

[12] 发明专利申请公开说明书

[21] 申请号 99127440.7

[43]公开日 2000年9月13日

[11]公开号 CN 1266246A

[22]申请日 1999.12.30 [21]申请号 99127440.7

[30]优先权

[32]1999.3.3 [33]JP [31]055031/1999

[32]1999.9.21 [33]JP [31]266929/1999

[71]申请人 富士通株式会社

地址 日本神奈川

[72]发明人 岩山尚美

[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

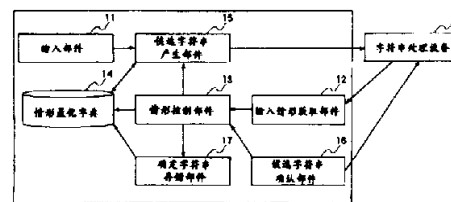
代理人 于 静

权利要求书 3 页 说明书 20 页 附图页数 24 页

[54]发明名称 输入字符串的设备和方法

[57]摘要

本发明提供了一种用于字符串输入的设备和方法，其中根据用户输入时的情形动态更新一个字典。当输入一个字符串时，获取该字符串的输入情形。根据所获取的输入情形，对用于候选字符串产生的一个字典或这种字典的一部分进行确定，以作为情形最优字典。通过使用这个情形最优字典，根据已输入的字符串的输入情形产生并输出最优输出候选字符串，并在对所输出的候选字符串进行确认和输出时，根据所获取的情形在该情形最优字典中存储所确定的字符串。



ISSN 1008-4274

权 利 要 求 书

1. 一种用于输入字符串的设备，包括：

一个用于输入字符串的输入部件；

一个用于获取输入字符串时情形的输入情形获取部件；

一个情形控制部件，用于根据输入情形获取部件得到的情形来确定用于产生候选字符串的字典或该字典某一部分并将其指定为情形最优字典；

一个候选字符串产生部件，用于根据用输入部件输入的字符串，使用由情形控制部件指定的情形最优字典产生并输出基于情形进行优化的候选输出字符串；

一个用于确定所输出的候选字符串的候选字符串确定处理部件；

以及

一个确定字符串存储部件，用于在由情形控制部件指定的情形最优字典中存储通过确定处理部件确定的字符串。

2. 根据权利要求 1 用于输入字符串的设备，其中由输入情形获取部件获取的情形至少包括一条从包含以下信息的信息组中选择的信息：

同给出输出候选字符串的字符串处理设备相关的信息；

同给出输出候选字符串的字符串处理设备可以输出的文本相关的信息；

同给出输出候选字符串的字符串处理设备可以输出的文本中的位置相关的信息；

同对输出候选字符串进行处理的字符串处理设备的处理模式相关的信息。

3. 根据权利要求 1 用于输入字符串的设备，其中情形控制部件根据输入情形获取部件所得到的情形，从多个字典中选择情形最优字典。

4. 根据权利要求 1 用于输入字符串的设备，其中情形控制部件根据输入情形获取部件所得到的情形，对从多个字典中选择情形最优字

典的方法进行更新。

5. 根据权利要求 1 用于输入字符串的设备，其中

确定字符串存储部件在对确定字符串进行存储时，将确定字符串的存储日期作为最后访问日期进行存储；

已存储的字符串的访问日期用于更新最后访问日期；并且

在根据已存储的访问日期和当前日期计算出字符串没被使用的时段后，一段时间没被使用的字符串会从情形最优字典中删除。

6. 根据权利要求 1 用于输入字符串的设备，其中确定字符串存储部件将一个经确定的字符串划分为多个必要的可优化的限制较低的单元。

7. 根据权利要求 1 用于输入字符串的设备，进一步包括一个情形最优字典产生部件，用于通过对同所确定的字符串方式相同的预先存在的电子文本中所使用的字符串进行处理来产生一个情形最优字典。

8. 一种用于输入字符串的方法包括：

输入一个字符串；

获取输入字符串的情形；

根据所获取的情形确定用于产生候选字符串的字典或该字典的一部分，并将其指定为情形最优字典；

根据所输入的字符串，通过使用所指定的情形最优字典产生并输出情形最优的输出候选字符串；

确定所输出的候选字符串；并

在一个情形最优字典中存储所确定的字符串。

9. 一个存储了在计算机上待执行的程序的计算机可读记录介质，该程序包括了以下步骤：

输入一个字符串；

获取输入字符串时的情形；

根据所获取的情形确定用于产生候选字符串的字典或该字典的一部分，并将其指定为情形最优字典；

根据所输入的字符串，通过使用所指定的情形最优字典产生并输

出情形最优的输出候选字符串;

确定所输出的候选字符串; 并

在一个情形最优字典中存储所确定的字符串。

输入字符串的设备和方法

本发明涉及一种用于输入字符串的设备和方法。特别地，本发明涉及一种用于输入字符串的设备和方法，其中有多于一个输出候选字符串同输入对应。

将字符串输入计算机的通常方法是用键盘输入字符。但是，在使用如日文这种字符丰富的语言的情况下，一般不可能用多个键实现同所有不同字符的对应。只能在键盘上键入“假名”字符串或拉丁文字符串（写为“罗马字”），并且字符通过假名-汉字转换来输入。

为了减少必须键入的字符数目，在已提出的方法中，可以使用输入预测功能来进行字符输入，可以不用键入整个字符串就可以输入它们。另外，已提出的方法可以不使用键盘而输入字符。使用这些方法，可以通过使用如手写字符识别或语言识别技术来输入字符。进而，已提出的用于输入的方法可以仅使用便携终端，如便携电话的数字键盘来输入字符。

使用这些输入字符的方法，不可能清楚地判断出来自输入方的字符串，以便通过引用合适的字典，从作为输出候选的多个字符串中确定正确的字符串。

例如，在特开昭 61-32186 号公报揭示了一种使用个体模式识别（individual pattern recognition）和上下文处理的可供每个用户定制的字典的方法。通过使用一个为每个用户特制的字典，可以增强字符串识别的精度。

特开平 9-179859 号公报揭示了一种方法，其中预先准备了几种转换方法，根据字符串处理设备使用时的情形通过选择合适的转换模式来确定候选字符串。

进而，特开平 11-3331 号公报揭示了一种方法，其中文本域是根据文本中正在输入的字符串来确定的，并且候选字符串使用同该域对

应的字典来确定。

这样，同输入对应的字符串通过引用一个字典来确定，并且根据字典可以有微小的不同。例如，如果仅允许指定域的输入，可以产生一个字典，其内容与用户的输入匹配。

但是，如果不指定输入，并且字典的内容与用户的输入不匹配，则存在一个问题，即候选字符的选择效率不会提高，并且输入字符串的命中率也不会提高。例如，即使预先准备了一个字典，它也不能适应用户新生成的缩写，新产品名或其他适当的名词等。

另外，可以想象，可以对字典进行填充以充分适应未指定的输入，但是这样会增加字典的大小，从而消耗宝贵的计算机资源，这样存在着一个问题，即在小型终端，如 PDA（个人数字助理）上联机笔（online pen）输入的手写字符识别一般不能很好地适应要求。

例如，在特开昭 61-32186 号公报中揭示的方法尝试通过为每个用户提供经定制的字典作为上下文敏感（context-sensitive）字典来提高上下文敏感性，但是它们的缺点是不能灵活适应使用情形的变化。

在特开平 9-179859 号公报中揭示的方法尝试通过选择一种在确定候选字符串时的情形下的转换模式来提高候选字符串的命中率。但是，它仅可能从预先准备好的转换模式中进行选择，这种方法不能适应不可预测的情形。另外，转换模式的选择留给了用户，并且不能保证用户会选择正确的转换模式。

进而，可以想象，可以存储许多转换模式以适应多种情形，但是这样会消耗宝贵的计算机资源，因此一般不可能在小型终端上实现。

进而，在特开平 11-3331 号公报中揭示的方法中，可以指定当前编辑文本的域，并选择对应指定域的一个字典，但是这里域本身必须预先指定，并且在用如一个便携电话上输入一个字符串以产生一条短信息的情形下，通常难于指定这样的域。另外，为了指定这样的域，同域对应的字典必须有足够的空间，这将消耗宝贵的计算机资源，因此一般不可能在小型终端上实现。

本发明的目标是解决前面技术的问题，并且提供了一种用于输入

字符串的设备和方法，其中字典根据用户的输入情形动态更新，并且可以输入同用户的输入情形相匹配的字符串。

为了达到这些目标，输入字符串的设备包括一个用于输入字符串的输入部件；一个用于获取输入字符串情形的输入情形获取部件；一个根据输入情形获取部件获取的情形确定用于产生候选字符串的字典或这种字典的一部分，并将其指定为情形最优字典的情形控制部件；一个候选字符串产生部件，用于根据用输入部件输入的字符串，使用由情形控制部件指定的情形最优字典产生并输出基于情形进行优化的候选输出字符串；一个用于确定所输出的候选字符串的候选字符串确定处理部件；以及一个确定字符串存储部件，用于在由情形控制部件指定的情形最优字典中存储通过确定处理部件确定的字符串。

使用这种配置，用户确定作为输入字符串的字符串可以存储在一个情形最优字典中以适应输入发生时的情形，因此根据用户输入情形可以更新情形最优字典，并且可以输入最适宜输入情形的字符串。

进而，在本发明用于输入字符串的设备中，由输入情形获取部件得到的情形最好从包含以下信息的信息组中至少选择一条信息：同输出候选字符串的字符串处理设备相关的信息；同输出候选字符串的字符串处理设备可以输出的文本相关的信息；同输出候选字符串所在的字符串处理设备可以输出的文本中的位置相关的信息；同对输出候选字符串进行处理的字符串处理设备的处理模式相关的信息。

由于根据字符串处理设备的差异，待输出的字符串可以限定在一定程度上内，这依赖于是否输入了一个时间表，还是输入了一个电子邮件等，因此同输出候选字符串所在的字符串处理设备相关的信息是有用的。由于待输出的字符串根据可以输出的文本也可以限定在一定程度上内，因此同输出候选字符串所在的字符串处理设备可以输出的文本相关的信息是有用的。同样，由于待输出的字符串根据可以输出的文本，根据待输入的域也可以限定在一定程度上内，因此同输出候选字符串的字符串处理设备可以输出的文本中的位置相关的信息是有用的。并且，由于根据处理模式是否为如搜索模式或存储模式，待输出的字符串也

可以限定在一定程度上内，因此同对输出候选字符串进行处理的字符串处理设备的处理模式相关的信息是有用的。

进而，在本发明用于输入字符串的设备中，情形控制部件最好根据输入情形获取部件得到的情形从多个字典中选择情形最优字典。当有一个以上的情形最优字典时这将十分有用，因为接下去可以选择一个最适合情形的情形最优字典。

进而，在本发明用于输入字符串的设备中，情形控制部件最好能根据输入情形获取部件所得到的情形，对从多个字典中选择情形最优字典的方法进行更新，因为这种方法可以选择一种最佳的候选字符串产生方法。

进而，在本发明用于输入字符串的设备中，确定字符串存储部件最好在对确定字符串进行存储时，将确定字符串的存储日期作为最后访问日期进行存储；已存储的字符串的访问日期用于更新最后访问日期；并且在根据已存储的访问日期和当前日期计算出字符串没被使用的时段后，一段时间没被使用的字符串会从情形最优字典中删除。在这种配置下，一段时间没有使用的字符串在用户认为没有用处后将被删除，这样情形最优字典不会过度增长。

进而，在本发明用于输入字符串的设备中，确定字符串存储部件最好将一个经确定的字符串划分为多个必要的可优化的限制较低的单元。这样，通过缩短字符串，可以提高命中率，并且确认所必需的处理时间也可以减少。

进而，本发明用于输入字符串的设备最好进而包括一个情形最优字典产生部件，用于对同所确定的字符串方式相同的预先存在的电子文本中所使用的字符串进行处理来产生一个情形最优字典。这样，即使当字符串输入设备首次使用时，输入字符串可以以最小的可靠性进行输入。

一个计算机可读记录介质存储了在计算机上待执行的程序，该程序包括了以下步骤：输入一个字符串；获取输入字符串时的情形；根据所获取的情形确定用于产生候选字符串的字典或该字典的一部分，

并将其指定为情形最优字典；根据所输入的字符串，通过使用所指定的情形最优字典产生并输出情形最优的输出候选字符串；确定所输出的候选字符串；并在一个情形最优字典中存储所确定的字符串。

在这种配置下，用户所确定的作为输入字符串的字符串可以在输入发生时通过装载并执行计算机上的程序存储在适合当时情形的情形最优字典中。这样，可以更新基于用户输入情形的情形最优字典，并且可以输入最适宜输入情形的字符串。

图 1 是根据本发明字符串输入设备的第一实施例的图。

图 2 示出了根据本发明字符串输入设备的第一实施例中一个情形最优字典的内容的例子。

图 3 示出了根据本发明字符串输入设备的第一实施例中一个情形最优字典的内容的例子。

图 4 示出了根据本发明字符串输入设备的第一实施例中一个情形最优字典的内容的例子。

图 5 示出了根据本发明在模式识别设备的一个实施例中一个情形最优字典的内容的例子。

图 6 是一张根据本发明字符串输入设备的第一实施例的流程图。

图 7 是一张根据本发明一个模式识别设备的图。

图 8 示出了一种手写字符输入模式的一个例子。

图 9 示出了可识别候选字符的例子。

图 10 是一张在识别操作开始时的处理流程图。

图 11 示出了根据本发明在一个模式识别设备的一个实施例中一个情形最优字典的内容的例子。

图 12 示出了一个对已识别出的候选字符串进行估计值计算的例子。

图 13 示出了一个手写字符输入模式的例子。

图 14 示出了根据本发明在一个模式识别设备的一个实施例中一个情形最优字典的内容的例子。

图 15 示出了已识别的候选字符的例子。

图 16 示出了一个对已识别出的候选字符串进行估计值计算的例子。

图 17 示出了根据本发明在一个模式识别设备的一个实施例中一个情形最优字典的内容的例子。

图 18 示出了一个对已识别出的候选字符串进行估计值计算的例子。

图 19 示出了一个对已识别出的候选字符串进行校正处理的例子。

图 20 示出了一个对已识别出的候选字符串进行校正处理的例子。

图 21 示出了一个对已识别出的候选字符串进行校正处理的例子。

图 22 示出了一个对已识别出的候选字符串进行校正处理的例子。

图 23 是一张根据本发明模式识别设备的一个实施例的流程图。

图 24 示出了记录介质的例子。

第一实施例

下面是参照附图对根据本发明的字符串输入设备的第一实施例进行的描述。图 1 是一张根据本发明字符串输入设备的第一实施例的方框图，它可以用于，如便携电话上。

在图 1 中，数字 11 表示一个隶属于便携电话的数字按键的输入部件，数字 12 表示一个输入情形获取部件，数字 13 表示一个情形控制部件，数字 14 表示一个情形最优字典，数字 15 表示一个候选字符串产生部件，数字 16 表示一个候选字符串确认部件，数字 17 表示一个确定字符串存储部件，及数字 2 表示一个用于输出候选字符串的字符串处理设备。

在图 1 中，待输入到字符串处理设备 2 的一个字符串在输入部件 11 输入。在第一实施例中所输入的是一个信号串，它通过按下数字键产生，也就是说，是一个数字串。所输入的数字串被发送到候选字符串产生部件 15。

日语使用一种中文文字的复杂书写系统（下面也称之为“日本汉字”）和两种语音音节字母（下面也称之为“假名”），每种语音音节字母包含 10 行，如表 1 所示：

a-行	ka-行	sa-行	ta-行	na-行	ha-行	ma-行	ya-行	ra-行	wa-行
a	ka	sa	ta	na	ha	ma	ya	ra	wa
i	ki	shi	chi	ni	hi	mi	-	ri	-
u	ku	su	tsu	nu	fu	mu	yu	ru	n
e	ke	se	te	ne	he	me	-	re	-
o	ko	so	to	no	ho	mo	yo	ro	wo

包括其他辅音字母的音节可以使用可区分的标记来提供：例如，“ga”行上的假名字符同“ka”行上的相同，但用另外的可区分的标记进行书写。这样，假名字符“gi”，例如，记为“ki”加可区分标记。第一实施例不需要区分带有区分标记的字符，并且对于带区分标记的字符和不带区分标记的字符的输入处理相同。因此，假名字符“gi”的输入处理，例如，同假名字符“ki”的输入处理相同。

在第一实施例中，将便携电话从 1 到 9 的数字分别分配给从“a”行到“ra”行的字符，并且将数字 0 分配给“wa”行和“n”。这样，任何日文字符串可以仅通过输入数字键来输入。

当字符串处理设备 2 已经启动并且有字符输入时，输入情形获取部件 12 从字符串所在的字符串处理设备 2 获取输入情形。这里，“输入情形”依赖于如字符串处理设备 2 的设备类型（例如，时间表输入设备，电子邮件输入设备等），输入到字符串处理设备 2 的文本类型（例如，当前编辑的邮件的“邮件地址”域），以及输入字符串处理设备 2 所处的模式（在一个电话号码处理设备中，例如，搜索模式，存储模式等）。

例如，如果一个时间表输入设备作为字符串处理设备 2 已被启动，选作字符串处理设备 2 的时间表输入设备通知输入情形获取部件 12。进而，当发生了在同时运行的电子邮件输入设备上输入的变化时，已被选作字符串处理设备 2 的电子邮件输入设备通知输入情形获取部件 12。

字符串输入设备自身也可以从字符串处理设备中获取信息，例如获取如在当前活动窗口中运行的程序的信息。

另外，当字符串处理设备 2 处于搜索模式时，正被搜索的对象也可以作为部分输入情形被获取。另外，当有一个以上的字符串输入设备并且开始输入时，也可以通过输入发生的输入方法进行获取。

所获取的输入情形被发送到情形控制部件 13。基于所获取的信息，情形控制部件 13 根据该情形假定一个候选字符串产生阶段。这里，“候选字符串产生阶段”意思是根据该情形更新情形最优字典 14 并且根据该情形更新用于产生候选字符串的方法。

作为一种根据情形更新情形最优字典 14 的方法，可能为不同情形提供不同的情形最优字典 14，并且可以在这些情形最优字典 14 之间切换。或者，不在字典间进行切换，也可能使用一种更新优先次序的方法，并提供一个域，说明在何种情形下情形最优字典 14 中的每一项都已使用并仅使用同该情形对应的域中的项，或者根据同该情形对应的域更新优先次序。

这个例子同在如便携电话或 PHS 中输入的字符相关。习惯上，在便携终端，象便携电话或 PHS 设备中输入的字符可以通过使用拨号键来完成，其中“1”对应“a”行，“2”对应“ka”行等。每个音节的元音由同一按键按下的次数确定。例如，为了输入日文“o-ku-ri-ma-su”，“1”被按下 5 次，“2”被按下 3 次，“9”被按下 2 次，“7”被按下 1 次，而“3”被按下 3 次（见表 1）。这样，为了输入 5 个日文假名字符，必须按键 14 次。

为了改善这种情形，最近已出现了经简化的输入方法“T9”。使用“T9”，便携电话或 PHS 设备的拨号键“1”对应“a”行，键“2”对应“ka”行等，这同上面相同，但它提供了一个内部语言数据库，并且根据所按下的键，对不同的所存储的词进行扫描，并可以很快选出想要的词。这样，当输入日文“o-ku-ri-ma-su”时，输入“12975”就足够了。

但是方便的“T9”可能远不完美。也就是说，短而言之，所选择的不同词的数目增长迅速，因此存在一个内在问题，即不想要的词被选择的概率增加了。

现在，第一实施例根据输入情形限制了通过“T9”选择的词。例如，图 2(a) 和 3(a) 示出了两种不同情形下情形最优字典 14 的内容。图 2 示出了用户输入一个时间表时的情形（以下称为“情形 1”），而图 3 示出了用户输入一条短信息时的情形（以下称为“情形 2”）。

在情形 1 的情形最优字典 14 中，当使用字符串处理设备 2 来输入一个时间表时，存储已输入的字符串。相似地，在情形 2 的情形最优字典中，当使用字符串处理设备 2 来输入一条短信息时，存储已输入的字符串。

在情形 1 中，当数字按键以“212”的顺序按下时，根据情形 1 的情形最优字典会产生一个候选字符串，因此候选字符串“ka-i-gi”（日文的“会议”）将输出到用于输入时间表的字符串处理设备 2。

另一方面，在情形 2 中，当数字按键以“212”的顺序按下时，根据情形 2 的情形最优字典会产生一个候选字符串，因此候选字符串“ke-i-ko”（一个日本人名）将输出到用于输入短信息的字符串处理设备 2。

如果字符串处理设备 2 处于搜索模式，则可以采用一种用于更新用于根据情形产生候选字符串的方法的方法来搜索情形获取部件 12 所得到的对象以更新情形最优字典 14 中的字符串，这样这些字符串将作为候选字符串来生成。另外，当字符串处理设备 2 处于输入模式时，如果所输入的字符串不在情形最优字典 14 中，可以产生一个不在情形最优字典 14 中的字符串作为候选字符串。

如果在第一实施例开始使用字符串输入设备时情形最优字典 14 没有实现任何字典功能，由于不可能动态更新情形最优字典 14，因此在安装中会包含一个常规标准字典。

图 4 是出了一个当用户为了打一个电话而在存储在便携电话中的电话目录中搜索一个名字时情形最优字典 14 的内容的例子。在这种情况下，情形最优字典 14 的内容与情形获取部件 12 所获取的作为电话目录的存储内容相匹配。这意味着如果数字按键以“212”的次序按下，则候选字符串“ko-i-ke”（一个日本人名）作为在电话目录名字搜索模

式下的输出。

此外，如果存在一个以上的字符串输入方法，字符串输入方法也由情形获取部件 12 获取。例如，如果识别出输入方法是手写字符，首先进行手写字符的模式识别，然后通过使用其内容已基于识别出的候选字符串更新了的情形最优字典 14 在候选字符串产生部件 15 中生成一个最优候选字符串。

此外，在输入方法依赖于数字按键的情况下，通过使用其内容已基于数字串更新了的情形最优字典 14 在候选字符串产生部件 15 中生成一个最优候选字符串。

在候选字符串产生阶段，由情形控制部件 13 对其进行设置，候选字符串生成部件 15 根据情形基于所输入的字符串产生一个最优候选字符串，并将它发送给字符串处理设备 2。

发送给字符串处理设备 2 的最优候选字符串由字符串处理设备 2 的输出部件进行输出。顺便提一句，最优候选字符串不必明确确定，可以在最优度评级后产生一个以上的最优候选字符串，用户可以用字符串处理设备 2 选择一个候选字符串。

关于由字符串处理设备 2 的输出部件输出的候选字符串，用户可以用候选字符串确定部件 16 执行一条确定命令。“确定命令”指用户给出一条可以完成的确定一个输出字符串的命令，例如，通过从多个候选字符串中选择某个候选字符串。用户也可能不进行任何特定的声明操作 (declaratory operation)，但是在用户进入到下一个输入操作时要进行确认。

当用户给出确认候选字符串的命令时，候选字符串确认部件 16 将经确认的字符串发送给情形控制部件 13。情形控制部件 13 将当前的输入情形和经确认的字符串发送给确认字符串存储部件 17。接着，确认字符串存储部件 17 根据当前输入情形用情形最优字典 14 存储所确认的字符串。

图 5 是出了一个情形最优字典 14 的例子。在图 5 中，对包括所输入的数字串，所存储的字符串，字符串最后使用的日期，以及字符串

的使用次数的存储项进行了存储。字符串的最后访问日期和字符串的访问次数在确认字符串存储部件 17 对所确认的字符串进行存储时刷新。应该注意，除了日期还可能包括字符串最后访问的时间。

根据字符串的最后访问日期，可以使用最近字符串访问的优先级或者字符串访问频率的优先级产生候选字符串。不用说可以结合这两种因素并且通过优先级权重来计算估计值以产生候选字符串。

此外，根据字符串的最后访问日期，在一段时间内没有访问的字符串将从情形最优字典 14 中删除。例如，字符串确认错误的或者偶尔访问一次可以在一段时间后删除，因此不是产生候选字符串所必需的数据可以从情形最优字典 14 中删除，并且可以对字典的容量进行经济使用。

如果在如英语中定义“空格”作为字符串分隔单元的划分词，则可以用空格分隔来存储文字单元。但是在如日语中难于对文字单元进行分隔，最好能在形态分析 (morphological analysis) 或使用一种合适的方法，如在字符类型 (例如，日文假名和日本中文字符) 变化的地方打点作为分隔之后进行文字的分隔。

作为一个例子，让我们查看一下日文词“pataanninshikisouchi”，这在英语里指“模式识别设备”。该词的“pataan”部分是用日文片假名写的，也就是两种日文语音假名字符类型之一，而该词的“ninshikisouchi”部分是用日本中文写的。当将“pataanninshikisouchi”分隔为字符串时，它可以通过形态分析分隔为“pataanninshiki” (模式识别) 和“souchi” (设备)。另一种分法是如果在字符类型变化 (这里，从假名字符“pataan”到日本中文字符“ninshikisouchi”) 的地方进行分隔，则它可以被分隔为“pataan”和“ninshikisouchi”。

情形最优字典 14 不仅受计算机资源的限制，而且也受某些物理约束。也就是说，不可能存储超过一定数量的数据。根据字符串的最后访问日期，具有最久访问日期的字符串可以从情形最优字典 14 中接连删除，因此可以持续维护情形最优字典 14 的字符串的最大存储数目或

者最大存储容量。

通过在最后访问日期中加入时间，可以对所存储的字符串进行更精细的管理。可以在一段时间过后，或者在输入一个新词并在情形最优字典 14 中存储时从情形最优字典 14 中删除字符串。

用户最好能根据计算机资源自由设置最大存储数目或最大存储容量。

也可以通过对在情形最优字典 14 中预先存在的与所确定的字符串方式相同的电子文本中使用的字符串进行处理来提供进一步的用于产生情形最优字典 14 的情形最优字符串字典产生部件。然而在这种情况下事实是，在用户第一次使用字符串输入设备的情形下，用户几乎不能期待根据这种情形产生候选字符串，根据用户第一次的使用情况在一定程度上可能产生候选字符串。

下面是参照图 5 和图 6，用于实现根据本发明的字符串输入设备的第一实施例的程序处理流程的解释。

在图 6 中，当用户输入一个字符串时（步骤 S61），获取字符串输入时的情形，即字符串处理设备 2 的设备类型信息（例如，时间表输入设备，电子邮件输入设备等），输入到字符串处理设备 2 的文本类型信息，该文本的当前位置信息（例如，在一封正在发出的邮件中的“邮件地址”域），以及字符串处理设备 2 所出的模式（在一个电话号码处理设备中，例如，搜索模式，存储模式等）信息（步骤 S62）。接着，根据所获取的情形，动态更新和引用情形最优字典 14。

根据情形通过选择和引用情形最优字典 14，产生根据该情形的一个最优候选字符串（步骤 S64）并将其作为候选字符串由字符串处理设备 2 输出（步骤 S65）。

接着，在执行完输入字符串的用户确认过程后（步骤 S66：确定），在情形最优字典 14 中存储所确认的字符串（步骤 S67），以便它可以应用到下一次处理中。

这样，在第一实施例中，根据字符串输入时的情形通过动态更新情形最优字典 14 中的内容，可以得到最适宜输入情形的内容，并且根

据使用情形最优字典 14 的情形通过进行候选字符串的生成处理，可以以较高的命中率输入字符串。

第二实施例

下面是参照附图根据本发明对字符串输入设备的第二实施例的说明。图 7 是根据本发明的字符串输入设备的第二实施例的一张方框图，它可以应用在如便携式计算机上。在这个例子中，字符串通过模式识别方式输入，如手写字符。在图 7 中，数字 70 表示一个模式输入部件，它包括一个手写输入部件，用来接受在部分显示屏上用笔写入的输入，数字 71 表示一个模式识别部件，数字 72 表示一个模式识别所用的字典，数字 73 表示一个候选字符串产生部件，数字 74 表示一个情形最优字典，数字 75 表示一个情形控制部件，可以根据情形假定候选字符串的产生状态，数字 76 表示一个输入情形获取部件，数字 78 表示一个确定字符串存储部件，而数字 79 表示一个包含有前述手写输入部件的显示部件。另外，还有一个字典管理部件 80，它是第二实施例的一个刻画特征，它包括候选字符串确认部件 77 和确认字符串存储部件 78，它还提供一个日期更新部件 81 和一个再现 (recurrence) 更新部件 82。

在图 7 中，待识别的模式通过模式输入部件 70 输入。在第一实施例中这是一个手写字符串。所输入的模式被发送到模式识别部件 71。

第二实施例通过作为手写输入的例词“shachou”进行说明。日文字“shachou”在英语中意指“公司总裁”，包含了两个中文字符，如图 8 所示。在图 8 中，左边的中文字符读作“sha”，右边的中文字符读作“chou”。当然，本发明并不局限于这种输入方法，可以使用其他的输入方法，如语音输入。图 8 示出了在指定格子中用手写方式输入字符的情况，但是对此没有具体限制，输入也可以是没有格子的输入区域中的手写字符串。

模式识别部件 71 通过参照用于模式识别字典 72 来查找所输入的模式，并将所识别的候选字符按精确性 (exactitude) 的次序输出给候选字符串产生部件 73。为了筛选所识别的候选字符，可以使用通常使

用的 OCR 引擎或联机识别引擎，或者其他类型的识别引擎。为了计算的精确性，可以使用几种方法，例如，计算所输入的模式和模式识别字典 72 中的模式之间的差距，或者根据刻画特征进行计算。图 9 是出了在图 8 所示例子，即通过手写输入字符“shachou”的情况下进行识别后的结果的一个例子。在图 9 中，列在所识别的候选字符旁边的数字表示所识别的相应候选字符的精确性，数字越高则精确性越高。

基于通过模式识别部件 71 所得到的识别结果，候选字符串产生部件 73 使用情形最优字典 74 确定最适宜所输入模式的模式识别候选字符串。

如图 10 流程图所示，情形最优字典 74 在模式识别开始时从一个外部存储器读到一个内部存储器中。首先，对所识别的候选字符串同存储在情形最优字典 74 中的已读入到内部存储器的字符串进行比较。根据所识别的候选字符串相对于情形最优字典 74 中的字符串的精确度，对所识别的候选字符串的估计值进行计算。在对所计算的估计值进行比较后，选择估计值最高的所识别的候选字符串。这样，可以在有两个以上的候选字符串的情况下确定一个经最优模式识别的候选字符串。此外，将存储在情形最优字典 74 中的最后访问日期或者访问次数考虑在内，可以加强关于所识别的候选字符串的估计值的可靠性，适宜用户的经识别的候选字符串可以作为经最优模式识别的候选字符串进行输出。

下面对估计值的计算进行说明。从模式识别部件 72 输出的一个经识别的候选字符串 W 可以通过 n 个候选字符 $C_1 \cdot C_2 \cdot C_3 \dots C_n$ （这里的 n 是一个正数）进行表示。 S_i 是识别候选字符 C_i 的精确度。如果所识别的候选字符串 W 在情形最优字典 74 中，则估计值 Z 可以表示为

$$\text{公式 1} \quad Z = \sum_{i=1}^n S_i + n \times \alpha \quad (\alpha > 0)$$

这里， α 是一个表示所识别的候选字符串 W 在情形最优字典 74 中的权重，它可以由用户进行修改。

如果所识别的候选字符串 W 不在情形最优字典 74 中，则估计值 Z 可以表示为

$$\text{公式 2} \quad Z = \sum_{i=1}^n S_i$$

通过确定公式 1 或公式 2 的估计值，即使所识别的候选字符的组合偶尔出现在情形最优字典 74 中，也可以进行包括从模式识别部件 72 得到的精确度的全面估算。这样，可以输出一个合适的字符串作为所识别的候选字符串。

如果所识别的候选字符串 W 在情形最优字典 74 中，则估计值 Z 可以表示为

$$\text{公式 3} \quad Z = \sum_{i=1}^n S_i + n \times (\alpha f(\text{freq}(W)) + \beta g(\text{passtime}(W)))$$

$(\alpha \geq 0, \beta \geq 0)$

这里， $\text{freq}(W)$ 表示存储在情形最优字典 74 中的所识别的候选字符串 W 的访问频率， $\text{passtime}(W)$ 表示从上次所识别的候选字符串 W 被访问开始已过去的时间。此外， $f(\text{freq}(W))$ 是一个访问频率的函数， $g(\text{passtime}(W))$ 是一个已过去时间的函数。通过因子 α 和 β 对函数 f 和 g 的加权计算，可以用所估算的访问频率和已过时间的加权来进行随意调整。

例如，如果如图 11 所示，所识别的候选字符串“shachou”（在日语中意思是公司总裁）和“rei-wo”（日语中意思是“谢谢”）被存储起来，并且用公式 1 以 α 值 30 计算估计值，接着如图 12 所示，通过将“sha”的精确度 748 和“chou”的精确度 759 的和同字符数量 $2 \times \alpha = 60$ 相加，得到字符串“shachou”的估计值 Z 为 1567。同样，计算出字符串“rei-wo”的估计值 Z 为 1500。结果，字符串“shachou”具有较高的估计值，以此得到了识别。

如图 13 所示，所输入的手写字符也可以是英文文本。同日文字符

串的情况类似，下面是用公式 3 对估计值进行计算的一个例子，如图 14 所示，这里在情形最优字典 74 中存储了两个所识别的英文候选字符串“hat”和“hut”。

根据图 15 所示的识别结果，对字符串“hat”来说，字母“h”的精确度 858，字母“a”的精确度 839 和字母“t”的精确度 888 的求和计算结果 2585。

这里，假设 $\alpha=5$ ， $\beta=10$ ， $f(1)=1$ ， $f(2)=2$ 。此外，执行日期是 1999 年 1 月 6 日，并且如果上次访问日期是 1999 年 1 月 3 日，则所过去的天数值为“3”，而如果上次访问日期是 1999 年 1 月 6 日则过去的天数为“0”。用这些值作为参数，可以假设 $g(0)=10$ ， $g(3)=7$ 。

如果对这种情况下的估计值进行计算，由于字符数为 3，公式 3 变为

$$2585 + 3 \times (5 \times f(1) + 10 \times g(3)) = 2585 + 3 \times (5 \times 1 + 10 \times 7) = 2810$$

相似地，如果对字符串“hut”的估计值进行计算，字母“h”的精确度 858，字母“u”的精确度 783 和字母“t”的精确度 888 的求和计算结果 2529，再加上 $3 \times (5 \times f(1) + 10 \times g(3)) = 3 \times (5 \times 1 + 10 \times 7) = 225$ ，结果是 2754。

如图 16 所示，通过对所有字符串的估计值进行计算，字符串“hat”的估计值最高，因此将作为所识别的字符串输出。

接下来，用户对所识别的结果进行正确性检查。可能出现这种情况，即输入的字符串没有当“hat”而是当“hut”来确认。

在这种情况下，如图 17 所示，当“hut”得到确认的时候情形最优字典 74 的内容发生变化。这意味着字符串“hut”的上次访问日期变为 1999 年 1 月 6 日，并且访问次数增加为“2”。

如果在这种情形下，再次按图 13 的模式进行字符输入，则识别结果如图 15 所示，按上面的方法进行估计值的计算，下面给出结果。

首先，如果字符串“hat”的估计值的计算结果同上面一样，字母“h”的精确度 858，字母“a”的精确度 839 和字母“t”的精确度 888 的求和计算结果 2585，在加上 $3 \times (5 \times f(1) + 10 \times g(3)) = 3 \times (5 \times 1 + 10 \times 7) = 225$ ，

结果为 2810。

另外，对于字符串“hut”，字母“h”的精确度 858，字母“u”的精确度 783 和字母“t”的精确度 888 的求和计算结果 2529，再加上 $3 \times (5 \times f(2) + 10 \times g(0)) = 3 \times (5 \times 2 + 10 \times 10) = 330$ ，结果为 2859。

结果，同上面一样，如果所有字符串的估计值计算结果如图 18 所示，可以确定字符串“hut”的估计值最高，因此“hut”作为所识别的字符串输出。

在候选字符串产生部件 73 中对情形最优字典 74 的引用并不局限在整个字符串都得以匹配的情况，它也可以抽取部分匹配的字符串作为所识别的候选字符串。这是因为不能保证字符串以存储在情形最优字典 74 中的字符串单元进行输入。

此外，如果候选字符串产生部件 73 对情形最优字典 74 的引用结果是这种情况，即存储在情形最优字典 74 中的字符串没有相应的字符串，则可以通过使用预先存在的词典或惯用的 N-gram 统计字典来进行候选字符串的确认处理，并将结果作为最优模式识别结果输出。

当模式识别结果输出给显示部件 79 时所识别的候选字符串由候选字符串产生部件 73 来确定。接着，如果用户判断由显示部件 79 显示的所识别的候选字符串是错误的，则用户可以用候选字符串确认部件 77 执行一个校正处理。

这里，“校正处理”指所识别的候选字符串按它们估计计算值的次序进行显示，并且用户可以选择正确的字符串作为所识别的字符串或者输入一个新的模式重新进行识别。例如，如果如图 19(a)所示，字符串“reichou”（这不是一个合法的日语文字，而是由两个日本汉字“rei”和“chou”组成）已经识别并且可能如图 19(b)所示，选择精确度高的字符“sha”。另外，如图 20(a)所示，如果如图 19 所示无意地识别了字符串“reichou”（这不是一个合法的日语文字），可以选择，例如“重新输入”按钮，此时所识别的候选字符“rei”被删除，并且通过输入一个新的手写字符串可以对其进行校正。应该注意的是对这些处理方法没有限制，只要是对用户想要输入的字符串进行正确模式识别的方法都可

以使用。

如果用户确定输出的所识别的字符串是正确的，则用户可以使用字典管理部件 80 中的候选字符串确认部件 77 进行确认处理。这里，“确认处理”指用户给出一条指令，对所识别的字符串进行确认。此外，对用户执行的所声明的确认指令的方法没有特定的限制，并且确认可以在用户接着进行下一次输入时进行而不用进行任何校正处理，或者可以在切换到另一个应用时对输出的字符串进行确认。

例如，如果手写字符串已被正确地识别为“shachou”，如图 21 所示，用户可以通过按“应用”或类似按钮以声明方式确认所识别的字符串。如果，如图 22 所示，列出并显示了包含手写字符串“shachou”的识别结果的整个句子，如果对所显示的另一个字符进行校正或者输入一个新的字符，则所显示的字符串“shachou”被作为一个经确认的所识别的字符串。

接着，如果候选字符串确认部件 77 对所识别的字符串进行确认的确认处理，则经确认的所识别的字符串由情形最优字典 74 中的确认字符串存储部件 78 进行存储。图 11 示出了一个情形最优字典 74 的例子。在图 11 中的情形最优字典 74 中，存储了字符串，字符串上次被访问的日期，以及它们被访问的次数。字符串上次被访问的日期，以及它们被访问的次数通过日期更新部件 81 和重现更新部件 82 来更新，并且通过确认字符串存储部件 78 存储在情形最优字典 74 中。除了日期，还可以包括字符串上次被访问的时间。

例如，如果候选字符串确认部件 77 通过执行图 21 所示的对所识别的字符串进行的校正处理确认字符串“shachou”为所识别的字符串，则确认字符串存储部件 78 将词“shachou”存储在情形最优字典 74 中。如果字符串“shachou”以前并没有存储在情形最优字典 74 中，则进行一次新的存储，并且如果以前已经存储了它，则上次访问日期被更新为当前日期，访问重现数增加“1”。

此外，根据上次字符串被访问的日期，也可以给最近被访问的字符串或访问重现数高的字符串以优先级的前提下进行候选字符串的确认

认处理。不用说可以结合这两种因素并且通过优先级权重来进行估计值的计算。

此外，当在情形最优字典 74 中存储由用户确定的所识别的字符串时，也可以在所有所确认的字符串中仅存储包含用户执行校正处理部分的所识别的字符串。既然可以期望下次对于具有相同模式的所识别的字符串，可以不进行校正处理而进行输入，因此该字符串可以再次得以正确地识别，没有必要在情形最优字典 74 中对其进行存储。结果，通过不在情形最优字典 74 中存储这样的所识别的字符串，存储在情形最优字典 74 中的字符串的数目得以减少，从而可以减少情形最优字典 74 的大小。

此外，如果有多个用户共享同一个终端，则可以相应提供用户特定的情形最优字典 74，例如，当用户变化时输入一个用户标识。这使得模式识别可以适应不同用户的使用。

通过上面所描述的第二实施例，同用户的访问情形相对应，情形最优字典的内容可以适应最适宜用户的实际情形，并且即使输入方式是手写方式，也可以执行具有高识别率的模式识别，并输入用户想要的字符串。

下面是参照图 23 根据本发明对用于字符串输入设备的第二实施例实现的程序的处理流程的说明。

在图 23 中，如果用户输入一个模式（步骤 S231），例如手写字符或语音输入，对所输入模式进行模式识别（步骤 S232），并且在情形最优字典中查找所识别的候选字符串（或者多个所识别的候选字符串的组合）（步骤 S233）。

如果所识别的候选字符串在情形最优字典中，则计算估计值，该计算将全面估算所识别候选字符的精确度，最后访问日期以及访问重现数（步骤 S234）。接着给出最合适的所识别的候选字符串（步骤 S235）。如果所识别的候选字符串不在情形最优字典中，则多个所识别的候选字符的组合将充当所识别的候选字符串。

接着，在用户执行完如所识别字符串的校正处理后，进行所识别

字符串的确定处理（步骤 S236）。接着，为了将结果用在以后的处理中，将所确定的所识别的字符串存储在情形最优字典中（步骤 S237）。

图 24 示出了记录介质的例子，该记录介质存储了根据本发明用于实现字符串输入设备的一个实施例的程序。如图 18 所示，这种记录介质不仅可以是便携记录介质 242，例如 CD-ROM242-1 或者软盘 242-2，也可以是通过通信电路连接的存储器设备 241，或者其他任何记录设备 244，例如计算机的硬盘或者 RAM。为了执行该程序，可以在主存储器中装载并执行该程序。

此外，如图 24 所示，存储同根据本发明字符串输入设备的一个实施例一起产生的情形最优字典的记录介质不仅可以是便携记录介质 242，例如 CD-ROM242-1 或者软盘 242-2，也可以是通过通信电路连接的存储器设备 241，或者其他任何记录设备 244，例如计算机的硬盘或者 RAM，它可以在使用例如根据本发明的字符串输入设备时读入计算机 243。

本发明可以在不偏离其精神或实质特征的前提下以其他特定形式进行使用。在本申请中所揭示的实施例在所有方面都是说明性而不是限制性的，本发明的保护范围是由所附的权利要求而不是由上面的描述来限定的，所有在同权利要求等同的范围和意义内的变更均可包含其中。

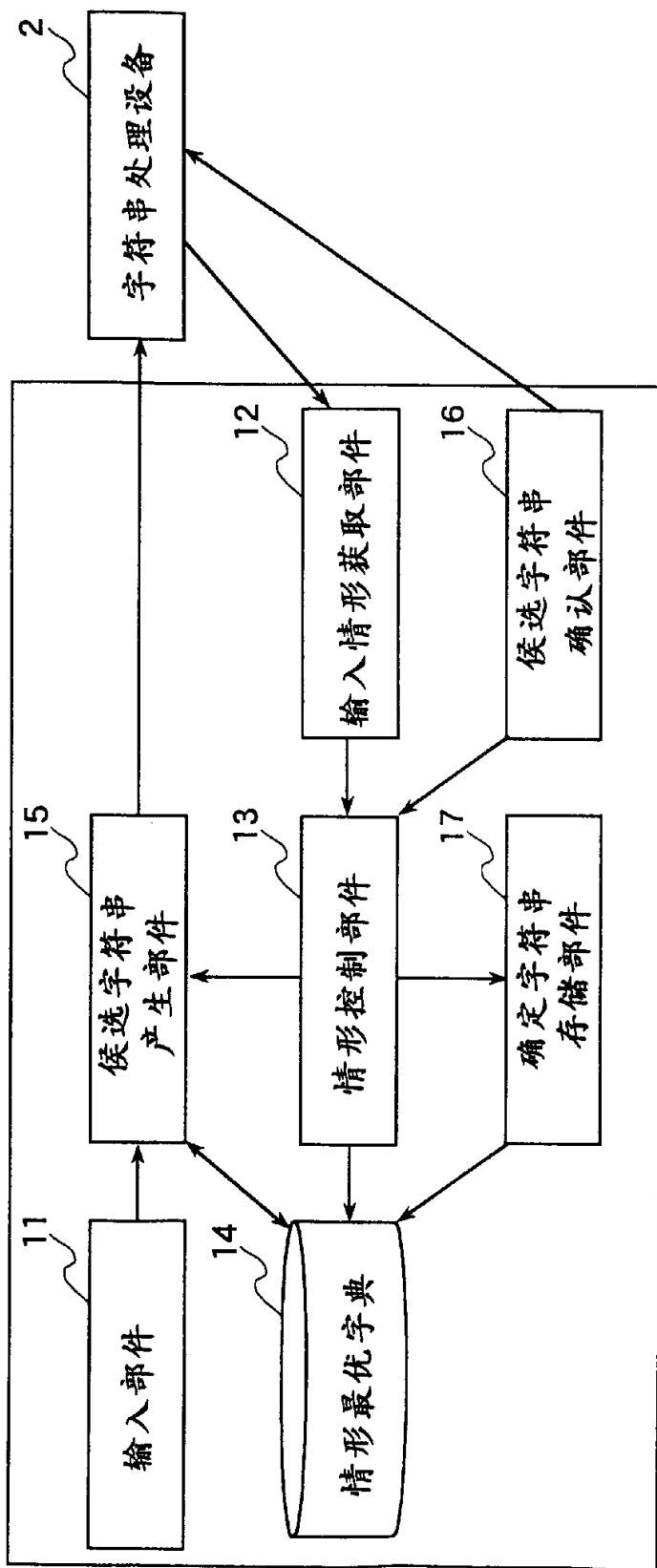


图1

入力	候选字符串
•	•
212	会議 [ka-i-gi]
•	•
•	•
•	•

(a)

时间表

7月23日 ■

会議

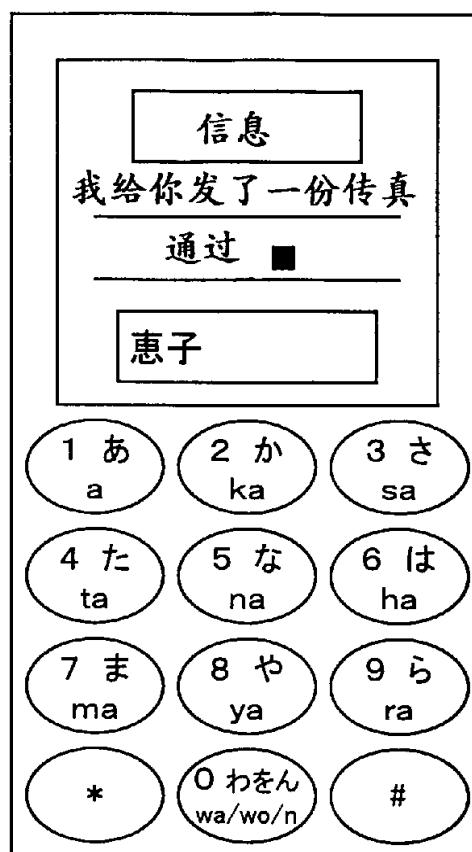
1 あ a	2 か ka	3 さ sa
4 た ta	5 な na	6 は ha
7 ま ma	8 や ya	9 ら ra
*	0 わをん wa/wo/n	#

(b)

图2

输入	候选字符串
•	•
212	惠子 [ke-i-ko]
•	•
•	•
•	•

(a)



(b)

图 3

输入	候选字符串
•	•
212	小池 [ko-i-ke]
•	•
•	•
•	•

(a)

电话簿

呼叫: 名字?

小池

1 あ a	2 か ka	3 さ sa
4 た ta	5 な na	6 は ha
7 ま ma	8 や ya	9 ら ra
*	0 わをん wa/wo/n	#

(b)

图 4

数字串	所存储的字符串	上次访问日期	访问重现数
.	.		
212	会議 [ka-i-gi]	99/9/3	1
.	.		
212	恵子 [ke-i-ko]	99/9/6	10
.	.		
.	.		

图5

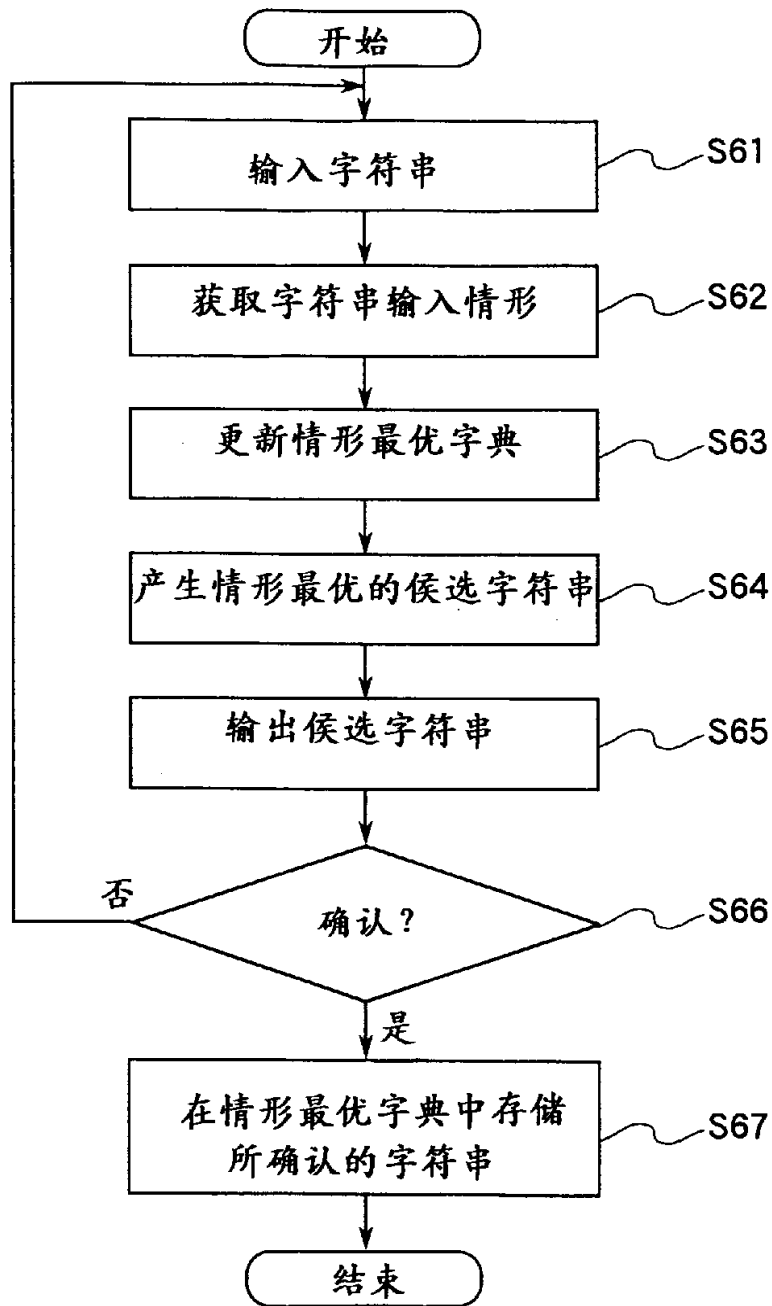


图6

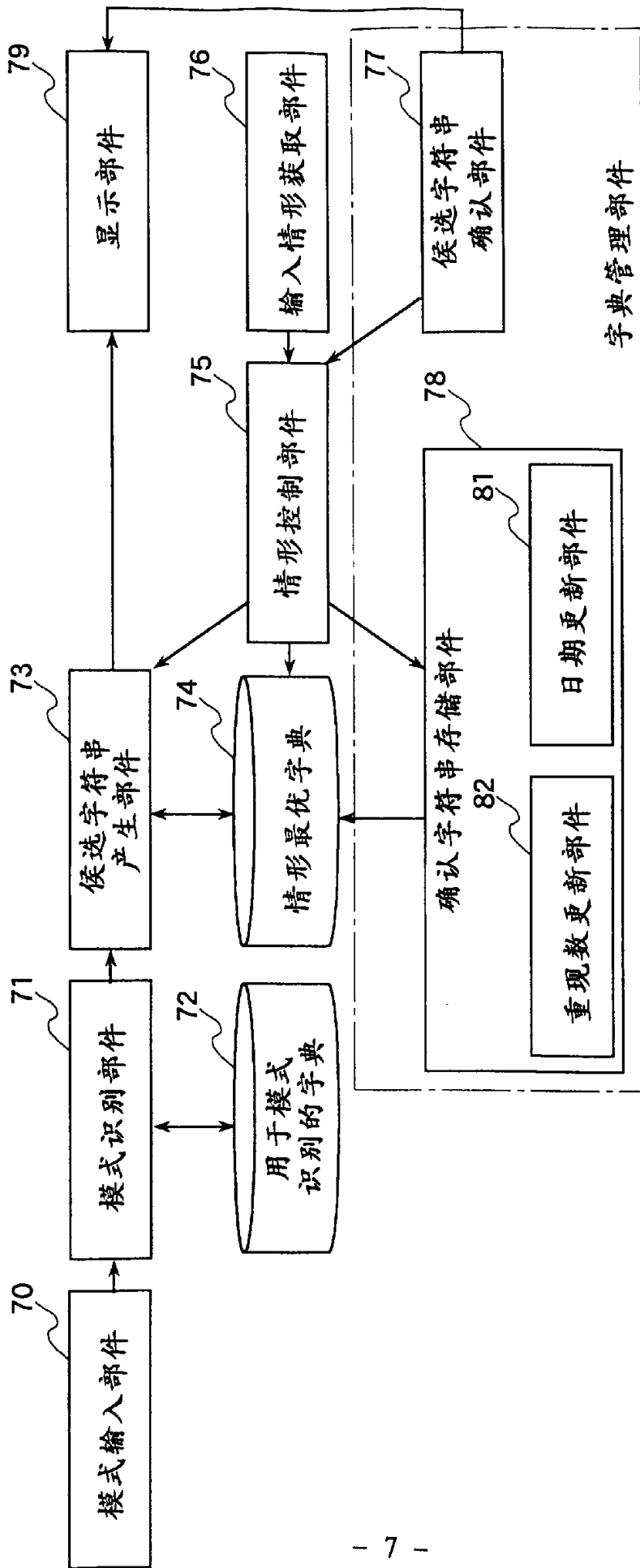


图7

社 [sha]	長 [cho]
------------	------------

图 8

(a)

礼 [rei]	797
社 [sha]	748
杠	669
杜	663
私	659

(b)

長 [cho]	759
δ	710
蒐	650
雇	648
を [wo]	643

图9

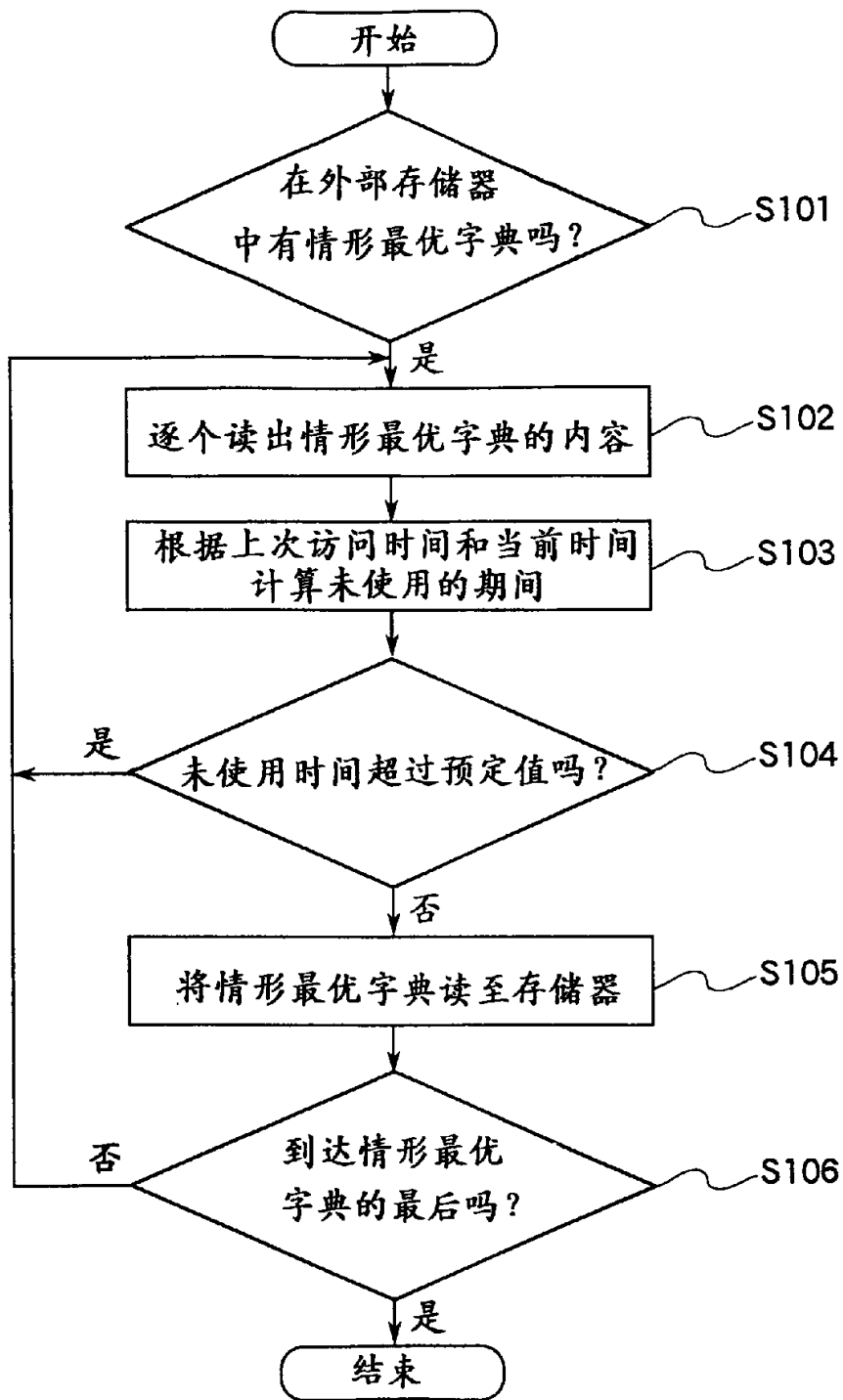


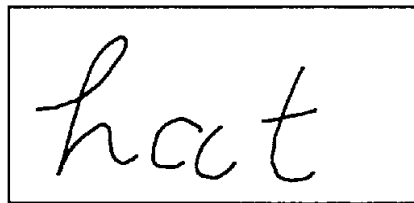
图10

所存储的字符串	上次访问日期	访问重现数
•		
礼を [rei-wo]	99/1/3	1
•		
社長 [sha-cho]	99/1/6	10
• •		

图 11

候选字符串	估计值
礼長 [rei-cho]	1556
社長 [sha-cho]	1567
•	
礼を [rei-wo]	1500
• •	

图 12



hat

图13

所登记的字符串	上次访问日期	访问重现数
.		
hat	99/1/3	1
.		
hut	99/1/3	1
.		
.		

图 14

(a)	(b)	(c)																		
<table border="1" style="border-collapse: collapse; width: 100%;"> <tr><td style="padding: 2px 10px;">h</td><td style="padding: 2px 10px;">858</td></tr> <tr><td style="padding: 2px 10px;">b</td><td style="padding: 2px 10px;">809</td></tr> <tr><td style="padding: 2px 10px;">l</td><td style="padding: 2px 10px;">792</td></tr> </table>	h	858	b	809	l	792	<table border="1" style="border-collapse: collapse; width: 100%;"> <tr><td style="padding: 2px 10px;">a</td><td style="padding: 2px 10px;">839</td></tr> <tr><td style="padding: 2px 10px;">u</td><td style="padding: 2px 10px;">783</td></tr> <tr><td style="padding: 2px 10px;">o</td><td style="padding: 2px 10px;">707</td></tr> </table>	a	839	u	783	o	707	<table border="1" style="border-collapse: collapse; width: 100%;"> <tr><td style="padding: 2px 10px;">t</td><td style="padding: 2px 10px;">888</td></tr> <tr><td style="padding: 2px 10px;">n</td><td style="padding: 2px 10px;">681</td></tr> <tr><td style="padding: 2px 10px;">x</td><td style="padding: 2px 10px;">675</td></tr> </table>	t	888	n	681	x	675
h	858																			
b	809																			
l	792																			
a	839																			
u	783																			
o	707																			
t	888																			
n	681																			
x	675																			

图 15

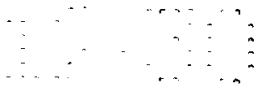


候选字符串	估计值
hat	2810
hut	2754
•	
bat	2536
• •	

图 16

所登记的字符串	上次访问日期	访问重现数
•		
hat	99/1/3	1
•		
hut	99/1/6	2
• •		

图 17



候选字符串	估计值
hat	2810
hut	2859
.	
bat	2536
.	
.	

图18



(a)

礼 [rei]	長 [cho]
------------	------------

礼 [rei]
社 [sha]
杠
杜
私

(b)

社 [sha]	長 [cho]
------------	------------

礼 [rei]
社 [sha]
杠
杜
私

图 19



(a)

社	長
[sha]	[cho]

(b)

	長
	[cho]

显示下一个 后选字符串

重新输入

图 20

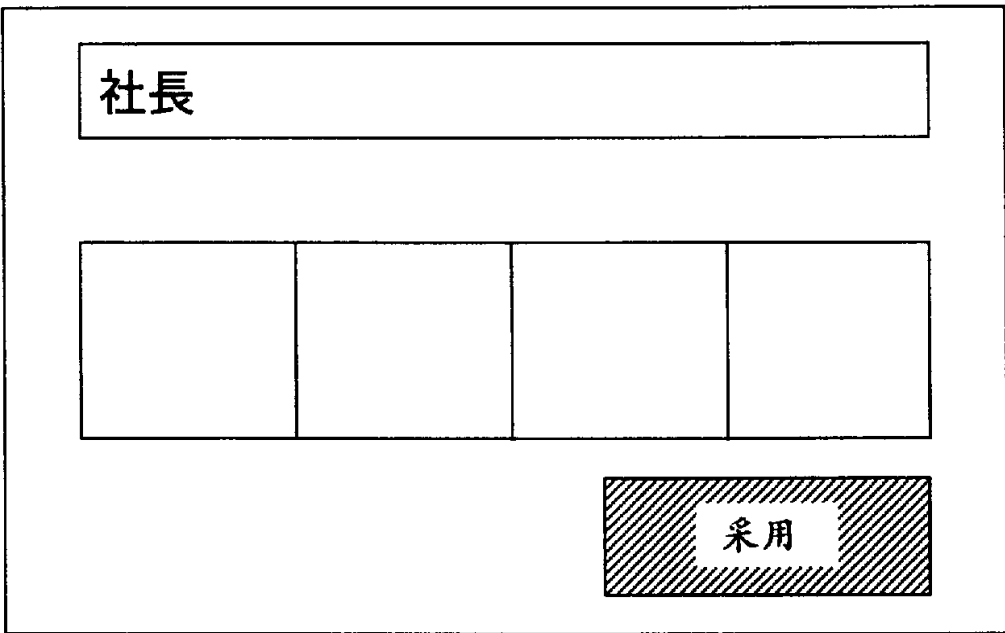


图 21

社	長					

图 22

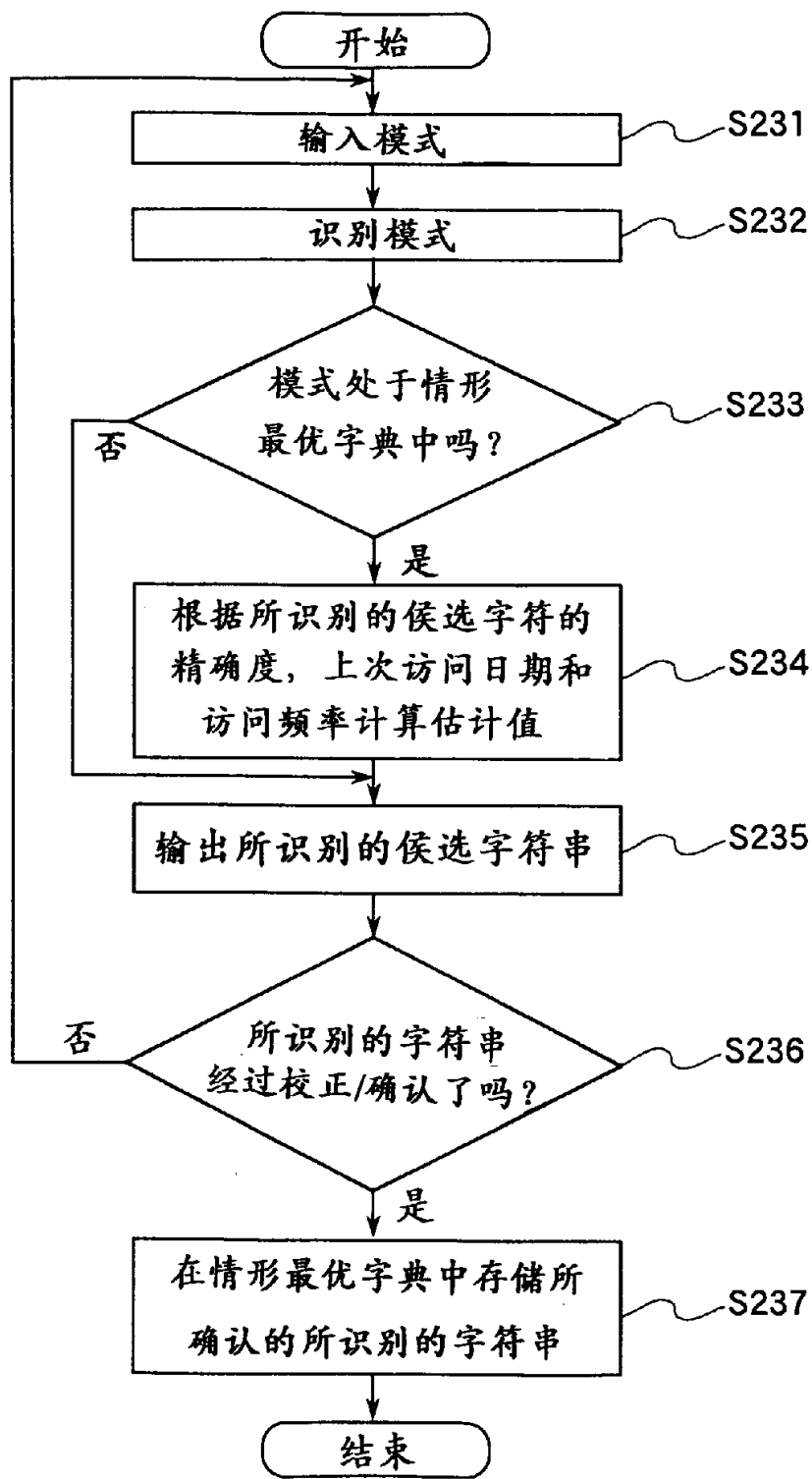


图23

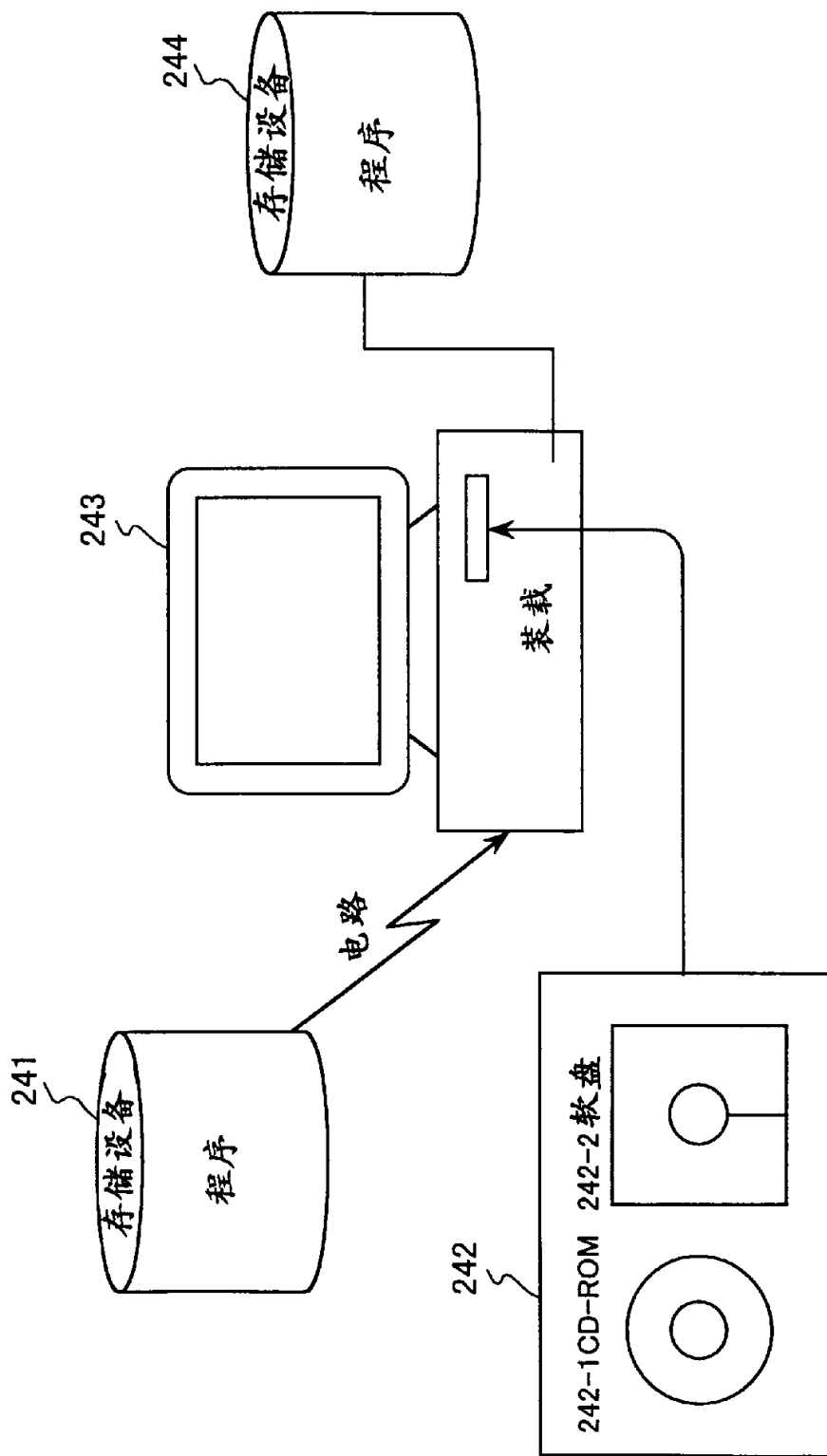


图 24