



(12) 发明专利

(10) 授权公告号 CN 101192231 B

(45) 授权公告日 2012.07.04

(21) 申请号 200610163022.5

23-26行, 第2栏第23-24行, 第32-33行, 第3栏第14-18行, 第4栏第55-57行, 附图1A.

(22) 申请日 2006.11.27

US 6219679 B1, 2001.04.17, 说明书第5栏第35-37行, 第55-65行, 第6栏第13-27行, 附图1A, 附图3F.

(73) 专利权人 国际商业机器公司  
地址 美国纽约

审查员 李劲娴

(72) 发明人 聂伯敏 马嶽

(74) 专利代理机构 中国国际贸易促进委员会专  
利商标事务所 11038  
代理人 袁玥

(51) Int. Cl.

G06F 17/30(2006.01)

G06F 17/24(2006.01)

(56) 对比文件

US 6219679 B1, 2001.04.17, 说明书第2栏第40-44行, 第5栏第35-37行, 第55-65行, 第6栏第13-27行, 第7栏第5-9行, 附图1A, 附图3F.

US 5619709 A, 1997.04.08, 说明书第1栏第

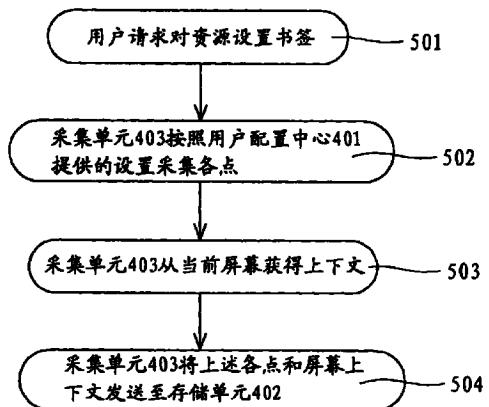
权利要求书2页 说明书8页 附图5页

(54) 发明名称

基于上下文的书签

(57) 摘要

本发明涉及对用户访问的资源设置书签以及使用该书签的方法和装置。书签设置方法,包括响应对资源的当前屏幕设置书签的请求,对资源的当前屏幕的实际文本采集屏幕上下文信息;以及存储所述资源的地址信息和屏幕上下文信息作为用于返回所述资源特定部分的书签。根据设置的书签提供资源的特定部分的方法包括:响应用户重新打开资源的请求,将由所述地址信息识别的资源内容下载到存储器;对下载的资源内容采集屏幕上下文信息,和比较所下载资源内容的屏幕上下文信息和所述书签指示的屏幕上下文信息,根据模糊匹配的规则选定所下载的资源的特定部分并在显示窗口中显示所选定的资源特定部分。



1. 一种在数据处理系统中对资源的特定部分设置书签的方法，包括如下步骤：

响应对资源的当前屏幕设置书签的请求，对资源的当前屏幕的实际文本采集屏幕上下文信息；

对所述资源的当前屏幕的实际文本编写摘要，所述摘要用于在对所述资源设置书签后再访问该资源时识别再访问的资源是否为有效资源；以及

存储所述资源的地址信息、屏幕上下文信息和所述摘要作为用于返回所述资源特定部分的书签。

2. 根据权利要求 1 所述的对资源的特定部分设置书签的方法，其中根据预定的采集率对资源的当前屏幕采集屏幕上下文信息。

3. 根据权利要求 2 所述的对资源的特定部分设置书签的方法，其中所述采集率是由用户输入的。

4. 根据权利要求 2 所述的对资源的特定部分设置书签的方法，其中所述采集率是从存储单元获取的。

5. 根据权利要求 1 所述的对资源的特定部分设置书签的方法，其中所述屏幕上下文信息包括文本定位器元件。

6. 根据权利要求 5 所述的对资源的特定部分设置书签的方法，其中所述文本定位器元件包括当前屏幕中的第一个词、某个特定词或者第一段。

7. 根据权利要求 5 所述的对资源的特定部分设置书签的方法，其中所述屏幕上下文信息还包括文本定位辅助信息。

8. 根据权利要求 7 所述的对资源设置书签的方法，其中所述文本定位辅助信息包括行号、段落号、屏幕分辨率和 / 或文本布置和显示方式中的至少之一。

9. 根据权利要求 1 所述的对资源的特定部分设置书签的方法，其中对当前屏幕的实际文本编写摘要的步骤包括根据预定的采集率对当前屏幕编写摘要的步骤。

10. 一种根据设置的书签提供资源的特定部分的方法，所述书签包括地址信息、通过采集所述资源特定部分获得的屏幕上下文信息，和对资源的实际文本编写的摘要信息，所述方法包括如下步骤：

响应用户重新打开资源的请求，将由所述地址信息识别的资源内容下载到存储器；

对下载的资源内容采集屏幕上下文信息；

对下载的资源内容编制摘要，并且比较该摘要与所述书签指示的摘要，进而判断下载的资源内容是否有效；和

比较所下载资源内容的屏幕上下文信息和所述书签指示的屏幕上下文信息，根据模糊匹配的规则选定所下载的资源的特定部分并在显示窗口中显示所选定的资源特定部分。

11. 根据权利要求 10 所述的根据设置的书签提供资源的特定部分的方法，其中所述书签的屏幕上下文信息是通过以预定的采集率采集所述资源特定部分获得的，

所述对下载的资源内容采集屏幕上下文信息的步骤包括使用相同的预定采集率对下载的资源内容采集屏幕上下文信息。

12. 根据权利要求 11 所述的根据设置的书签提供资源的特定部分的方法，其中所述书签的屏幕上下文信息包括文本定位器元件，

所述使用预定采集率对下载的资源内容采集屏幕上下文信息的步骤包括：

在下载的资源内容中检索与所述文本定位器元件匹配的部分；

基于所述与文本定位器元件匹配的部分使用预定采集率对下载的资源内容采集实际文本，获得各个用于与所述书签的屏幕上下文信息进行比较的瞬态模拟屏幕。

13. 根据权利要求 12 所述的根据设置的书签提供资源的特定部分的方法，其中所述文本定位器元件包括当前屏幕中的第一个词、某个特定词或者第一段。

14. 根据权利要求 12 所述的根据设置的书签提供资源的特定部分的方法，其中所述屏幕上下文信息还包括文本定位辅助信息。

15. 根据权利要求 14 所述的根据设置的书签提供资源的特定部分的方法，其中所述文本定位辅助信息包括行号、段落号、屏幕分辨率和 / 或文本布置和显示方式中的至少之一。

16. 一种在数据处理系统中对资源的特定部分设置书签的书签设置装置，包括：

采集单元，用于响应对资源的当前屏幕设置书签的请求，对资源的当前屏幕的实际文本采集屏幕上下文信息；

摘要编写器，对所述资源的当前屏幕的实际文本编写摘要，所述摘要用于在对所述资源设置书签后再访问该资源时识别再访问的资源是否为有效资源；以及

存储单元，用于存储所述资源的地址信息、屏幕上下文信息和所述摘要作为用于返回所述资源特定部分的书签。

17. 根据权利要求 16 所述的书签设置装置，其中摘要编写器根据预定的采集率对当前屏幕编写摘要。

18. 一种根据设置的书签提供资源的特定部分的装置，所述书签包括地址信息、通过采集所述资源特定部分获得的屏幕上下文信息，和对资源的实际文本编写的摘要信息，所述装置包括：

比较单元，用于响应用户重新打开资源的请求，将由所述地址信息识别的资源内容下载到存储器，并对下载的资源内容采集屏幕上下文信息，比较所下载资源内容的屏幕上下文信息和所述书签指示的屏幕上下文信息，以及在下载资源内容之后对下载的资源内容编制摘要信息，并且比较该摘要信息与所述书签指示的摘要信息，进而判断下载的资源内容是否有效；

定位器，用于根据模糊匹配的规则选定所下载的资源的特定部分并在显示窗口中显示所选定的资源特定部分。

## 基于上下文的书签

### 技术领域

[0001] 本发明涉及对用户访问的资源设置书签以及使用该书签的方法和装置。

### 背景技术

[0002] 众所周知,万维网或因特网是支持超文本标注语言格式 (HTML) 的文档的服务器系统。HTML 支持通向文档以及图形、音频和视频文件等的链接。从其它文档引用指向文档的链接,该链接允许用户仅仅通过鼠标点击容易地从一个文档或 Web 页转入另一个文档或 Web 页。因此链接是非常有用的因特网导航工具。文档或 Web 页可被互换使用。

[0003] 另一个有用的因特网导航工具是书签。书签应用在大部分 Web 浏览器中,Web 浏览器是用于定位 Web 站点并显示 Web 页的软件应用程序。在 Web 浏览器中,书签允许用户在文件夹例如书签文件夹中存储 Web 地址或 Web 页的 URL(统一资源定位符)来指定协议如 HTTP 或 FTP 以及对象、文档、万维网网页或其他目标在 Internet 或 Intranet 上的位置,以便随后访问。

[0004] 当用户在访问那些希望以后再访问的 Web 页时可以对该 Web 页设置书签。但是当随后访问该 Web 页时,该 Web 页经常从其开端开始显示。如果用户希望返回到 Web 页的中间或结尾时需要向下滚动滚动条来找寻相关段落,这个过程非常费力,尤其是当显示的文档很大时。

[0005] 因此需要使书签具有显示访问对象的特定部分的功能。现有技术中针对 Office 文档等可编辑文档提供了利用光标的位置设置书签的功能。具体地在定位书签时,编辑者需要将光标移动到可编辑文档中希望设置书签的位置进行书签设置,从而在其它可编辑文档中以 UNC 路径(通用命名规则)引用该书签时可以显示已设置书签的文档的特定部分。

[0006] 但是这种可以指向对象的特定部分的书签的设置取决于光标的位置,其书签系统要求指出书签所在的光标位置。对于可编辑文档,这是自然的,但是对于网页以及其它不可编辑文档例如只读 PDF 却无法做到这一点。现在随着网站提供的电子书籍以及软件材料的普及,希望提出一种对这些不可编辑文档设置书签的新书签技术。

[0007] 业界也曾提出了对网页的各种解决方案。例如 US2004/0205543A1 公开了允许用户在设置书签时,在存储页面 URL 的同时还存储显示该页面的窗口的尺寸、该窗口内的水平和垂直滚动块的位置以及用于显示该页面的字体属性,这样在重新访问设置了书签的 Web 页时允许仅仅显示具体的页面部分。US6219679B1 教导了使用显示区域左上角位置和右下角位置的坐标对作为书签定位符的一部分,由此限定了再访问时显示的页面部分。

### 发明内容

[0008] 考虑到以上的现有技术,本发明致力于解决现有技术中利用光标、坐标等可定位信息设置书签的缺陷,提出一种既可为不可编辑文档设置书签、还可为可编辑文档设置书签的技术方案。

[0009] 根据本发明的一个方面,一种在数据处理系统中对资源的特定部分设置书签的方

法,包括如下步骤:

[0010] 响应对资源的当前屏幕设置书签的请求,对资源的当前屏幕的实际文本采集屏幕上下文信息;以及

[0011] 存储所述资源的地址信息和屏幕上上下文信息作为用于返回所述资源特定部分的书签。

[0012] 本发明提出通过采集屏幕上下文信息来反映显示屏幕的文档内容和结构,将此屏幕上下文信息作为书签的一部分内容,由此在再次访问设置了书签的资源内容时可以利用已记录的屏幕上下文信息与再次访问的资源内容相比较,从而选定所下载的资源的特定部分。

[0013] 其中根据用户输入或从存储单元获取的一定的采集率对资源的当前屏幕采集屏幕上下文信息。

[0014] 所述屏幕上下文信息可以包括文本定位器元件,例如当前屏幕中的第一个词、某个特定词或者第一段;还可包括用于辅助文本定位器元件定位文档的文本定位辅助信息,例如行号、段落号、屏幕分辨率和/或文本布置和显示方式中的至少之一。

[0015] 优选地,本发明的对资源设置书签的方法还包括对资源的实际文本编写摘要信息的步骤,并且存储所述编写的摘要信息作为所述书签的一部分。对当前屏幕编写摘要的步骤可以包括根据预定的采集率对当前屏幕编写摘要的步骤。

[0016] 根据本发明的另一方面,一种根据设置的书签提供资源的特定部分的方法,所述书签包括地址信息和通过采集所述资源特定部分获得的屏幕上下文信息,所述方法包括如下步骤:

[0017] 响应用户重新打开资源的请求,将由所述地址信息识别的资源内容下载到存储器;

[0018] 对下载的资源内容采集屏幕上下文信息,和

[0019] 比较所下载资源内容的屏幕上下文信息和所述书签指示的屏幕上下文信息,根据模糊匹配的规则选定所下载的资源的特定部分并在显示窗口中显示所选定的资源特定部分。

[0020] 类似地,所述对下载的资源内容采集屏幕上下文信息的步骤包括使用相同的预定采集率对下载的资源内容采集屏幕上下文信息。

[0021] 另外,所述书签的屏幕上下文信息包括文本定位器元件,相应地所述使用预定采集率对下载的资源内容采集屏幕上下文信息的步骤包括:

[0022] 在下载的资源内容中检索与所述文本定位器元件匹配的部分;

[0023] 基于所述与文本定位器元件匹配的部分使用预定采集率对下载的资源内容采集实际文本,获得各个用于与所述书签的屏幕上下文信息进行比较的瞬态模拟屏幕。

[0024] 同样优选地所述书签还包括对资源的实际文本编写的摘要信息,相应地所述根据设置的书签提供资源的特定部分的方法在下载资源内容的步骤之后还包括:

[0025] 对下载的资源内容编制摘要信息,并且比较该摘要信息与所述书签指示的摘要信息,进而判断下载的资源内容是否有效的步骤。

[0026] 本发明还提出了一种在数据处理系统中对资源的特定部分设置书签的书签设置装置,包括:

[0027] 采集单元,用于响应对资源的当前屏幕设置书签的请求,对资源的当前屏幕的实际文本采集屏幕上下文信息;以及

[0028] 存储单元,用于存储所述资源的地址信息和屏幕上下文信息作为用于返回所述资源特定部分的书签。

[0029] 另外,本发明也提出了一种根据设置的书签提供资源的特定部分的装置,所述书签包括地址信息和通过采集所述资源特定部分获得的屏幕上下文信息,所述装置包括:

[0030] 比较单元,响应用户重新打开资源的请求,将由所述地址信息识别的资源内容下载到存储器,对下载的资源内容采集屏幕上下文信息,并且比较所下载资源内容的屏幕上文信息和所述书签指示的屏幕上下文信息;

[0031] 定位器,用于根据模糊匹配的规则选定所下载的资源的特定部分并在显示窗口中显示所选定的资源特定部分。

[0032] 本发明使用屏幕上下文信息来记录文档中的读取或编辑位置,可在无需光标、坐标等可定位信息的帮助下对文档设置书签;在再访问时利用模糊匹配方法比较已记录的屏幕上下文信息与再次访问的资源内容,从而精确地定位文档中的读取或编辑位置。

[0033] 本发明采用了反映文档内容和结构的屏幕上下文信息,相比 US6219679B1 等使用显示区域左上角位置和右下角位置的坐标来定位具体的显示区域的技术方案,可以不受文档实际布置方式或部分内容变化导致的坐标等信息变化的影响,因而可以精确地定位文档中的具体位置。

[0034] 另外本发明提出了对资源内容编制摘要信息,在再访问时通过比较已记录的摘要信息和再访问时编制的摘要信息,可检测文档的有效性。

[0035] 利用本发明的书签设置方法,可以使用 URL 地址或 UNC 路径转向万维网中的各种文件、文件的位置或网页,或是 Intranet 上的网页等。同样由于在同一计算机内 Office 文档 (Word、Excel、PowerPoint、Access 等) 相互之间以及 PDF 之间可以利用类似的 UNC 路径 (通用命名规则) 对文件进行定位,因此本发明的书签设置方法完全可以应用于同一计算机内 Office 文档之间、PDF 文档之间以及 Office 文档和 PDF 文档的交互访问。

[0036] 根据下面示意性实施例的说明 (参照附图) 本发明的其它特点将显而易见。

## 附图说明

[0037] 图 1 示出了根据本发明的分布式数据处理系统的示意性框图。

[0038] 图 2 示出了根据本发明的服务器设备的示意性框图。

[0039] 图 3 示出了根据本发明的客户机设备的示意性框图。

[0040] 图 4 示出了根据本发明的优选实施例用于对资源设置书签的元件的示意图。

[0041] 图 5 示出了用于根据本发明的优选实施例生成书签的方法的流程图。

[0042] 图 6 示出了用于根据本发明的优选实施例,响应对书签的选择呈现资源的方法的流程图

## 具体实施方式

[0043] 下面,将参照附图等详细描述本发明的具体实施方式。但是,本发明可以通过多种不同的方式来实施,本领域人员可以很容易地理解其方式和详细内容可以被变换为各种各

样的形式,而不脱离本发明的宗旨及其范围。因此,本发明不应该被解释为仅限定在实施方式所记载的内容中。

[0044] 参考附图,其中图 1 示出了实施本发明的数据处理系统 100 的网络示意图。数据处理系统 100 包括作为在彼此相连的各种设备和计算机之间提供通信链接的媒介的网络 102。该网络 102 可以包括有线、无线通信链接或光纤。

[0045] 在所示实例中,服务器 104 以及存储单元 106 连接到网络 102。另外,服务器 104 的客户机 108、110 和 112 可以是个人计算机和网络计算机。在所示实例中,服务器 104 提供例如引导文件、操作系统图像以及客户机 108、110 和 112 的应用程序。网络数据处理系统 100 中,网络 102 表示使用 TCP/IP 协议彼此通信的网络和网关集合的因特网。在因特网中心是在主节点或主计算机之间的高速数据通信线路的骨干,由若干商业、政府、教育和其它发布数据信息的计算机系统组成。当然数据处理系统 100 还可以在 intranet、局域网 (LAN)、广域网 (WAN) 中。

[0046] 现在根据本发明的一个实施例参考图 2 描述可以作为例如图 1 的服务器 104 的服务器实施的数据处理系统的框图。数据处理系统 200 可以是包括连接到数据总线 206 的多个处理器 202 和 204 的对称多处理器 (SMP)。或者可以应用单个处理器系统。同样连接至数据总线 206 的是存储器控制器 / 高速缓冲存储器 208,用以提供与局部存储器 209 的接口。I/O 总线桥 210 也连接到数据总线 206,用以提供与 I/O 总线 212 的接口。存储器控制器 / 高速缓冲存储器 208 和 I/O 总线桥 210 可以集成在一起。

[0047] 连接至与 I/O 总线 212 的周边元件互连 (PCI) 总线桥 214 提供了与 PCI 局部总线 216 的接口。可以有多个调制解调器连接到 PCI 局部总线 216。典型的 PCI 总线实现可以支持 4 个 PCI 扩展槽或者插入式 (add-in) 连接器。与图 1 中网络计算机 108、110 以及 112 的通信链接可以由调制解调器 218 和通过插入式主板与 PCI 局部总线 216 相连的网络适配器 220 来提供。

[0048] 附加的 PCI 总线桥 222 和 224 为附加的 PCI 局部总线 226 和 228,由此可以支持附加的调制解调器和网络适配器。以这种方式,数据处理系统 200 允许连接到多个网络计算机。存储器映射图形适配器 230 和硬盘 232 也可以如图所示直接或间接连接到 I/O 总线 212。

[0049] 现在参考图 3 描述可以应用本发明的数据处理系统的框图。数据处理系统 300 是一个客户机计算机的例子。该数据处理系统 300 应用周边元件互连 (PCI) 局部总线构架。尽管所示的例子应用了 PCI 总线,其它诸如加速图形端口 (AGP) 以及工业标准构架 (ISA) 等也可以使用。处理器 302 和主存储器 304 通过 PCI 桥 308 连接到 PCI 局部总线 306。PCI 桥 308 也包括用于处理器 302 的集成的存储器控制器和高速缓冲存储器。可以通过直接元件互连或通过插入式主板形成与 PCI 局部总线 306 的附加连接。在所示的实例中,局域网适配器 320,SCSI 主机总线适配器 312 以及扩展总线接口 314 通过直接元件连接连接到 PCI 局部总线 306。相反,音频适配器 316、图形适配器 318 以及音视频适配器 319 通过插在扩展槽中的插入式主板连接到 PCI 局部总线 306。扩展总线接口 314 提供用于键盘和鼠标适配器 320、调制解调器 332 以及附加存储器 324 的连接。小计算机系统接口 (SCSI) 主机总线适配器 312 提供了用于硬盘驱动器 326、磁带驱动器 328 以及 CD-ROM 驱动器 330 的连接。典型的 PCI 局部总线实现可支持三个或四个 PCI 扩展槽或插入式连接。

[0050] 操作系统运行在处理器 302 中,用于协调和提供对图 3 中数据处理系统 300 的各种元件的控制。该操作系统可以是市场上可获得的 Windows2000 等。面向对象编程系统例如 Java 可与该操作系统一起运行,以便提供从 Java 程序或该数据操作系统 300 中其它应用程序对操作系统的调用。操作系统的指令、面向对象操作系统以及应用程序或程序可以在存储设备例如硬盘驱动器 326 中,可以在由处理器 302 执行时加载到主存储器 304 中。

[0051] 作为另外一个实例,数据操作系统 300 可以被配置成无需依赖网络通信接口的类型而启动。另外,数据操作系统 300 也可以是个人数字助理 (PDA) 设备,与 ROM 和 / 或闪速 ROM 一起用于提供存储操作系统文件和 / 或用户生成文件的非易失性存储器。

[0052] 图 3 所示的实例以及上述实例仅仅是例举,例如数据操作系统 300 也可以是笔记本或采用 PDA 形式的手持式计算机。数据操作系统 300 还可以是资讯站 (kiosk) 或环球网设备 (web appliance)。

[0053] 以下首先简介本说明书使用的术语:

[0054] 屏幕上下文信息:屏幕上下文记录了当前屏幕的信息,包括屏幕中的第一个词,屏幕分辨率,第一段以及行号中的至少之一。对于不同文档屏幕上下文信息可能不同,例如对 html(超文本链接标示语言) 文档,屏幕上下文信息可以具有标题信息。

[0055] 采集率:采集率涉及采集文档内容。例如如果采集率是 100%,则整个文档将被记录。

[0056] 重新显示模式:重新显示模式是告诉书签设置装置,对设置了书签的文档的位置进行重新定位的方式。例如“由第一个词定位”,则在显示屏幕中的第一个词被存储为在重新定位时屏幕中的第一个词。

[0057] 有效文档:文档只有在书签设置装置自上一次对该文档设置书签后将其识别为仍然是同一个文档时才是有效文档。

[0058] 系统体系结构

[0059] 本发明提供了一种对资源设置书签的方法和装置,本发明可以应用在图 1 所示客户机设备 108、110 和 112(即个人计算机或网络计算机)中或者在服务器 104 中。本发明可以存在于由计算机系统使用的各种数据存储介质(例如软盘、光盘、硬盘、ROM、RAM 等)中。

[0060] 本发明的对资源设置书签的装置可以由图 2 或图 3 所示的数据操作系统 200 或数据操作系统 300 实现,具体地,如图 4 所示,根据设置的书签提供资源的特定部分的装置以及根据设置的书签提供资源的特定部分的装置包括 5 个单元,分别为用户配置中心 401,存储单元 402,采集单元 403,比较单元 404 和定位器 405。

[0061] 用户配置中心 401 用于设置采集整个文档的采集率、以及按照书签模式选择文档重新显示的方式,例如可以通过各种编辑工具 (Office 文档工具或 Acrobat Reader 工具) 或者浏览器的图形用户接口由终端用户借助鼠标键盘来设置,由此通过设置整个文档的采集率可使终端用户定义比较文档的准确性。

[0062] 终端用户设置的采集率可以存储在存储单元 402 中,该存储单元 402 如图 3 中所示的主存储器 304。

[0063] 采集单元 403、比较单元 404 以及定位器 405 可以是例如图 2 和图 3 中处理器 202、204 或处理器 302 中的功能模块。

[0064] 采集单元 403 与用户配置中心 401 以及存储单元 402 相连, 用于按照用户配置中心 401 提供的设置采集文档各点, 并从当前屏幕获得上下文, 然后将上述各点和屏幕上下文信息发送至存储单元 402。

[0065] 比较单元 404 用于判断设置书签的文档显示仍然有效, 具体地从存储单元 402 检索 / 取出设置了书签的该文档各点, 通过使之与抽取实际文档的内容相比较来识别该书签是否有效。

[0066] 定位器 405 根据比较单元的比较结果选定所下载的资源的特定部分并在显示窗口中显示所选定的资源特定部分。

[0067] 图 5 示出了用于根据本发明的优选实施例生成书签的方法的流程图。该方法可以由图 4 所示的对资源设置书签的装置实现。

[0068] 该方法首先接收用户对资源设置书签的请求 (步骤 501)。优选地, 用户对资源请求设置书签的同时可以输入请求设置的采集率, 以便于步骤 502 中的采集。用户通过用户配置中心 401 提供的采集率可以存储到存储单元 402, 也可以直接传送到采集单元 403 进行采集, 实现针对不同文档定制不同的采集率。

[0069] 响应该书签设置请求, 采集单元 403 开始采集整个文档信息 (步骤 502)。在此以 Internet 或 Intranet 网页为资源对象, 当然资源对象也可以是网络上或者同一计算机内诸如 Word、Excel、PowerPoint、Access 等 Office 文档或者 PDF 文档等。

[0070] 在步骤 502, 采集单元 403 首先从网页抽取实际文本, 即不抽取 HTML 标签以及一些图片和一些多媒体数据。

[0071] 随后采集单元 403 从存储单元 402 或者采集单元 403 获得采集率, 并根据该采集率开始采集各点, 从而形成该文档的摘要。或者, 由于摘要编写器是目前的成熟技术, 在此本发明还可以使用文本编辑器中的摘要编写器工具来编写摘要。在此情况下, 用作摘要编写器工具的采集单元对文档进行分析并为每个句子指定分数, 以此确定文档中的要点。那些包含文档中常用词汇的文本内容 (例如句子) 得分较高。然后由用户按百分比选择部分得分最高的句子, 将其显示在摘要中。在此编写摘要器时使用的“百分比”与采集率的含义一样, 表示“相当于原长的百分比”。

[0072] 编写摘要的目的是为了在对文档设置书签后再访问该文档时通过模糊匹配识别该文档是否仍然是同一个文档, 即为了识别再访问的文档是否为有效文档。该步骤 502 是优选步骤, 本发明的方法可以在不编写摘要的情况下直接使用文本定位器元件分析资源对象, 获得屏幕上下文信息。

[0073] 随后在步骤 503, 采集单元 403 利用文本定位器元件描述当前屏幕的上下文。屏幕上下文信息记录了当前屏幕的信息, 包括屏幕中的第一个词、某个特定词或者第一段等中的至少一个文本定位器元件。在此实例中, 使用屏幕中的第一个词作为文本定位器元件, 步骤 503 的处理记录第一个词例如“< complexType”, 另外从第一个词例如“< complexType”开始在当前屏幕上下文中任意拾取其它词, 其中的拾取率是由终端用户在用户配置中心 401 定义的采集率。

[0074] 对于不同的文档屏幕上下文信息可能不同, 例如对于 html(超文本链接标示语言) 文档, 屏幕上下文信息还可以具有标题信息。

[0075] 附加地, 行号、段落号以及屏幕分辨率也可以作为屏幕上下文信息的一部分, 辅助

上述与实际内容相关的可检索的文本定位器元件来描述当前屏幕的上下文。另外，在屏幕有图片、尤其是图片可能覆盖整个屏幕的情况下，除了纯文本信息之外，屏幕上下文信息还可以包括文本布置和显示的方式，在此采集单元获取的文本布置和显示方式包括图片信息例如图片信息源。

[0076] 最后在步骤 504，采集单元 403 将设置了书签的资源的地址信息、摘要信息和屏幕上下文信息传送到存储单元 402 进行记录。设置书签的流程结束。

[0077] 图 6 示出了用于根据本发明的优选实施例，响应对书签的选择呈现资源的方法的流程图。

[0078] 首先在步骤 601 用户请求重新打开设置了书签的文档。响应用户定位该书签的请求，在步骤 602 中比较单元 404 首先根据书签中指示的超链接地址抽取文档的实际文本。书签中包括的超链接地址例如是通用资源标识符 (URI) 的路径信息，在本实施例中，针对网页，URI 可以是统一资源定位符 (URL) 地址，用于指定协议（如 HTTP 或 FTP）以及对象、文档、万维网网页或其他目标在 Internet 或 Intranet 上的位置，也可以是 UNC 路径（通用命名规则路径），用于对文件的地址进行定位。如果针对同一客户机内的其它文档作为书签对象，则书签中包括的路径信息为 UNC 路径信息。

[0079] 在步骤 603，比较单元 404 从存储单元 402 检索 / 取出设置了书签的文档的摘要信息。随后在步骤 604 比较单元 404 根据从存储单元 402 检索 / 取出的采集率编制该文档的摘要，并且比较该实际文档的摘要与从存储单元 402 取出的已记录的摘要，以便识别这两个摘要是否一致，从而确定该文档是否是有效文档。与步骤 502 一样，在步骤 603 中比较单元 404 还可以使用文本编辑器中的摘要编写器工具直接编写摘要。并且在不需要判断再访问的文档是否有效的情况下或者设置书签时未编写摘要信息的情况下，在本发明响应对书签的选择呈现资源的方法中，步骤 603 和步骤 604 都是可选的步骤，本发明的方法可以在不判断再访问的文档是否有效的情况下即省略步骤 603、604 以及步骤 605，直接执行步骤 606 的处理。

[0080] 如果在步骤 605 判断为该文档并非有效文档即文档已过期，则流程结束。

[0081] 如果文档仍然有效，则比较单元 404 在步骤 606 采用已获取的书签中包括的屏幕上下文信息针对实际文档生成瞬态 (snapshot) 模拟屏幕。例如如果所获取的书签中的屏幕上下文信息以第一个词“< complexType”作为与实际内容相关的可检索的文本定位器元件，则比较单元 404 在实际文档中查找“< complexType”，在发现词“< complexType”后从该词“< complexType”开始形成瞬态模拟屏幕。随后，比较单元 404 从按照与采集单元 403 拾取一样的方式从瞬态模拟屏幕拾取词，即使用相同的与实际内容相关的可检索文本定位器元件如第一个词、某个特定词或者第一段以及相同的采集率来拾取瞬态模拟屏幕中的词；并且比较拾取的这些词与采集单元 403 已采集的作为屏幕上下文信息存储在存储单元中的各词。如果他们匹配，则利用定位器在显示窗口中显示该屏幕，完成书签定位。这里的匹配取决于模糊匹配算法，无需 100% 匹配。如果它们不匹配，则返回步骤 606 重新生成瞬态模拟屏幕进行匹配，直到到达文档结尾。

[0082] 在书签的屏幕上下文信息还包括行号、段落号、屏幕分辨率和 / 或文本布置和显示方式的情况下，比较单元 403 结合采用这些辅助信息来匹配瞬态模拟屏幕和已存储的屏幕上下文信息。

[0083] 优选的,在屏幕有图片的情况下,比较单元可以通过根据屏幕上上下文信息中的图片信息判断图片的大小和形状,不必在下载实际文档或形成瞬态模拟屏幕的过程中下载图片。若最终选定的瞬态模拟屏幕包括有图片,则可以继续下载的内容。在此,根据具体的要求可允许用户或不允许用户上下滚动查阅页面的其余部分,或者当用户上下滚动查阅页面的其余部分时,选定屏幕之外的图片可以在继续下载之后显示在屏幕上。

[0084] 尽管已经结合实施例具体说明了由本发明的发明人所做出的发明,但是本发明不限于这样的实施例,在不脱离本发明主旨的情况下,可以进行各种修改。例如,在本发明的实施例中,以显示窗口即当前屏幕的第一个词描述屏幕上上下文信息,即以当前屏幕的第一个词开始采集文本内容,获得屏幕上上下文信息。当然还可以使用其它某个特定词如第二个词、最后一个词或者第一段等等作为文本定位器元件。在本发明公开的范围内,本领域的普通技术人员可以设想到其它获取反映显示屏文档内容和结构的技术方案。

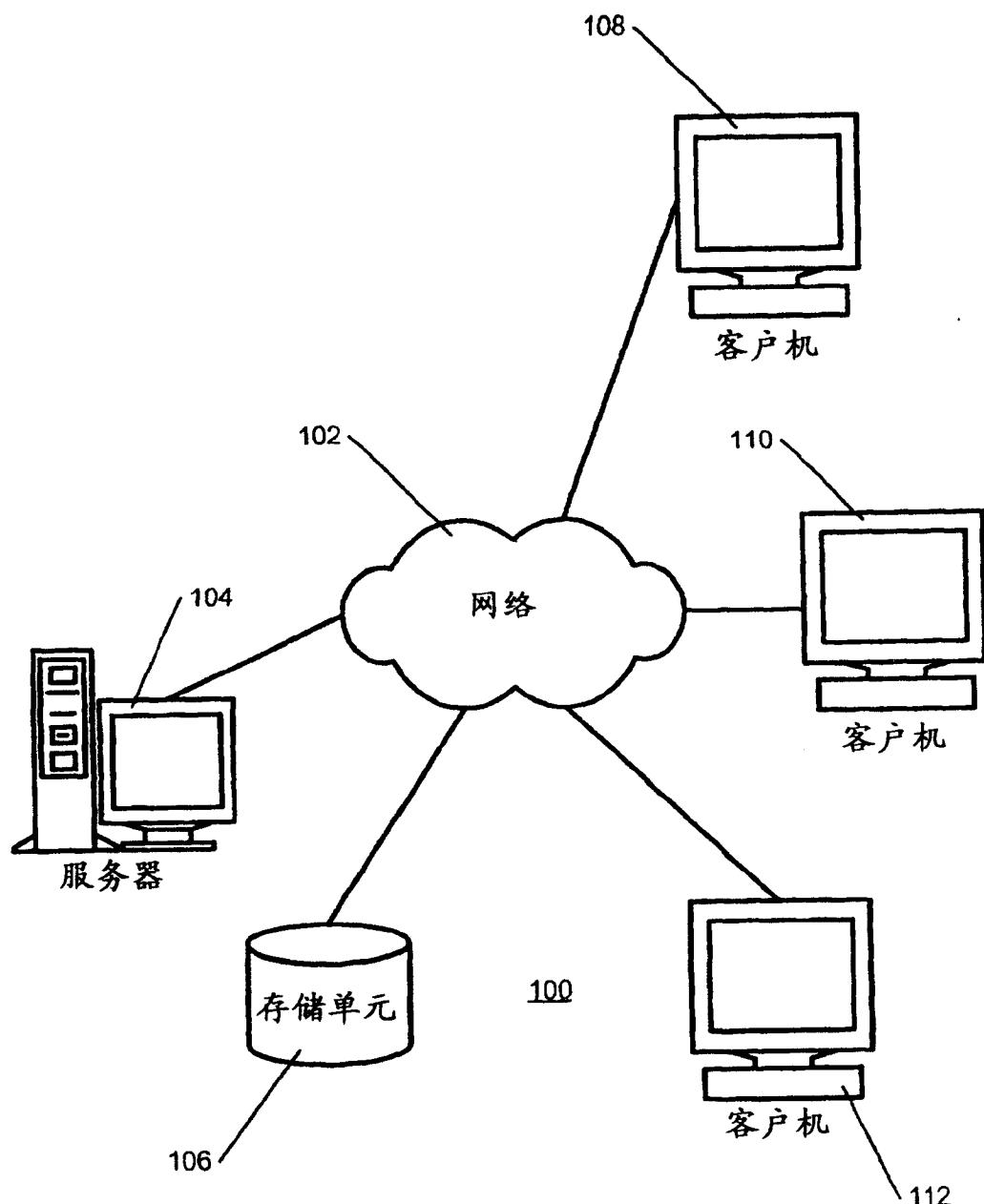


图 1

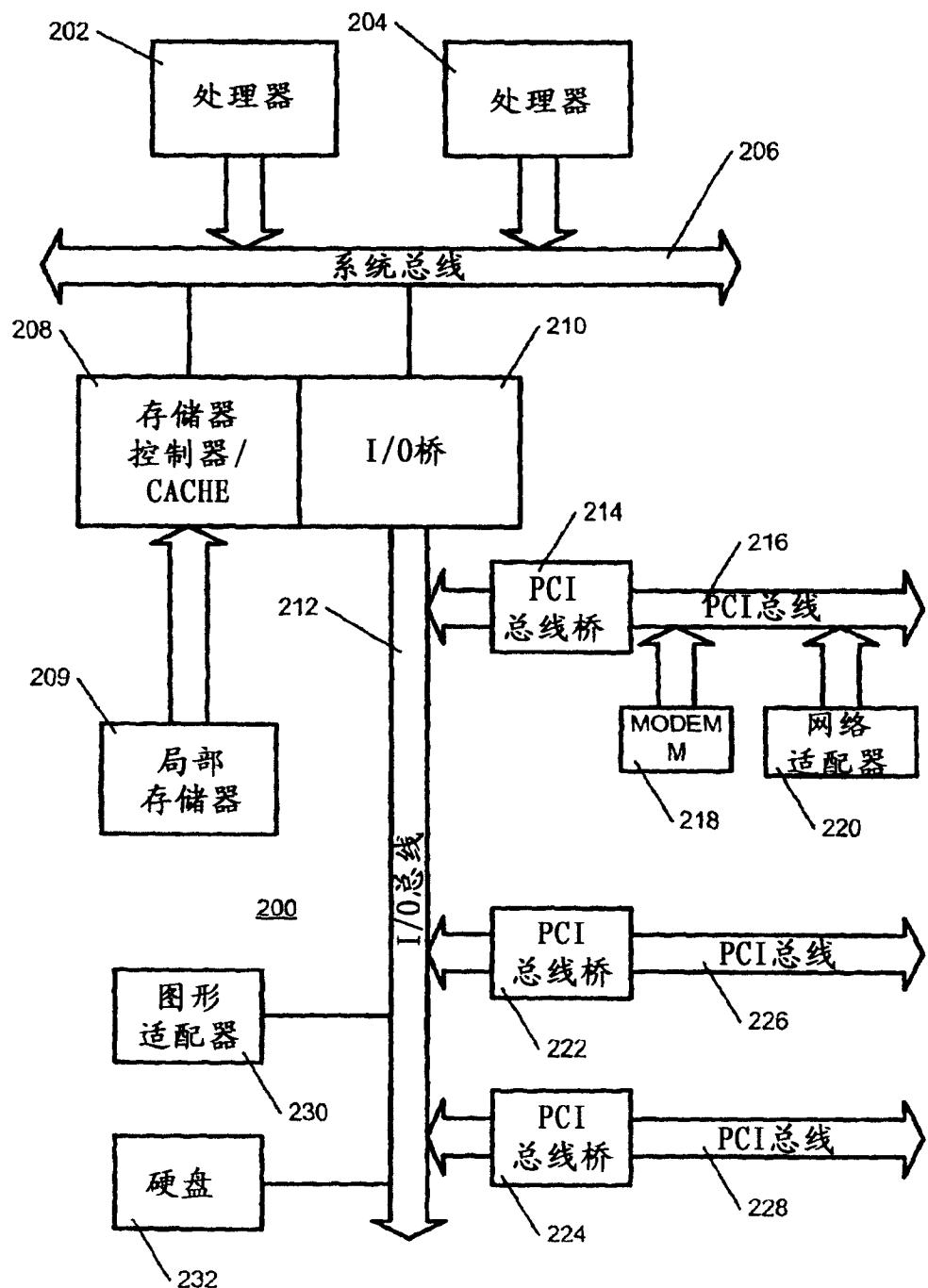


图 2

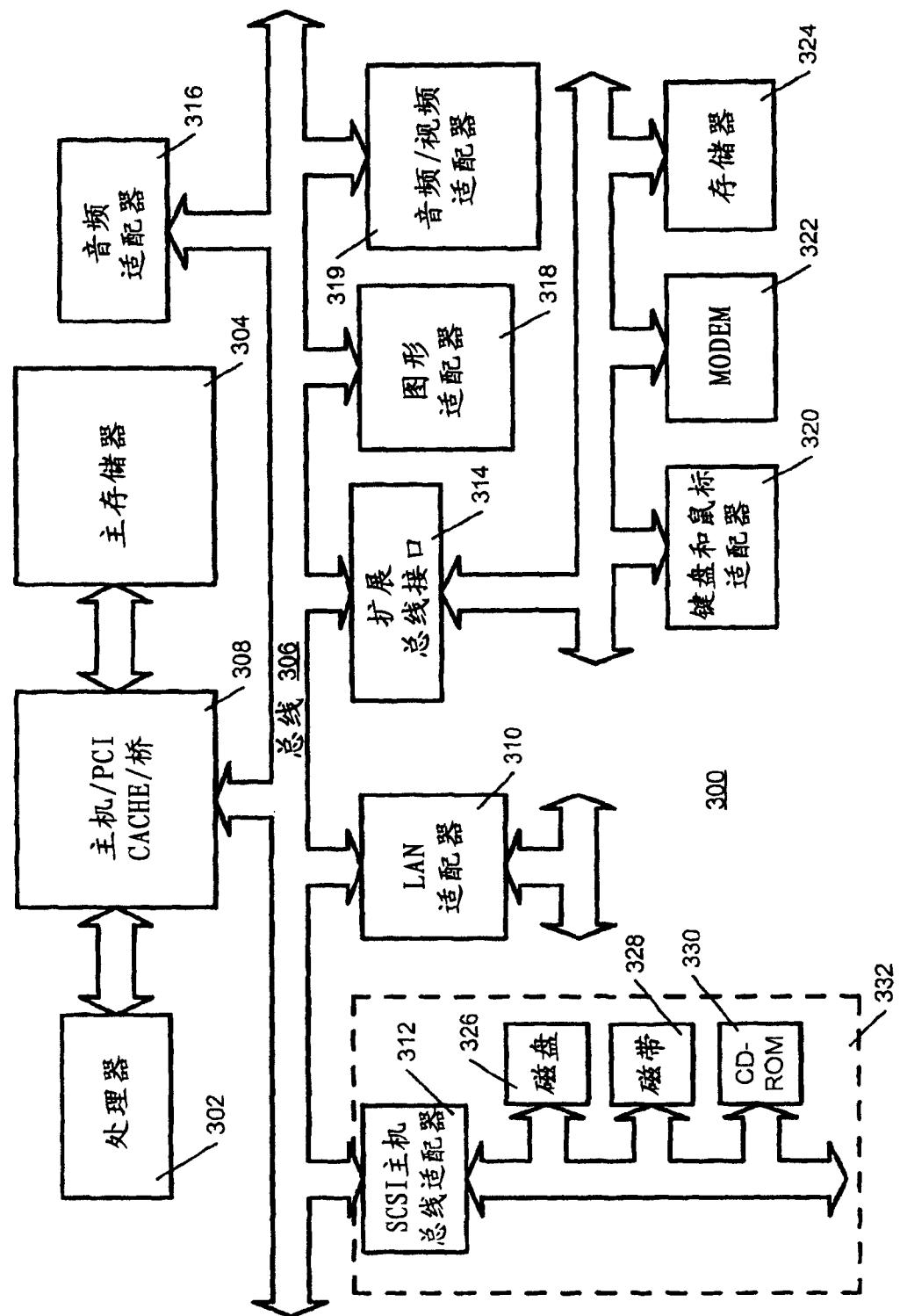


图 3

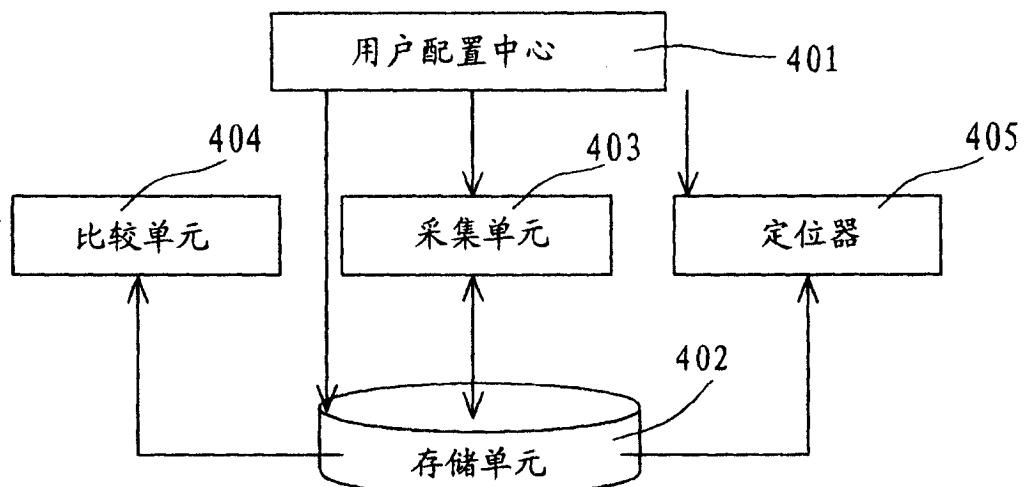


图 4

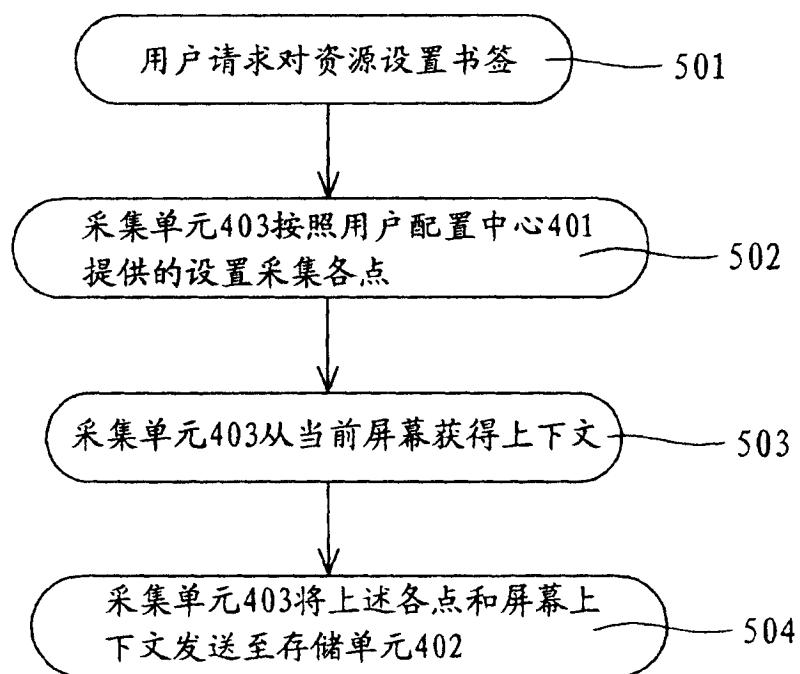


图 5

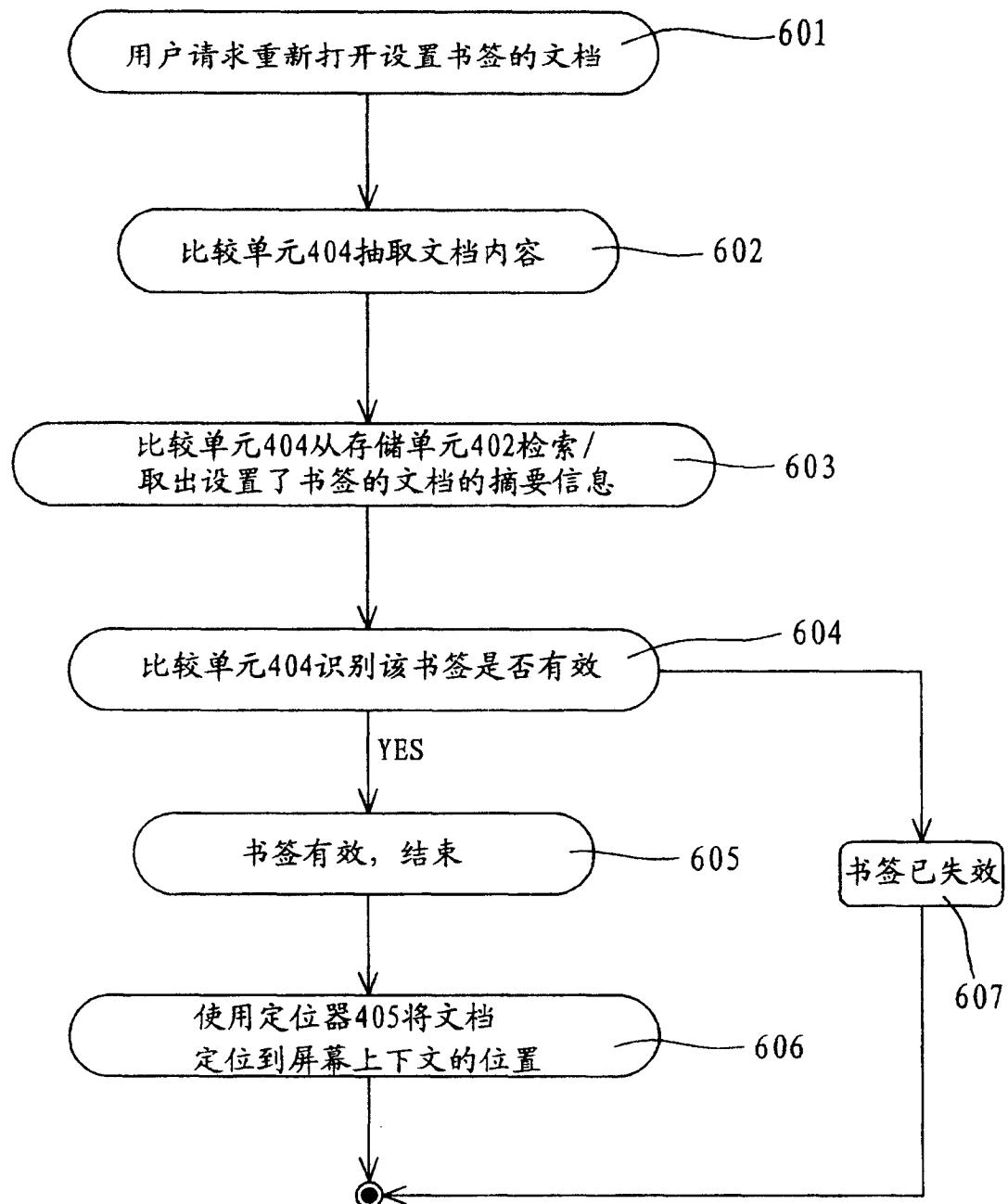


图 6