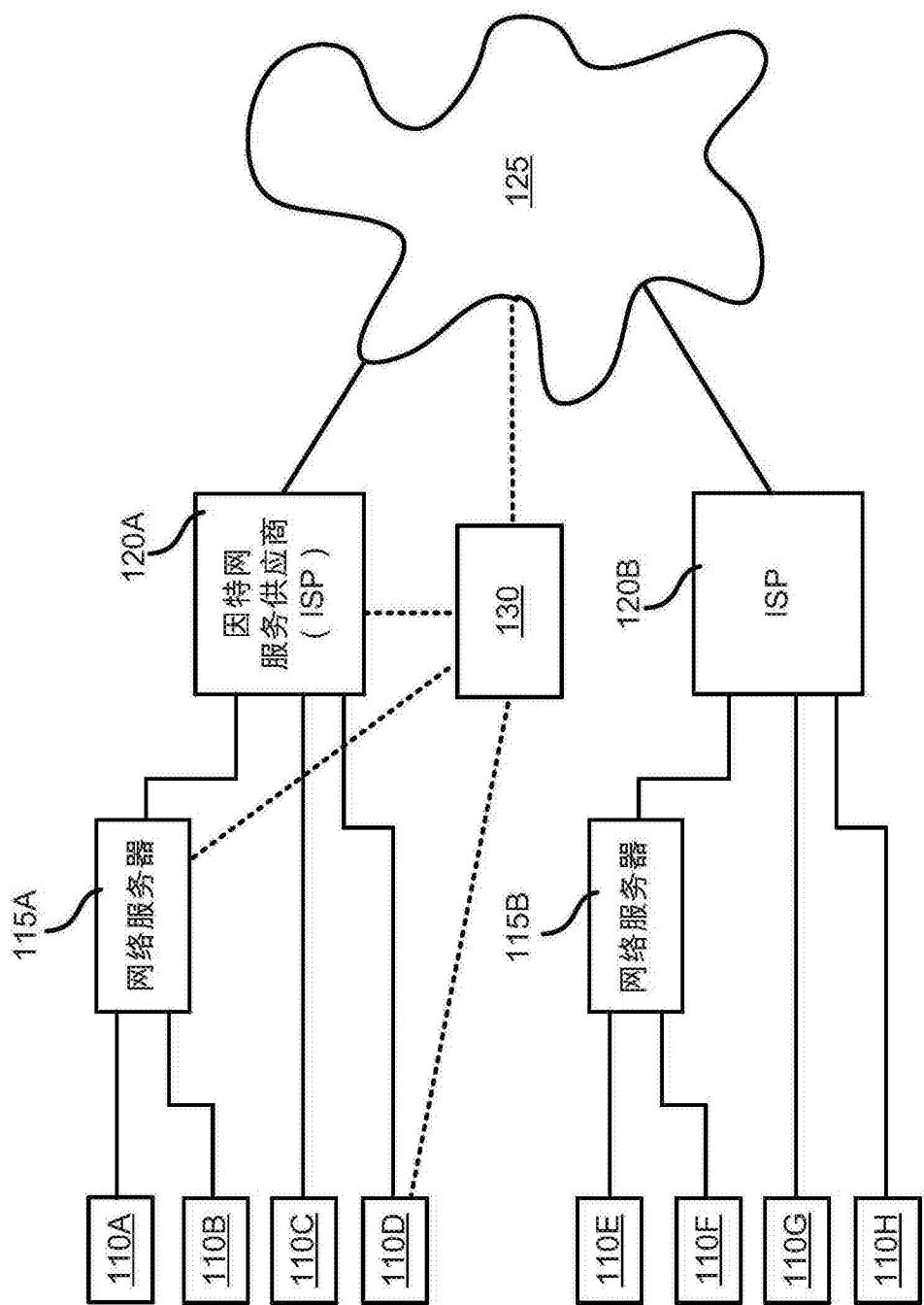


图1

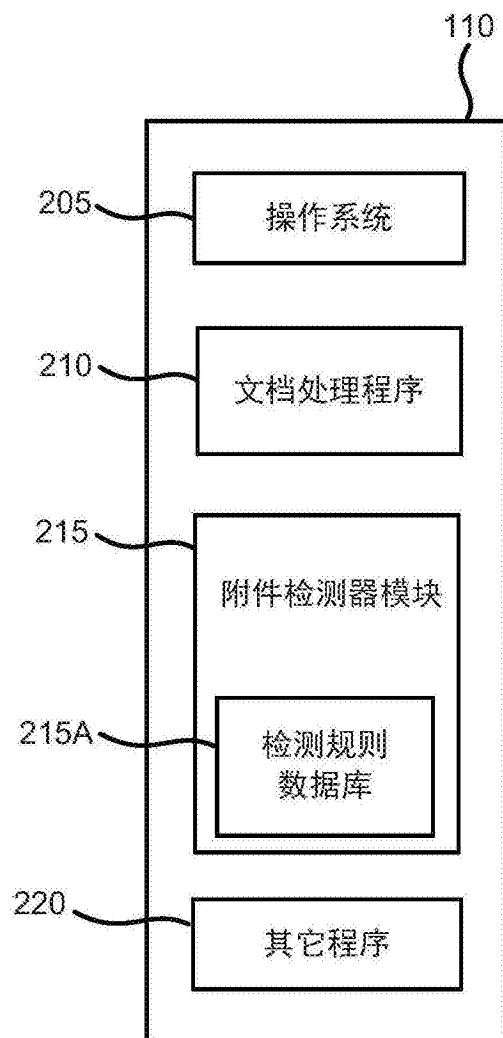


图2A

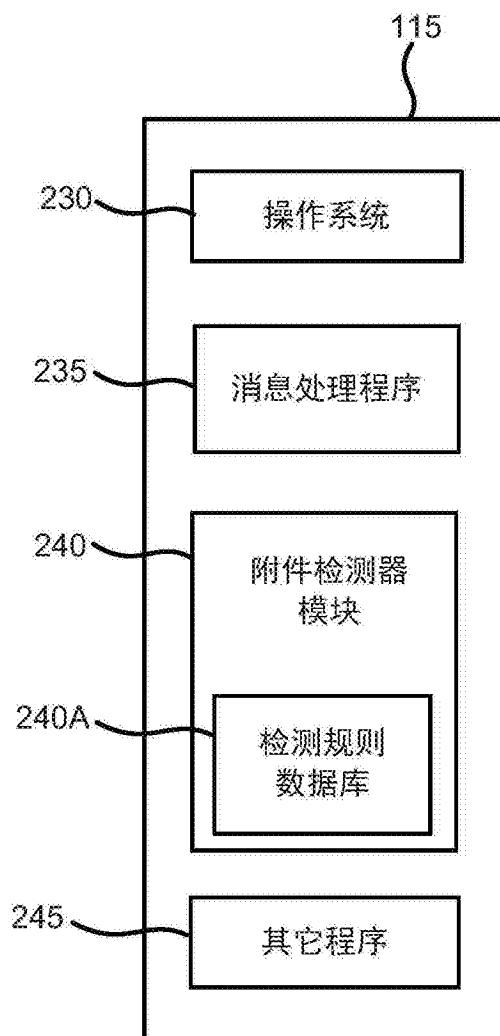


图2B

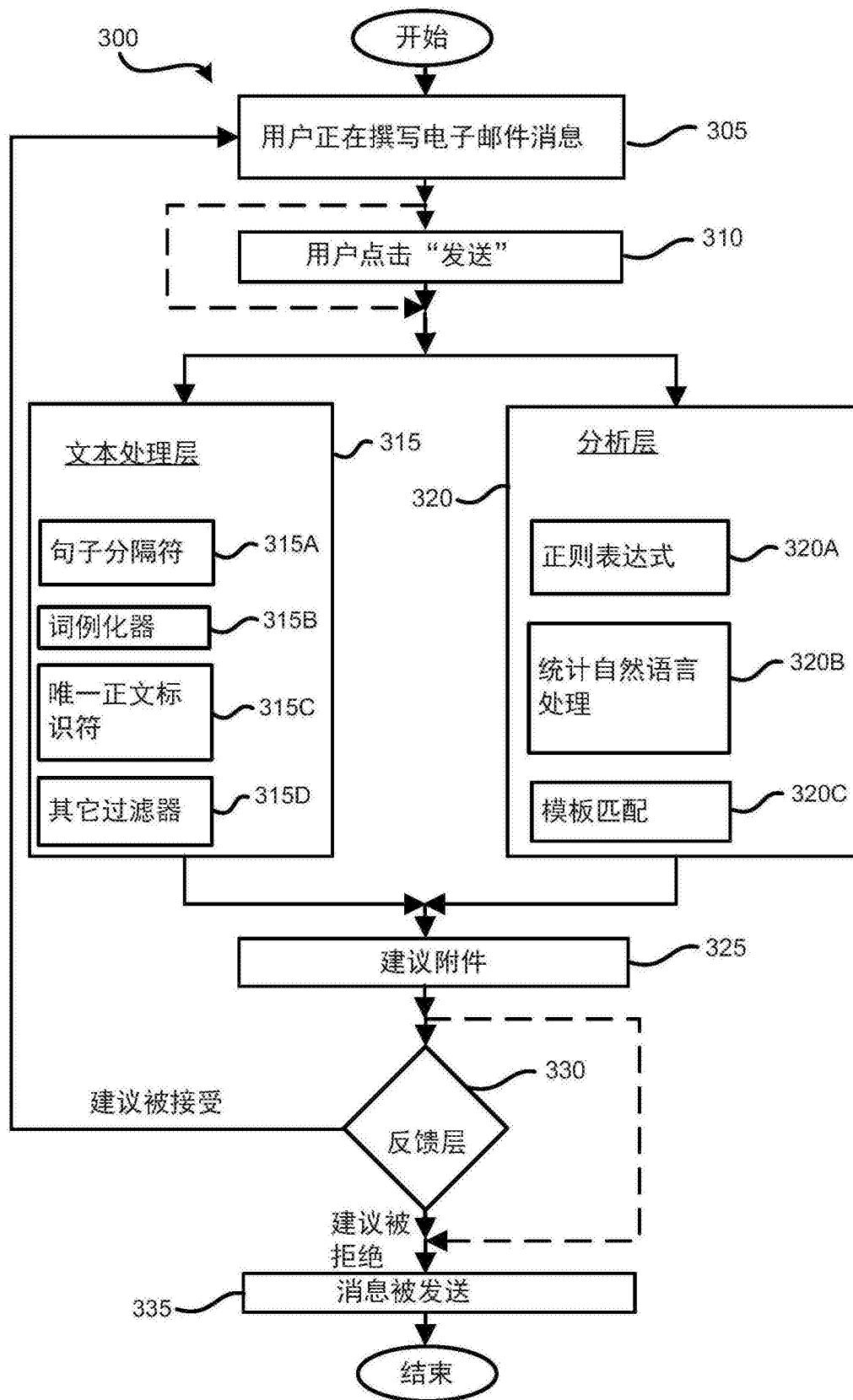


图3

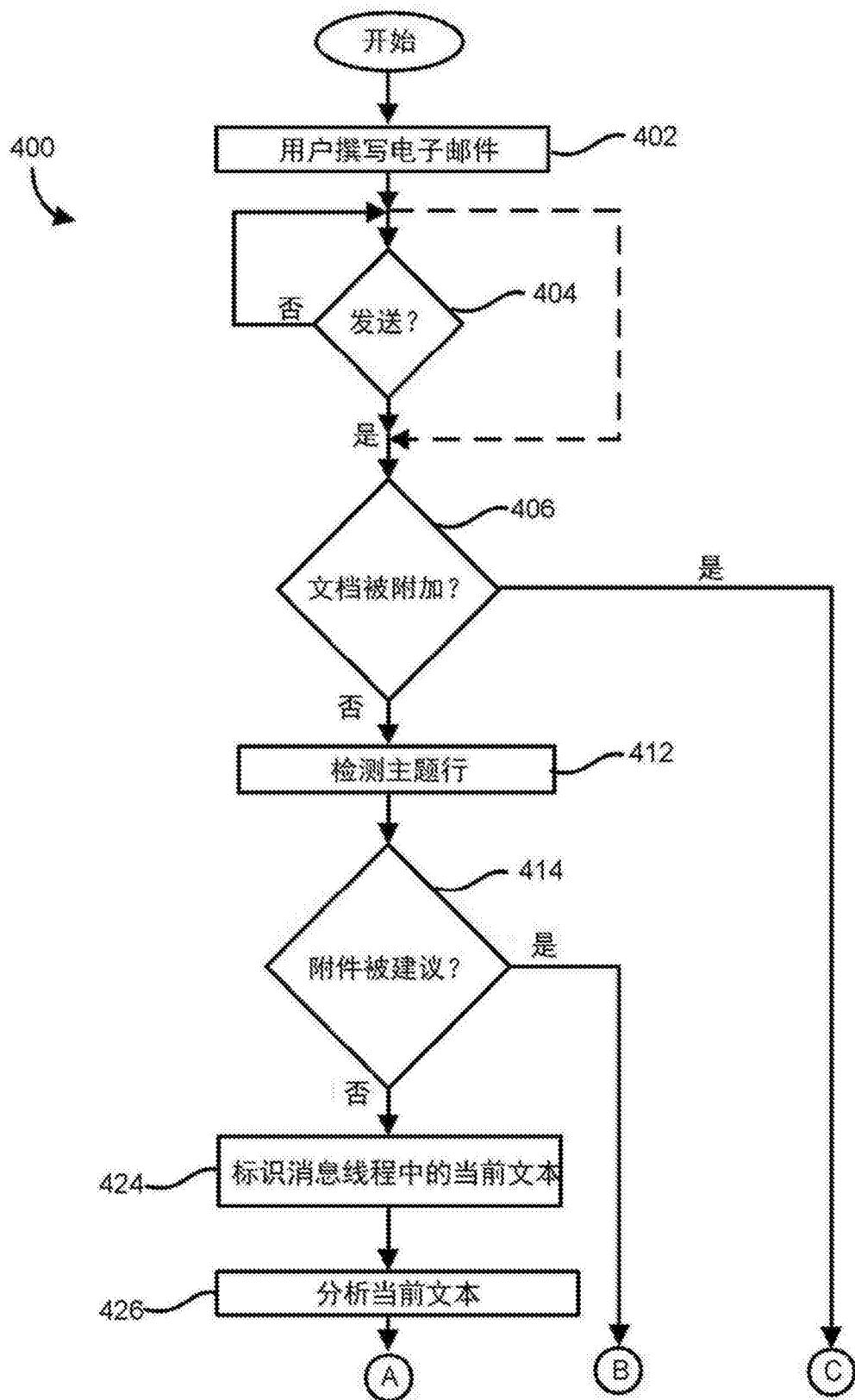


图4A

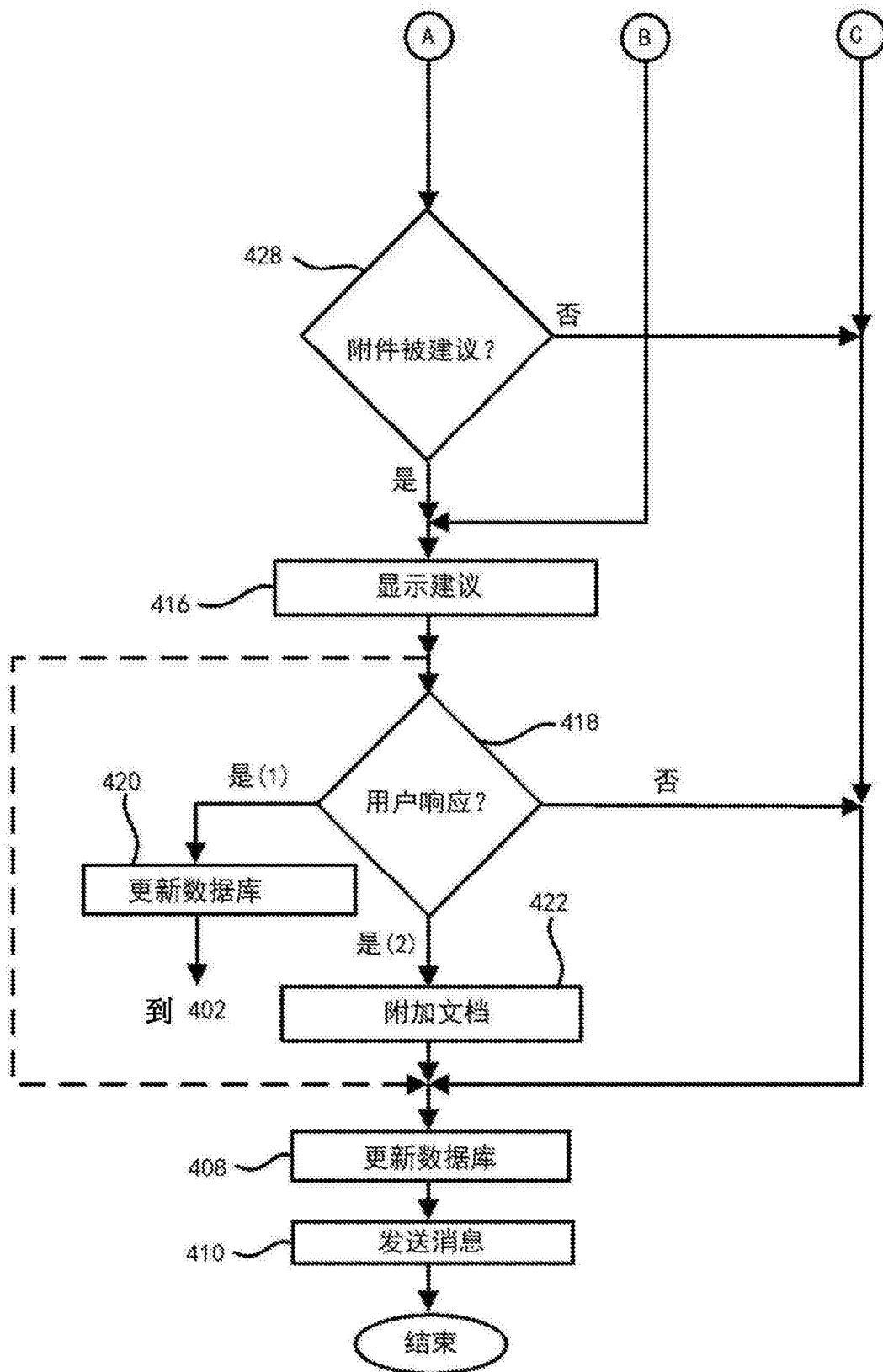


图4B

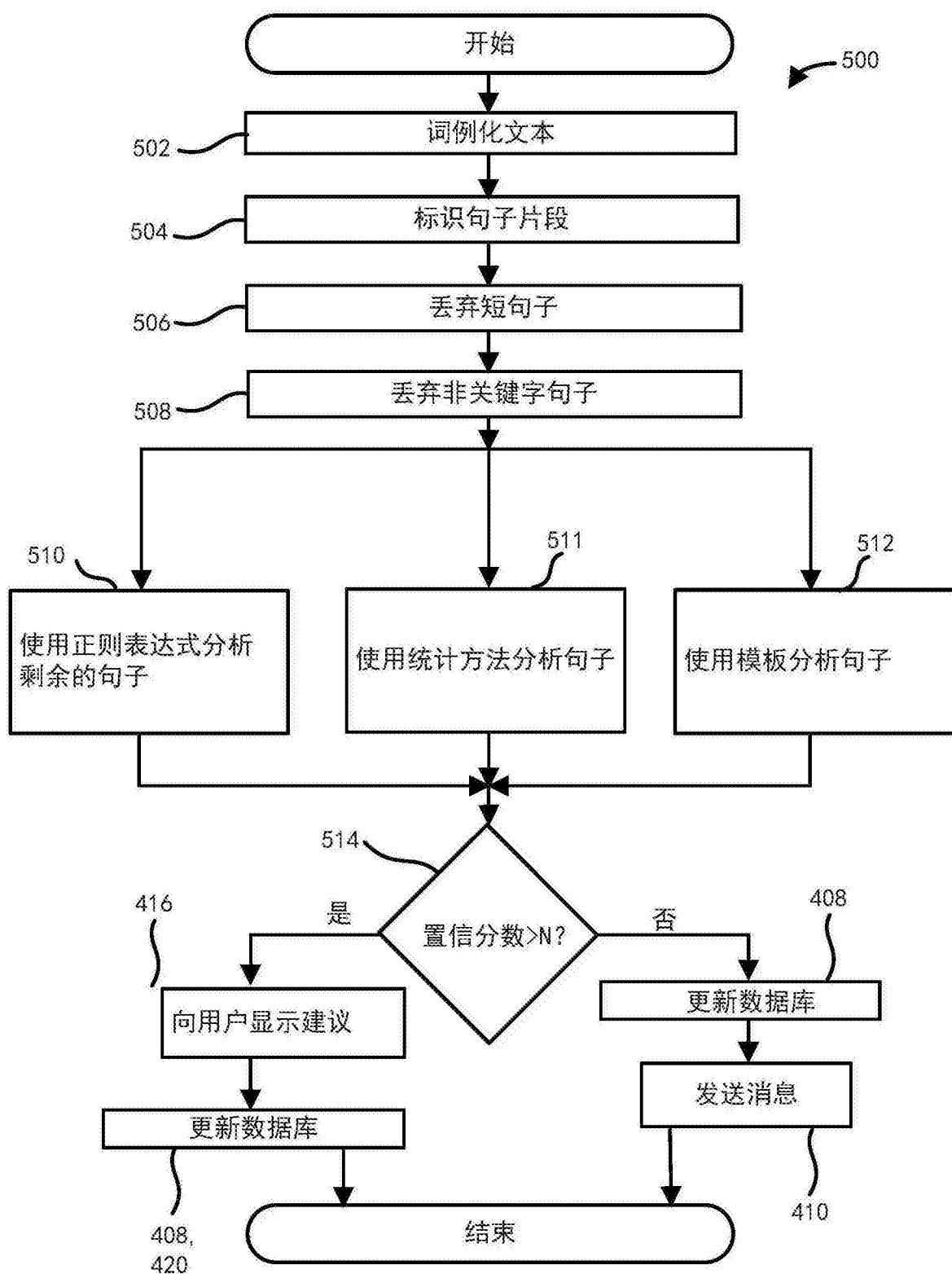


图5

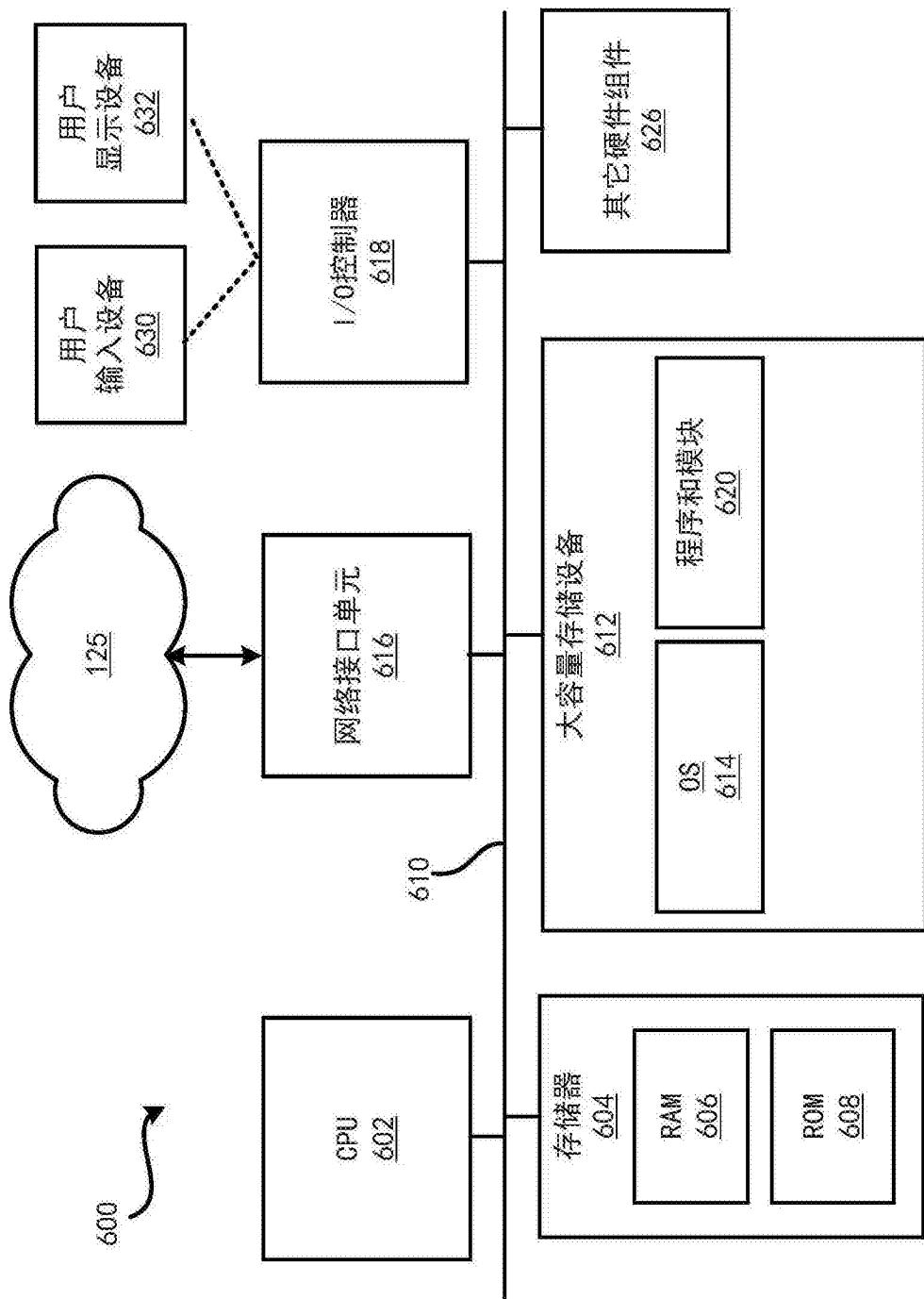


图6