



[12] 发明专利说明书

[21] ZL 专利号 95105403.1

[43] 授权公告日 2003 年 4 月 2 日

[11] 授权公告号 CN 1104677C

[22] 申请日 1995.5.4 [21] 申请号 95105403.1

[30] 优先权

[32] 1994. 9. 16 [33] JP [31] 221543/1994

[71] 专利权人 国际商业机器公司

地址 美国纽约州

[72] 发明人 山下晶夫 平山唯树

审查员 杨 蕊

[74] 专利代理机构 中国专利代理(香港)有限公司

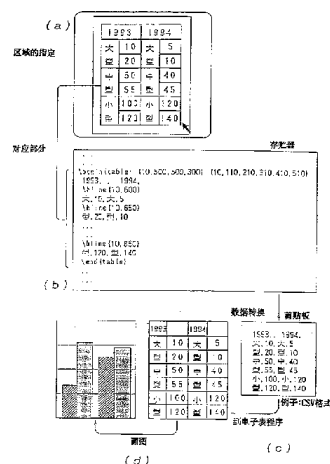
代理人 杜有文 马铁良

权利要求书 11 页 说明书 17 页 附图 9 页

[54] 发明名称 表中的剪贴方法与数据处理系统

[57] 摘要

本发明使把文本编辑程序中的排成表格式而不是形成几何格式的文本显示数据作为普通单元数据粘贴到一个电子表程序的一个表区域成为可能。上述目标是这样实现的：在执行剪和贴操作的文本编辑程序一方提供一个表格式数据转换装置以处理要送到剪贴板的或要通过 DDE 传送给一个电子表程序的数据。表格式数据转换装置通过检测直线数据或文本编辑器的一个表的列之间的空白来把数据转换成反映表结构的格式，例如 CSV 格式。



1. 一种表数据的剪贴方法，使用一个剪贴操作把表结构数据从一个文本编辑程序贴到一个应用程序的一个表格式区域，该应用程序能够以表格式显示数据，并能够把包括表结构数据的预定格式的数据通过一个计算机系统的处理贴到表格式区域中，所述表格式数据剪贴方法包含如下步骤：

(a) 通过一个指点装置指定包含被所述文本编辑程序所编辑的文本的表的一个区域的范围；

(b) 由所述计算机系统分析在所述指定区域范围中的表结构数据；

(c) 根据所述被分析的表结构数据，由所述计算机系统把所述指定区域范围内的文本数据转换成能被所述应用程序识别为包括表结构的数据的格式的数据；以及

(d) 把所述步骤(c)所转换的数据以可以被识别为包括表结构的数据的格式贴到所述应用程序的表格式区域中。

2. 权利要求1提出的表数据剪贴方法，其特征在于所述应用程序是一个电子表程序。

3. 权利要求1提出的表数据剪贴方法，其特征在于所述应用程序是一个能够在一部分文本中以表格式显示数据的文本处理程序。

4. 权利要求1提出的表数据剪贴方法，其特征在于所述应用程序是一个能够以表格式显示多个记录及其项目的数据库程序。

5. 权利要求1提出的表数据剪贴方法，其特征在于所述把转换成的可被识别为包括表结构的数据的格式的数据贴到所述应用程序的

所述表格式区域中的步骤是通过一个剪贴板完成的。

6. 权利要求1提出的表数据剪贴方法, 其特征在于所述把转换成的可被识别为包括表结构的数据的格式的数据贴到所述应用程序的所述表格式区域中的步骤是通过动态数据交换完成的。

7. 权利要求1提出的表数据剪贴方法, 其特征在于所述能被识别为包括表结构的数据的格式是一个这样的格式, 它的列方向由逗号分开, 其行方向由一个换行符分开。

8. 权利要求1提出的表数据剪贴方法, 其特征在于所述能被识别为包括表结构的数据的格式是这样一个格式, 其列方向由一个制表符分开, 其行方向由一个换行符分开。

9. 权利要求1提出的表数据剪贴方法, 其特征在于对在所指定区域范围内的表结构数据的分析是根据包含或指定在所述区域中的直线数据完成的。

10 权利要求9提出的表数据剪贴方法, 其特征在于对在所指定范围区域中的表结构数据的分析具有延长假想直线的步骤, 这样所述区域被网格化, 确定网格区域与一个实际数据区域的相关性, 并确定基于所述相关性的表结构数据。

11. 权利要求1提出的表数据剪贴方法, 其特征在于对在指定范围区域内的表结构数据的分析是基于包含在所述区域中的数据列中存在的一个空白区域而完成的。

12. 权利要求11提出的表数据剪贴方法, 其特征在于所述步骤(c) 具有一个空白块的相对端到达所述指定范围区域的边界时在所述空白块的区域中画一条假想直线的步骤。

1 3 . 权利要求 1 2 提出的表数据剪贴方法, 其特征在于所述步骤(c) 具有当一个空白块的一端到达所述指定范围区域的边界或者所述假想直线, 并且所述空白块的另一端到达一条假想直线时, 在所述空白块的一个区域中画一条假想直线的步骤。

1 4 . 在一个计算机系统中, 其中在一个支持多窗口环境的操作系统下, 在一个窗口中正有一个文本编辑程序被访问, 在另一个窗口中正有一个应用程序被访问, 该应用程序能够以表格式显示数据并且把包括表结构数据的预定义格式数据贴到所述列表格式的一个显示区域中, 一个表数据剪贴方法, 用于把正被所述文本编辑程序编辑的文本中的一个表区域由一个剪贴操作贴到所述应用程序的表格式显示区域中, 它包含如下步骤:

(a) 通过一个指点装置指定一个包含正被所述文本编辑程序编辑的文本的表的区域的范围;

(b) 由所述计算机系统分析在所述指定区域的范围内的表结构数据;

(c) 根据所述被分析的表结构数据, 由所述计算机系统把所述指定区域范围内的文本数据转换成能被所述应用程序识别为包括表结构的数据的格式的数据;

(d) 把所述步骤(c) 中被转换的数据由所述计算机系统存入一个剪贴板中; 以及

(e) 把在步骤(c) 中被转换成的、能被识别为包括表结构的数据的格式的数据由所述计算机系统通过从所述剪贴板取得数据来贴到所述应用程序的所述表格式显示区域上。

1 5 . 权利要求 1 4 提出的表数据剪贴方法, 其特征在于所述应用程序是一个电子表程序。

1 6 . 权利要求 1 4 提出的表数据剪贴方法, 其特征在于所述应用程序是一个能够在一部分文本上以表格式显示数据的文本处理程序。

1 7 . 权利要求 1 4 提出的表数据剪贴方法, 其特征在于所述应用程序是一个能够以表格式显示多个记录及其项目的数据库程序。

1 8 . 权利要求 1 4 提出的表数据剪贴方法, 其特征在于所述能被识别为包括表结构的数据的格式是这样一个格式, 其列方向由逗号分开, 其行方向由一个换行符分开。

1 9 . 权利要求 1 4 提出的表数据剪贴方法, 其特征在于所述能被识别为包括表结构的数据的格式是这样一个格式, 其列方向由一个制表符分开, 其行方向由一个换行符分开。

2 0 . 权利要求 1 4 提出的表数据剪贴方法, 其特征在于所述能被识别为包括表结构的数据的格式是一个SYLK 格式。

2 1 . 权利要求 1 4 提出的表数据剪贴方法, 其特征在于对指定区域范围内的表结构数据的分析是根据包含或指定在所述区域中的直线数据完成的。

2 2 . 权利要求 2 1 提出的表数据剪贴方法, 其特征在于对指定区域范围内的表结构数据的分析包含延长假想直线的步骤以使该区域网格化, 确定网格化区域与实际数据区域的相关性并根据所述相关性确定表结构数据。

2 3 . 权利要求 1 4 提出的表数据剪贴方法, 其特征在于对指定区域范围内的数据的分析是基于在包含在所述区域中的数据列中存在的一个空白区域来完成的。

24. 一个表数据剪贴方法，使用一个剪贴操作用来把表结构数据通过一个计算机系统的处理从一个第一应用程序贴到一个第二应用程序的一个表格式区域，该第一应用程序能够在一个预定的存储区域内保存文本数据，并在至少一部分所述文本数据被显示在屏幕上的条件下能够访问和取得显示位置上的相关的文本数据；第二应用程序能够以表格式显示数据，并能把包括表结构数据的预定格式的数据贴到一个所述表格式区域中，所述表数据剪贴方法包含下列步骤：

- (a) 由一个指点装置指定包含正被所述第一应用程序显示在屏幕上的一个表的区域的范围；
- (b) 由所述计算机系统取得与所述指定区域范围相关的文本数据；
- (c) 基于所述取得的文本数据，由所述计算机系统分析在所述指定区域范围内的表结构数据；
- (d) 基于所述被分析的表结构数据，由所述计算机系统把在所述指定区域范围内的文本数据转换成能被识别为包括表结构的数据的格式的数据；以及
- (e) 把所述步骤(d)中被转换的数据以可被识别为包括表结构的数据的格式贴到所述第二应用程序的所述表格式区域中。

25. 权利要求24提出的表数据剪贴方法，其特征在于所述第一应用程序是一个能够把从扫描仪取得的图像作为一个图像显示在所述屏幕上的应用程序。

26. 权利要求24提出的表数据剪贴方法，其特征在于所述第二应用程序是一个电子表程序。

27. 权利要求24提出的表数据剪贴方法，其特征在于所述第二应用程序是一个能够以表格式显示一部分文本中的数据的文本处理

程序。

28. 权利要求24提出的表数据剪贴方法，其特征在于所述第二应用程序是一个能够以表格式显示多个记录及其项目的数据库程序。

29. 权利要求24提出的表数据剪贴方法，其特征在于所述把转换成的可以被识别为包括表结构的数据的格式的数据贴到所述第二应用程序的所述指定区域的步骤是由动态数据交换完成的。

30. 权利要求24提出的表数据剪贴方法，其特征在于所述把转换成的可以被识别为包括表结构的数据的格式的数据贴到所述第二应用程序的所述表格式区域的步骤是由剪贴板完成的。

31. 权利要求24提出的表数据剪贴方法，其特征在于所述能被识别为包括表结构的数据的格式是这样的格式：它的列方向由一个逗号分开，它的行方向由一个换行符分开。

32. 权利要求24提出的表数据剪贴方法，其特征在于所述能被识别为包括表结构的数据的格式是这样的格式：它的列方向由一个制表符分开，它的行方向由一个换行符分开。

33. 在一个计算机系统中，其操作系统支持多窗口环境，一个用于观察一个光学字符阅读机（OCR）的识别结果的显示器正在一个窗口中被访问，而一个能够以表格式显示数据并能将预定义格式的数据包括表结构数据贴到所述表格式区域中的应用程序正在另一个窗口中被访问，一种表数据的剪贴方法，用来把所述显示器中的一个表区域由一个剪贴操作贴到所述应用程序的表格式显示区域中，它包含如下步骤：

（a）由一个指点装置指定包含由所述显示器显示的图象中包括一个表的区域的范围；

(b) 由所述计算机系统访问与所述指定区域相关的OCR 识别结果中的标记文本；

(c) 相应于所述指定区域由所述计算机系统分析标记文本中的表结构数据；

(d) 基于所述被分析的表结构数据，由该计算机系统把在所述指定区域范围内的文本数据转换为可以被所述应用程序识别为包括表结构的数据的格式；以及

(e) 由所述计算机系统把在所述步骤(d) 中被转换的数据存到剪贴板中；以及

(f) 由所述计算机系统把在所述步骤(d) 中转换的能被识别为包括表结构的数据的格式的数据通过从所述剪贴板取数据来贴到所述应用程序的所述表格式显示区域上。

3 4 . 权利要求 3 3 提出的表数据剪贴方法，其特征在于所述步骤(d) 包括从所述标记文本中清除一个标记的过程。

3 5 . 权利要求 3 3 提出的表数据剪贴方法，其特征在于所述能被识别为包括表结构的数据的格式是这样的格式：它的列方向由一个逗号分开，它的行方向由一个换行符分开。

3 6 . 权利要求 3 3 提出的表数据剪贴方法，其特征在于所述能被识别为包括表结构的数据的格式是这样的格式：它的列方向由一个制表符分开，它的行方向由一个换行符分开。

3 7 . 一个数据处理系统，它包含

(a) 一个程序存储装置，用于存储一个文本编辑程序和一个应用程序，使得它们能够被访问，该应用程序能够以表格式显示数据，并能把预定格式数据包括表结构数据贴到一个所述表格式的区域中；

(b) 一个指点装置用于指向屏幕上的任意一点；

(c) 一种通过所述指点装置指定所述文本编辑程序的一个屏幕显示区域的区域的装置；

(d) 一种装置，用于响应被所述范围指定装置所指定的所述区域而分析所述区域内的表结构数据；

(e) 一种装置，用来根据所述分析的结果，把指定区域范围内的文本数据转换成能被所述应用程序识别为包括表结构的数据的格式的数据，并用来暂时存储所转换的数据；以及

(f) 一种装置，用来把暂存于所述存储装置中的数据贴在所述应用程序的表格式显示区域中。

38. 权利要求37提出的数据处理系统，其特征在于所述应用程序是一个电子表程序。

39. 权利要求37提出的数据处理系统，其特征在于所述应用程序是一个能够把一部分文本中的数据以表格式显示出来的文本处理程序。

40. 权利要求37提出的数据处理系统，其特征在于所述应用程序是一个能够以表格式显示多个记录及其项目的数据库程序。

41. 权利要求37提出的数据处理系统，其特征在于所述指点装置是一个鼠标器。

42. 权利要求37提出的数据处理系统，其特征在于所述用于贴数据的装置通过一个剪贴板传送数据到所述应用程序的所述表格式显示区域。

43. 权利要求37提出的数据处理系统，其特征在于所述用于

贴数据的装置通过动态数据交换传送数据到所述应用程序的所述表格式显示区域。

4 4 . 权利要求 3 7 提出的数据处理系统，其特征在于所述能被识别为包括表结构的数据的格式是这样的格式：它的列方向由一个逗号分开，它的行方向由一个换行符分开。

4 5 . 权利要求 3 7 提出的数据处理系统，其特征在于所述能被识别为包括表结构的数据的格式是这样的格式：它的列方向由一个制表符分开，它的行方向由一个换行符分开。

4 6 . 权利要求 3 7 提出的数据处理系统，其特征在于所述能被识别为包括表结构的数据的格式是一个SYLK 格式。

4 7 . 权利要求 3 7 提出的数据处理系统，其特征在于对指定区域范围内的表结构数据的分析是基于包含或指定在所述区域中的直线数据进行的。

4 8 . 权利要求 4 7 提出的数据处理系统，其特征在于对指定范围区域内的表结构数据的分析具有延长假想直线的步骤，以便使所述区域网格化，确定网格区域与实际数据区域的相关性，并确定基于所述相关性的表结构数据。

4 9 . 权利要求 3 7 提出的数据处理系统，其特征在于对指定范围区域内的表结构数据的分析是基于包含在所述区域内的数据列中存在的空白区域来进行的。

5 0 . 一个数据处理系统，它包含：

(a) 程序存储装置，用于存储一个第一应用程序和一个第二应用程序，使得它们能够被访问，其第一应用程序能够在一个预定存储区

域内保存文本数据，并在至少有一部分所述文本数据显示在屏幕上的条件下能够访问和取得显示位置上的相关文本数据；第二应用程序能够以表格式显示数据并能把预定格式数据包括表结构数据贴到一个所述表格式的区域中；

(b) 一个指点装置，用于指向位于屏幕上的一个任意点；

(c) 一个用于通过所述指点装置指定所述应用程序的一个屏幕显示区域的范围的装置；

(d) 一个装置，用于访问被保存的文本数据并响应于被所述范围指定装置所指定的所述区域来取得相应于所述指定区域的数据；

(e) 一个用于根据所取得的数据分析所述区域内的表结构数据；

(f) 一个装置，用于根据所述分析结果把指定区域范围内的文本数据转换成可被所述第二应用程序识别为包括表结构的数据的格式的数据，并用于暂时存储所转换的数据；以及

(g) 一个装置，用于把由暂存装置所存的数据贴到所述第二应用程序的表格式显示区域中。

5 1 . 权利要求 5 0 提出的数据处理系统，其特征在于所述第一应用程序是一个OCR 的显示器。

5 2 . 权利要求 5 0 提出的数据处理系统，其特征在于所述第二应用程序是一个电子表程序。

5 3 . 权利要求 5 0 提出的数据处理系统，其特征在于所述第二应用程序是一个能够以表格式显示一部分文本中的数据的文本处理程序。

5 4 . 权利要求 5 0 提出的数据处理系统，其特征在于所述第二应用程序是一个能够以表格式显示多个记录及其项目的数据库程序。

5 5 . 权利要求 5 0 提出的数据处理系统，其特征在于所述贴数据的步骤通过一个剪贴板传送数据到所述应用程序的所述表格式显示区域中。

5 6 . 权利要求 5 0 提出的数据处理系统，其特征在于所述贴数据装置通过动态数据交换传送数据到所述应用程序的所述表格式显示区域。

表中的剪贴方法与数据处理系统

本发明与在一台计算机的多窗口环境中运行的应用程序（比如字处理程序与电子表程序（spreadsheet program））之间对文档和数据的剪贴技术有关。

在个人计算机发展的最初阶段，由于CPU 的性能低和主存储器容量有限，操作系统只能处理简单的任务。个人计算机因而只能在同一时间执行字处理程序，文本编辑器和电子表程序中的一个。

然而在某种情况下可能需要把一个由字处理器产生的文档用于一个电子表程序，或者电子表程序产生的表被一个字处理器利用。这最终可以实现，但要用麻烦的方法。就是说，一个文档或者一个指定了范围的表的区域被赋予一个作为文本格式数据的文件名，并通过访问字处理器或者字处理器程序的一个特殊功能写入磁盘。该程序结束后，运行另一个程序，并指定存在磁盘上的那个文本文件的文件名。这样文本或者表的数据便被传给了运行的程序。

但在最近，随着CPU 性能的提高和CPU 地址空间的扩大，廉价的半导体存储器和大容量硬盘也已方便获得。因为这个，支持多任务和多窗口系统的操作系统例如OS/2（IBM 的商标）和Windows（Microsoft 的商标）也被开发出来被广泛使用。

例如在这些操作系统下，不仅可以在一个窗口运行一个字处理器

而同时在另一个窗口运行一个电子表程序，而且可以把一个应用程序的数据剪下并贴到另一个应用程序中。

具体地是这样实现的：首先使一个应用程序的窗口活跃，使用一个指点设备例如鼠标器指定一个矩形区域范围。然后从菜单条上拉下编辑菜单，并选择“复制”或者“剪下”。该指定区域内的数据（文本或位图）便被存入剪贴板一次。

然后使另一个应用程序的窗口活跃，把光标移到想要的位置，从菜单条上拉下编辑菜单并选择“粘贴”。剪贴板中存的文本或位图数据便被贴到光标位置上。

这种接口的优点在于尽管数据格式取决于应用程序，屏幕上看到的文本数据或位图还是可以通过剪贴板贴到另一个应用程序的窗口中。例如把一个图像的一部分从正在编辑TIFF 格式的位图文件的图像编辑器使用图象粘贴功能贴到一个字处理器窗口的预定位置上是不可能的。在粘贴时，数据贴入的字处理器是否支持TIFF 图象格式是无关紧要的。

另外，经常会产生这样的需要，即一个应用程序例如字处理器要使用以印刷文档所描述的表，但是这样的文档很少有可能以电子格式存储，在文档是通用杂志的情况下，便没有这种可能。

PUPA 4 - 278634 公开了一种用扫描仪把印刷文档转换成图象文件并执行文档识别的技术，以便反映出图象中的信息。与该技术相适应，可以从一个印刷文档获得一个电子文件，并保持了原文档的表和行组。

PUPA 2 - 58970 进一步公开了一个来自主计算机的图象和从扫描仪读入的图象可以被任意移动，剪裁和合并以输出一个图象。

还有，PUPA 2-301860 公开了一个文档输入装置，一个图象输入装置，及一个图象存储器。图象存储器包含一个屏蔽装置，输入图象数据存在其中，它还屏蔽图象存储器中图象数据的存储，还有一个用于指定图象存储器中的一个文档的分布位置的指定装置，以及一个分布装置（spread means），它用于把相应于文档数据的字符分布（spread）到所考虑的图象存储器中的文档所分布的位置。

结合这些先有技术，一方面有可能通过例如剪贴的方法把显示器上的从一个打印文档获得的图象与字符数据结合在一起，并做成一个所期望格式的电子文档。

另一方面，电子表程序经常用于金融与商业管理领域并成为许多企业的主要运作手段。但是，因为先有技术所描述的剪贴的数据是图象数据，它们不能作为一种规则的单元数据形式贴到电子表程序中。至多可能作为电子表程序所显示的馅饼图装饰来粘贴数据。即使费尽周折做出了一个表格式的图象，当观察一个表时为了将输入数据给电子表程序就必须进行包括手工输入字符的低效率的处理。

但是，这并不是由以位图图象形式来剪贴数据所带来的唯一麻烦。例如，如果一个以表格式写入的部位由一个普通字处理器来剪贴，文本数据一般会存在剪贴板中。但是，如果该部位被贴入一个电子表程序，原文档的整个一行被存入一个单元，因此即使有一个部位其数字按长度方向排列，也不会不可能计算一个平均值。或者比如只根据这一行的数字做一个条形图。

本发明的一个目标是提供一种方法和一个系统，使得可以把从OCR读入的表数据作为一个普通的单元数据贴到一个电子表程序的表区域。

本发明的另一个目标是提供一种方法和一个系统，使得可以把以表格式排列而不是形成一个几何格式的文本显示数据作为普通的单元数据贴到一个电子表程序的表区域中。

上述目标在本发明中是这样达到的：在启动剪贴操作的应用程序一边提供一种机制来处理发送到剪贴板或通过动态数据交换（DDE）直接传送到目标程序的数据。

更具体地，当要执行剪贴操作的一个应用程序是一个OCR 识别结果的显示器时，该应用程序被设计得使识别结果能以标记文本来保存。这里使用的标记目的是指出诸如一个表的大小，高度，以及起始点坐标，该表以可识别的格式出现，表示成由一个特殊符号例如“\.”开始。因此，当要剪贴的范围由鼠标的拖动操作来指定时，可以参考所保存的标记文本的坐标数据来计算指定了标记文本中的哪一部分。于是，在标记文本的指定部分被存到应用程序的缓冲区域中后，缓冲区域的内容并不立即传送到剪贴板，而是访问把表数据处理成比如CSV 格式的函数。这样，数据被转换成文本数据格式。其中行结束是一个换行符而列的结束是一个逗号（或者标记）。通过以这种格式执行的粘贴操作从剪贴板到一个典型的电子表程序如Excel（Microsoft 的商标）或Lotus 1-2-3（Lotus Development 的商标）的Windows 版，使把想要的数据存在表的 每个单元中并保持其表格式成为可能。

如果剪贴操作是根据先有技术进行，当要执行剪贴操作的应用程序是一个普通文本处理器如字处理器或文本编辑器时，用比如鼠标器指定的文本不转换而直接传送给剪贴板。然而根据本发明，指定区域的内容在传送到剪贴板之前曾被存在缓冲区中，并且该区域根据直线

数据被网格化。然后，访问一个类似于在上述OCR 的情况下把标记文本数据转换为CSV 格式的函数。最后，要送到剪贴板的文本被转换成这样的格式：行结束是一个换行符，列结束是一个逗号（或制表符）。

本发明的一个实施例将参考图来描述。

图 1 是一个显示本发明赖以实现的硬件的结构图；

图 2 是一个显示本发明的处理单元的方框图；

图 3 是一个显示怎样把一个表从一个OCR 显示器剪贴到一个电子表程序的图。

图 4 是一个显示本发明的处理过程的流程图。

图 5 是一个显示根据直线数据分析字处理器或文本编辑器的表的过程的流程图。

图 6 是一个显示在剪下粘贴一个表之前的OCR 显示器和电子表程序的屏幕的例子图。

图 7 是一个显示在剪下粘贴一个表之后的OCR 显示器和电子表程序的屏幕的例子图。

图 8 是一个显示在一个字处理器或文本编辑器中包括直线的表的例子图；

图 9 是一个显示基于直线的一个假想网格区域的图；

图 10 是一个显示一个表中的一个实际数据与一个网格单元之间的对应关系的图；

图 11 是一个显示一个表中的CSV 格式的转换数据的图；

图 12 是一个显示这个表在被贴入电子表程序后的状态的图。

图 13 是一个用来解释独立于直线数据的表结构的分析方法的图；

图 1 4 是一个用来解释独立于直线数据的表结构的分析方法的图；
图 1 5 是一个用来解释独立于直线数据的表结构的分析方法的图；
图 1 6 是一个用来解释独立于直线数据的表结构的分析方法的图；
图 1 7 是一个用来解释独立于直线数据的表结构的分析方法的图；
图 1 8 是一个用来解释独立于直线数据的表结构的分析方法的图；
图 1 9 是一个用来解释独立于直线数据的表结构的分析方法的图；
图 2 0 是一个用来解释独立于直线数据的表结构的分析方法的图；

A. 硬件组成

参考图 1，示出了实现一个本发明的系统配置。该配置众所周知，包含一个系统设备1002，它系统设备包括主存（RAM），一个CPU，一个控制器和硬盘单元；一个键盘1004，用于向系统设备1002 输入键盘命令或字符串；一个能够显示字符和彩色位图图形的显示单元1006；一个鼠标器1008，用来指向显示单元1006 的屏幕上的一个任意位置并传送定位信息给CPU。另外，系统设备1002 还通过一个预定接口卡接到一个光学字符阅读器（optical character read-OCR）的图象扫描仪1010。

在硬盘单元中存储了一个操作系统和一些应用程序，例如电子表程序，字处理器，数据库，以及文本编辑器。

可以该操作系统，可采用支持GUI 多窗口环境的程序，该程序还有通过剪贴板剪贴的功能或者在不同窗口的应用程序之间的动态数据交换（DDE）功能，例如Windows（Microsoft 的商标），OS/2（IBM 的商标），在AIX（IBM 的商标）上的X-WINDOW 系统（MIT的商标）。

B. 系统配置

本发明的系统配置将在下面参考图 2 做更为详尽的描述。

首先，一个图象输入单元2002 接到一个光学字符阅读器（OCR）2001，一个由扫描仪1010 读入的图象通过OCR 2001、图象输入单元2002 及一个输出控制单元2012 显示在显示单元1006 的一个窗口内。

一个文本输入单元2004 从一个处理单元2010 获得文本显示数据并把它传到输出控制单元2012，这样其内容就能显示在显示单元1006 的屏幕的一个窗口内。

输入单元2006 是一个处理单元，用来检测一个事件，诸如鼠标器的按动，拖动和两次按动，光标或键盘的运动，按动回车键，这个检测信号实际上是在一个输入确定单元2007 中确定的。

输入确定单元2007 检测光标的当前位置以及根据光标位置而指定的范围。所指定的范围通常为矩形，比如说，通过按下鼠标键并从光标的某个位置拖动鼠标，屏幕上便出现一个矩形区域，如果按着键继续移动鼠标，矩形区域便扩大。当矩形区域扩大到覆盖住预定区域时释放鼠标键，范围的指定便完成了。

接到OCR 2001 的处理单元2008 存储由OCR 2001 分析的标记文本数据，特别是被翻译并表现为一个表的那部分标记文本数据被本发明所利用。

处理单元2010 包括一个缓冲存储器以存储文本数据用于执行本发明的处理的字处理器或文本编辑器，其内容的一部分通过文本输入单元2004 和输出控制单元2012 显示在屏幕上的一个窗口中。

基于由输入确定单元2007指定的范围，一个表数据获取单元2014 访问存在处理单元2008 中的标记文本数据，并从标记文本数据中取得相应于一个表的一部分。

一个直线和字符数据获取单元2016 访问处理单元2010 以存储用于字处理器或文本编辑器的内部数据，并在一个由输入确定单元2007 指定的区域中取得直线与字符数据。

一个表分析单元2018 检测一个字符，该字符要存入一个通过以表格形式分割一个指定区域而得到的一个单独的分割区域中。注意表分析单元2018 中的处理将在后而描述。

一个电子表数据转换单元2020 从表数据获取单元2014 或表分析单元2018 接收数据并把数据转换成电子表格式，该格式中数据由CSV 或制表符分开并能把数据存在一个独立单元中，并考虑到典型的电子表程序如Excel (Microsoft 的商标) 或Lotus 1-2-3 (Lotus Development 的商标) 的Windows 版而保持表格式。

经这种方式转换的电子表数据通过一个外部存储传送单元2022 传送到剪贴板或者通过动态数据交换 (DDE) 直接传送到电子表程序。作为替代，嵌入在一个电子表程序中的一个表中的对象的数据由对象链接嵌入 (OLE) 来改变。

C. 本发明的处理

后面将参考图 3 中的操作、图和图 4 和 5 中的流程图来描述本发明的处理过程。由于针对OCR 数据显示器处理和针对字处理器或文本编辑器的处理略有不同，对它们将分别描述。

OCR 数据显示器

如果在根据本发明的OCR 数据显示器中，比如通过揷动标题条上的“文档”来访问一个下拉菜单并选择其中的“表复制”，就开始了图 4 的流程图所示的处理。在图 4 中的步骤4002，通过一个指点设备例如鼠标器1008 指定一个如图 3 (a) 所示的矩形范围。在图 2 中这

是由输入单元2006 完成的。指定一个区域的方法有多种，例如上述的拖动方法以及从一个下拉菜单的菜单条选择“范围指定”的方法，扩大一个矩形区域是通过简单移动鼠标或按着光标键来完成，指定一个矩形区域则是用拖动鼠标或按回车键完成。本发明对这些特定的范围指定方法并无限制，还可以采用自定的范围指定方法。

如果用此方法指定范围，在步骤4004，指定的矩形范围的四个角的坐标将由图 2 中的输入确定单元2007 确定。

下一步在步骤4006，被OCR分析的数据2008 被数据获取单元2014 根据指定矩形范围的四个角的坐标来访问。于是，被OCR 分析的文本数据，例如图 3（b）所示，始于“`\begin {table}`”，止于“`\end {table}`”，被表数据获取单元2014 取得。因为产生标记文本数据的方法例如图 3（a）所示的通过OCR，并非本发明的主要论题，并由例如PUPA 4 - 278634 的出版物所公开，所以此处略去对该方法的描述。

注意在图3(b)中，对应于直线数据的是例如在“`\begin{table}`” 一行中的{10,500,500,300} 表现为一条竖线的数据，以及表现为一条水平线的例如“`\hline{10,600}`” 的数据。

在步骤4008，由数据获取单元2014 提供的如图 3（b）所示的数据被由电子表数据转换单元2020 转换成比如图 3（c）所示的CSV 数据格式。因此在最简单的情况下所需的处理就只是清除标记。但是在一个专用的电子表程序的情况下，某些时候需要不是由逗号分开而是用制表符分开的文本数据，来粘贴表数据。另外，需要一个引在引号（“ ”）中的字符串，并且如果它是基于用在Multiplan（Microsoft 的商标）中的SYLK 格式，就会有很多变化。因此最好在本发明的OCR

数据显示器的菜单中提供多种表数据格式，这样可以从中任意选择一个。

CSV 格式和SYLK 格式的相互关系是这样的：就是说表 1 所示的数据采用逗号的CSV 格式的。

表 1

“TOKYO” , 1, 110.21

“OSAKA” , 2, 89.05

表达式 1 示出了SYLK 格式：

表达式 1

B; Y 2 ; X 3

C; Y 1 ; X 1 ; K “TOKYO”

C; X 2 ; K 1

C; X 3 ; K110.21

C; Y 2 ; X 1 ; K “OKASA”

C; X 2 ; K 2

C; X 3 ; K89.05

E

在表达式 1 中，B; Y 2 ; X 3 指示表数据在 Y 方向的大小为 2 而在 X 方向的大小为 3，在以 “C; ” 开始的行中象 Y_n; X_m; K_a 这样的格式指示 “a” 的数据存在第 n 行第 m 列的单元中。还有，表达式 “Y_n” 被省略的行指示它与紧上面的一行具有相同的列位置。

在步骤4010，图 3（c）所示数据被外部数据传送单元2022 写入剪贴板。然后使被访问的电子表程序所在的窗口活跃，通过揪动鼠标使光标定位于想要的单元，然后，如果从电子表应用程序的菜单条拉

下“编辑”菜单并选择“粘贴”，在步骤4010写入的表数据将被贴入从而使得当前单元的位置保持在左上角。因为以这种方式贴入电子表应用程序的表保持其原来的表数据格式，所以有可能根据安排有数字的单元来绘出条形图，如图3（d）所示，或者指定包含数字的单元的多个范围并计算其平均值或标准差。

本例将参考图6和图7中示出的屏幕加以更为具体的描述。图6中，OCR识别结果显示器被显示在窗口6010中，电子表程序则显示于窗口6020中。注意该显示器具有符合本发明的剪贴板传送功能。

然后，比如说如果用鼠标指定了显示器中的表6012，通过掀动表6012的标题条中的“文档（Document）”便显示出一个下拉菜单（未示出），然后选择“表复制”（Table Copy）（未示出），那么根据本发明，在剪贴板中所存的数据将既不是显示器中指定范围的位图映象，也不是与表的显示器的显示内容相关联的主存或硬盘中所保存的表的位置的识别结果的一个标记文本，而是具有例如表6012所反映的结构的CSV格式的文本数据。

下一步，如果窗口6020被选为活跃窗口，掀动鼠标把当前单元置为A1，掀动菜单条上的“编辑（Edit）”以拉下“编辑”（Edit）菜单，然后选择“粘贴”（Paste），那么A1单元被置于左上角，表7010将被贴入电子表7002，同时反映表6012的结构。

字处理器或文本编辑器

如果在字处理器或文本编辑器中有本发明的表复制功能，例如通过掀动标题条中的“编辑”Edit（未示出）访问一个下拉菜单中的“表复制”（Table Copy），那么图4中的流程图中的操作4002将开

始执行。在这种情况下，由鼠标或键盘上的光标执行的“范围选择”（Range Specifiication）操作与OCR 显示器的情况是相同的。

下一步在步骤4004，利用表数据获取单元2014 根据所确定的矩形区域的四个角的坐标访问字处理器或文本编辑器的内部数据2010。然后在步骤4006 利用直线与字符获取单元2016 取得所指定范围内的数据。从表达式“直线”（Ruled line）可以理解，直线与字符获取单元2016 取得直线数据以及字符数据，其直线数据用来确定一个指定范围内的一个表的结构。就是说，在根据本发明的字处理器或文本编辑器的情况下，步骤4008 包含着如图 5 所示的多个步骤5002—5010，与上面描述的OCR 数据显示器的情况不同。

更具体地，参考图 5，在步骤5002 执行一个延长直线的操作。在字处理器或文本编辑器的情况下，之所以需要这一步，是因为每一列（或行）中的块的长度不总是相等的，如图 8 所示。（实际上，对于一个被OCR 识别的表，定每一列中块的长度并不总是相等。但是因为这样一个表的识别和翻译是事先由OCR 完成并反映在标记文本中，如图 3（b）所示，作为一个产生的结果，鉴于直线数据，所以表数据获取单元2014 不需要再次翻译该表）。由于在字处理器的情况下，直线数据是通过一个特殊符号或者在一个单独的与文档相联的属性文件中说明的方式来嵌入一个文档中。该特殊符号或属性文件被翻译。因为文本编辑器不具备一个特殊的直线功能，而是象“T”，“├”或“L”这样的直线字符通常作为一个文本嵌入，所以这样的直线字符被翻译。

如果以这种方法检测到直线的存在并且在步骤5002 中被扩展，那么在步骤5004 中一个区域将如图 9 所示被网格化。在图 9 中，实

线是实际的直线而虚线是假想的直线。而且，图 9 中的图 1 - 2 0 不是一个表的实际数据，而是赋予相应于表中的假想网格单元的区域序号。

在步骤5006 中，完成表中的实际文本数据到表中的假想的网格区域的映射。结果如图 1 0 所示。需要注意，特别是在图 1 0 中，字符 F 或 H 占据了多个网格单元。当映射是这样取得时，将对每一个网格单元按号码顺序执行下面的处理。就是例如当 $F = \{9, 10, 11\}$ 时，某些字符占用了多个网格单元，它就被存在号码最低的网格单元而且后面的单元置为空白。

当这一相关完成时，网格单元将如图 1 2 所示被字符占有。这一对单元的占有是在步骤5008 中完成的。如果如图 1 2 所示的格式的数据以这种方式被产生了，那么在步骤5010 中，可以很容易地产生电子表所要求的任意格式的数据。

如果比如说存有实际数据的网格单元被转换成实际数据被引号（“）括起来的字符，相邻的网格边之间的边界用逗号表示。一列的结束用一个换行符表示，象图 1 1 那样的CSV 格式的文本数据将被得到。依赖于在其中执行粘贴的电子表程序，产生很多变化，例如逗号用制表符来表示的方法，以及省略不用引号的方法，如上所述。

还可以想到，图 1 2 也展示了这样一种状态，即当如图 8 所示的字处理器或文本编辑器中的表真正贴在一个电子表程序上时该表出现在电子表程序中。

注意在图 2 中，上述步骤5002-5010 是由表分析单元2018 执行的处理过程。

当步骤5010 结束时，本发明对字处理器或文本处理器的情况的处理将返回图 4 中的步骤4010，但是对紧接着的处理的描述将被省略，因为它与上述OCR 显示器的情况相同。

注意因为上述步骤5002-5010 使用了直线数据，所以似乎有必要在字处理器或文本编辑器的情况下本发明的处理具有直线数据，但是也可能不用直线数据反映表数据。那么对于直线数据不总是出现在从字处理器或文本编辑器切下的一个表区域中的情况，即不依赖于直线的情况，将参考图 1 3 - 2 0 描述一个处理方法。

表分析处理的另一个实施例

现在假设如图 1 3 所示的一段表格式的文本被剪贴操作剪下，该文本是要分析的对象。按照本实施例的一个表分析程序，首先获得表中在垂直和水平方向的一组空白区域（空白块）。在一条直线（包括连续的减号以及┌，┐和└┘）已经存在的情况下，它被认为是一个空白。

垂直方向的空白由图 1 4 示出，水平方向的空白由图 1 5 示出。在图 1 4 和 1 5 中，空白块是由垂直方向和水平方向的箭头（← →）指示的。

然后按照如下规则在指定范围的表中绘出假想直线。

（1）当垂直空白块的相对端接触了外周边或者已画的直线时，在垂直空白块的位置画一条假想直线。

（2）当水平空白块的相对端接触了外周边或者已画的直线时，在水平空白块的位置画一条假想直线。

下面将针对图 1 3 的区域描述遵循上述规则（1）和（2）画一条假想直线的过程。

第一步如图 1 6 所示是遵照规则 (1)，对其相对端到达了外周边的线画一条假想垂直直线。

下一步如图 1 7 所示是遵照规则 (1)，对其一端到达外周边另一端到达实际的直线的线画一条假想的垂直直线。

然后，遵照规则 (2) 对其对端到达外周边的线画一条水平直线，如图 1 8 所示。

垂直和水平直线是以这种方式绘出，其结果如图 1 9 所示。注意图 1 9 并未在栏之间显示出边界线，如“Tanaka”列和“yamada”列之间，但实际上是假设在电子表中“Tanaka”和“yamada”是存储在垂直方向边界相接触的不同单元中的。因此，在所有列中的相邻列之间已经绘出了一条假想水平直线。如果图 1 9 未显示的位于列之间的假想水平直线由一条单线表示，并且在图 1 3 中事先绘出的一条实际直线和在图 1 8 中在空白列绘出的假想水平直线用一条双线表示，就得到了如图 2 0 所示的一个表结构。从此结构中，可以用伴随图 8 - 1 2 所描述的方法得到如图 1 1 所示的CSV 格式的文本。

在应用了图 8 - 1 2 所示的方法时，单线的直线和双线的直线在图 2 0 中被认为是一样的，但在某些情况下，可以进行一个分析以便使单线绘出仅是一个单元的边界，而双线绘出包含一个空白单元的一列。作为替代，包括属性数据的一个表结构的数据也可以被贴到一个电子表上，这样双线的位置变成了电子表中的一条直线。

· 在本发明中最好根据字处理器或文本处理器的表格式提供多种不同的表分析方法，并且可以从一个菜单中选取合适的表分析方法。若是这样，通过根据表的特性（比如包括直线的表，数字或字符仅仅按矩阵形式排列的表，以及已经用制表符分开的表）在菜单中选取一个

适当的表分析方法，一个具有一个结构的表便有更大可能贴到一个电子表中。

尽管在上述实施例中，字处理器和电子表程序在分离的窗口中被访问，最新型的一些字处理器可以在任意位置设置一个具有电子表功能的表。需要指出本发明适合于这种情况：即把表结构数据从同一窗口的一个应用程序的文本部分剪贴到其电子表部分。

另外，最近在一些电子表程序或字处理器之外的数据库程序中，记录可以表格式显示，并且比如CSV 格式的数据可以通过剪贴板或者DDE 被粘贴。

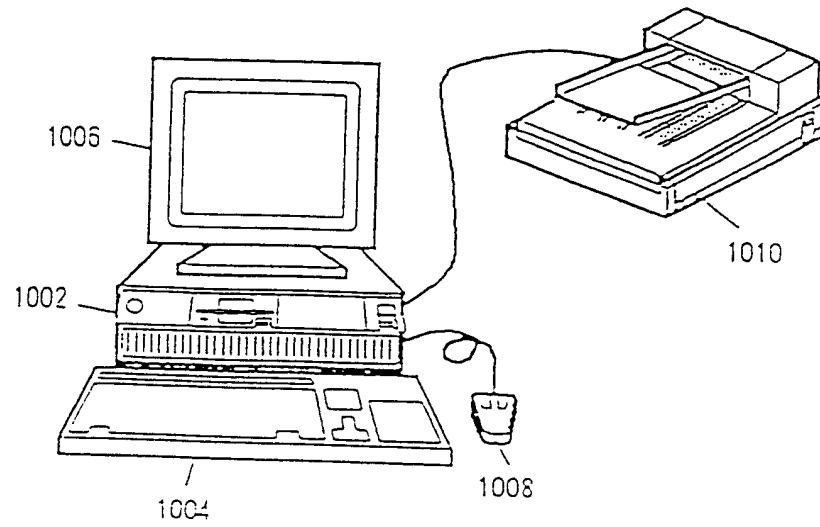
因此，依据本发明，表结构数据要贴入的目标应用程序并不局限于一个电子表程序，而是可以理解为它可以是任意一个能够以表格式显示数据并且能以一种特定的具有表结构的格式粘贴数据的应用程序。

还有，尽管在上面的实施例中，假设一个OCR 显示器，一个字处理器，及一个文本编辑器是获得表结构数据的源方的应用程序，但需要指出本发明当然不会局限于这些应用程序，而是适用于任意一个能够根据屏幕上的显示位置访问主存或辅存以获取与显示内容相关的文本数据的应用程序。此时，源方应用程序的显示屏幕可能是一个能够显示位图数据、图象数据、向量画线数据等等的图形显示屏，也可能是一个仅能显示文本字符的显示屏，例如一个文本编辑器。长话短说，必需的是源方应用程序能够存储与屏幕上的文本字符部分相关的文本数据，并能根据在一个指定范围内的字符数据的位置数据访问相应部分的文本数据。注意，当符合本发明的表格式被转换时，如有必要，指称的便是象直线数据那样的数据。因此一般地，象位图数据那样的

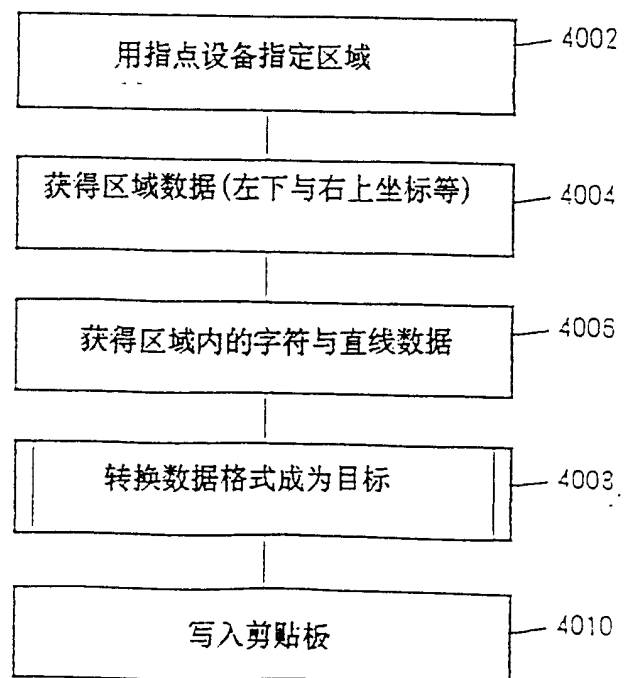
不在一个指定范围与文本数据相关的数据便不被传送到表数据转换单元，例如图 2 所示的电子表转换单元2020。

如上所述，根据本发明，不须用户另外费力，字处理器，文本编辑器和OCR 显示器中的表格式显示数据被直接通过一个剪贴操作贴到一个电子表程序中并保持其表结构成为可能。

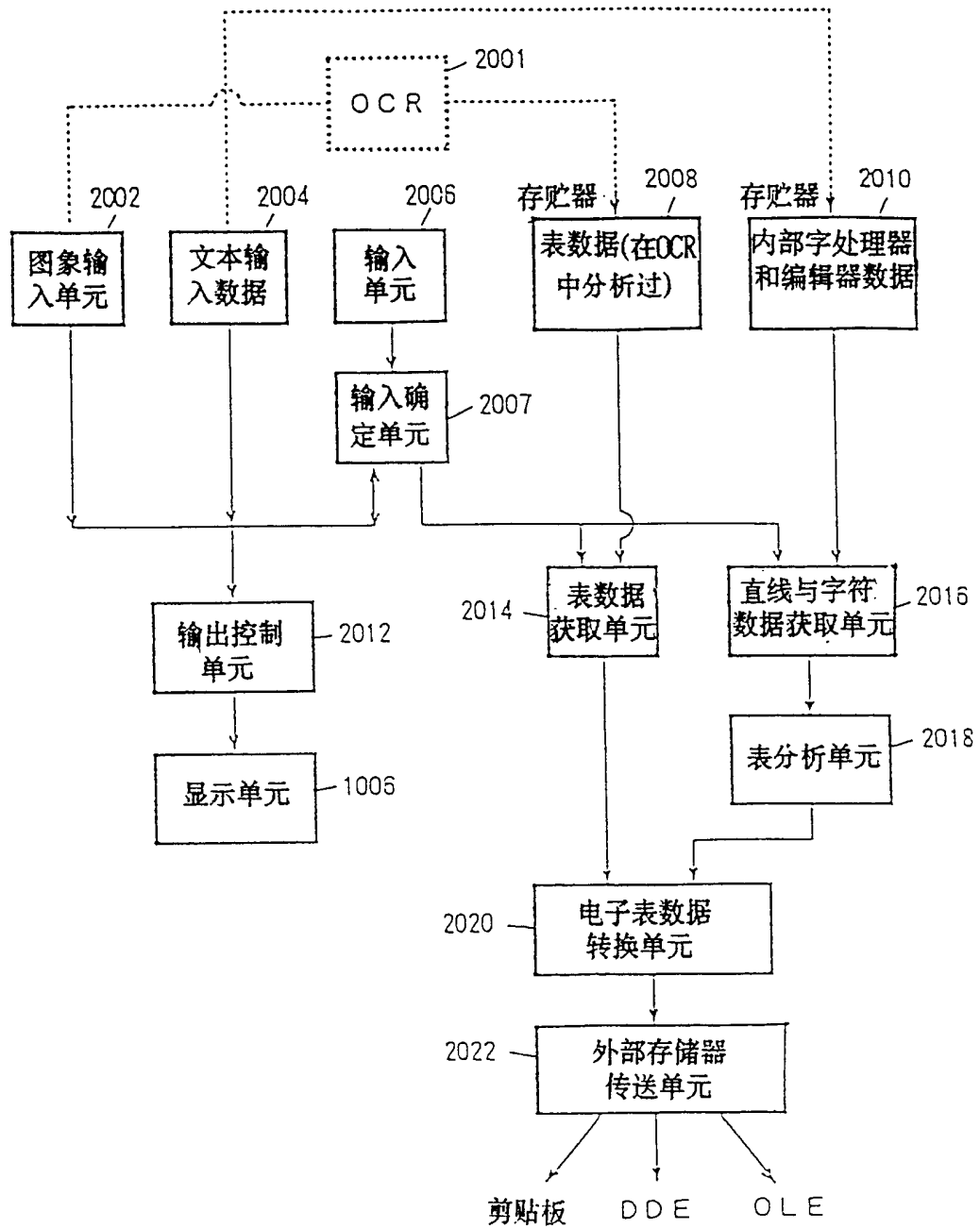
[图 1]



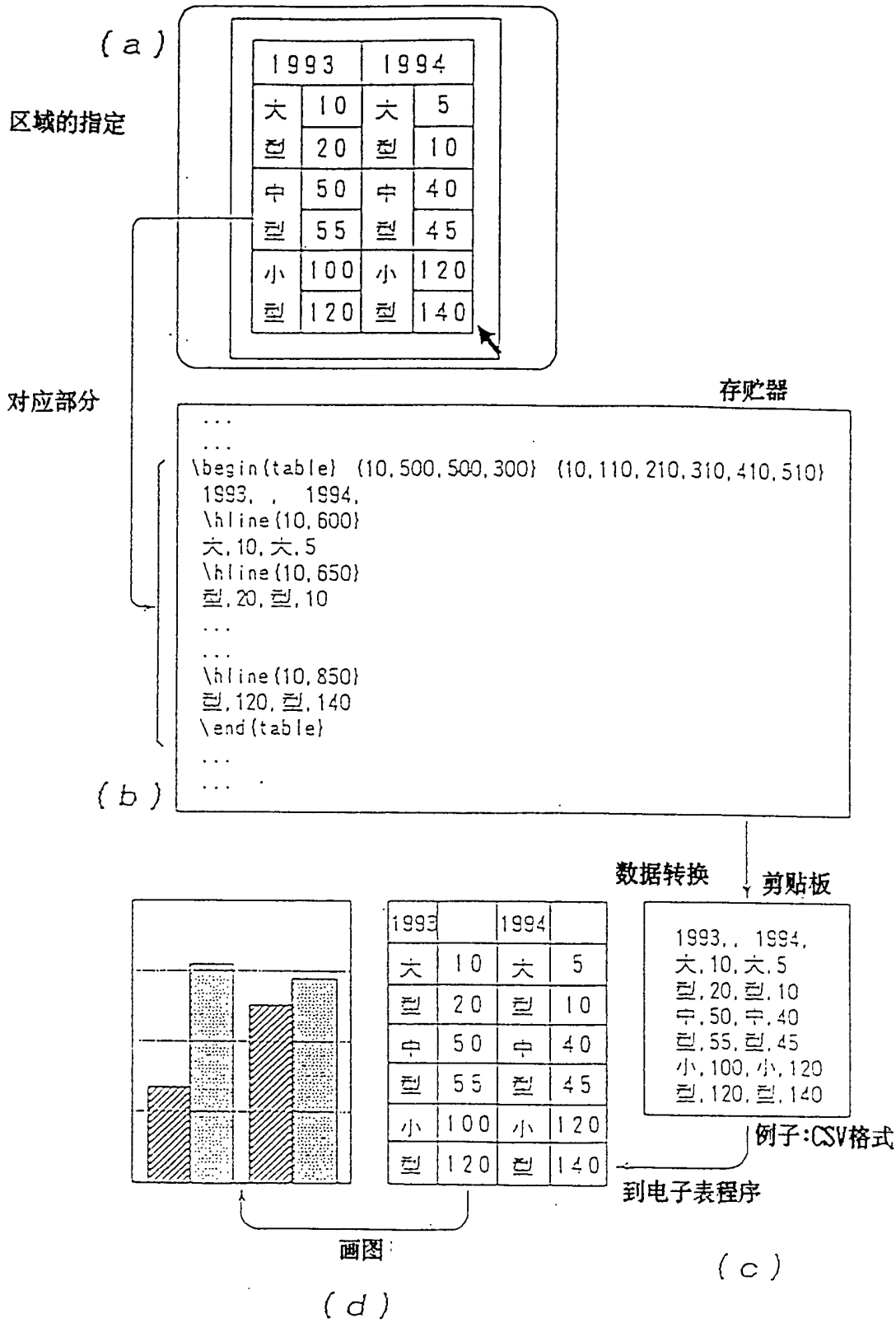
[图 4]



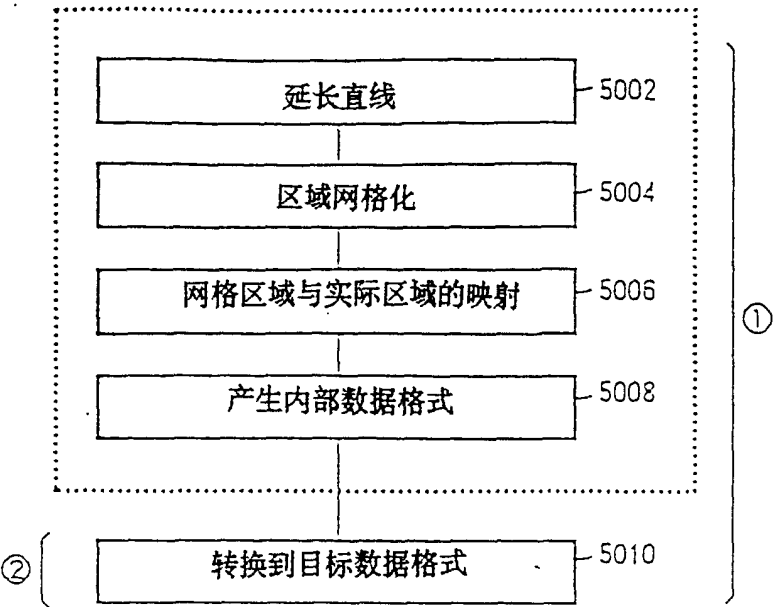
[图 2]



[图 3]



[图 5]



[图 8]

A	B	C
D	E	
F	G	
H	I	J
	K	L

[图 9]

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16
17	18	19	20

[図 7]

6020

スプレッドシート - Book 1								
ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(T) ツール(O) データ(D) ウィンドウ(W) ?								
	A	B	C	D	E	F	G	H
1	調査年度	コンピュータ部門委員			一般社員			
2		委員数	1社当た	教育費	1人当た	社員数	1社当た	教育費
3		費記入数	り平均額	委員数	り平均額	費記入数	り平均額	社員数
4			(千円)	総記入数(千円)			(千円)	総記入数
5	1992	547	3,333.50	541	76.3	266	7,775.20	5,256
6	1991	565	3,626.40	553	76.5	263	5,883.20	5,006
7	1990	439	2,389.70	434	60.2	224	6,218.40	4,864
8	1989	503	2,616.30	493	58.1	249	5,314.70	4,324
9	1988	450	1,942.20	445	47	239	5,029.10	4,025
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
	Sheet1	Sheet2						

文書(D) 表示(V) 挿入(R) 設定(S) ウィンドウ(W)							
調査年度	コンピュータ部門委員				一般社員		
	委員数	1社当た	教育費	1人当た	社員数	1社当た	
	費記入数	り平均額	委員数	り平均額	費記入数	り平均額	
		(千円)	総記入数	(千円)		(千円)	
1992	547	3,333.50	541	76.3	266	7,775.20	
1991	565	3,626.40	553	76.5	263	5,883.20	
1990	439	2,389.70	434	60.2	224	6,218.40	
1989	503	2,616.30	493	58.1	249	5,314.70	
1988	450	1,942.20	445	47	239	5,029.10	

これは、サンプルテキストです。 本日は、晴天なり。 明日は、晴天の真昼でしょう。 本日は、晴天なり。 明日は、晴天の真昼でしょう。 本日は、晴天なり。 明日は、晴天の真昼でしょう。 本日は、晴天なり。 明日は、晴天の真昼でしょう。 本日は、晴天なり。 明日は、晴天の真昼でしょう。 本日は、晴天なり。 明日は、晴天の真昼でしょう。 本日は、晴天なり。 明日は、晴天の真昼でしょう。 本日は、晴天なり。	これは、サンプルテキストです。 本日は、晴天なり。 明日は、晴天の真昼でしょう。 本日は、晴天なり。 明日は、晴天の真昼でしょう。 本日は、晴天なり。 明日は、晴天の真昼でしょう。 本日は、晴天なり。 明日は、晴天の真昼でしょう。 本日は、晴天なり。 明日は、晴天の真昼でしょう。 本日は、晴天なり。 明日は、晴天の真昼でしょう。 本日は、晴天なり。 明日は、晴天の真昼でしょう。 本日は、晴天なり。
---	---

[图 10]

$A=\{1\}$	$G=\{12\}$
$B=\{2,3\}$	$H=\{13,17\}$
$C=\{4\}$	$I=\{14\}$
$D=\{5,6\}$	$J=\{15,16\}$
$E=\{7,8\}$	$K=\{18,19\}$
$F=\{9,10,11\}$	$L=\{20\}$

[图 11]

"A", "B", "C"
"D", "E"
"F", "G"
"H", "I", "J"
"K", "L"

[图 12]

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
1	A	B		C
2	D		E	
3	F			G
4	H	I	J	
5		K		L

[图 13]

	Change in height		Change in weight	
Tanaka	170.5	172.5	65.5	66.0
Yamada		170.2	63.5	64.0
Shimizu	172.3	173.5	67.5	67.0
Mean	171.4	172.1	65.4	65.7

[图 14]

	Change in height		Change in weight	
Tanaka	170.5	172.5	65.5	66.0
Yamada	↕	170.2	63.5	64.0
Shimizu	172.3	173.5	67.5	67.0
↕	↕	↕	↕	↕
Mean	171.4	172.1	65.4	65.7

[图 15]

	Change in height		Change in weight	
Tanaka	170.5	172.5	65.5	66.0
Yamada		170.2	63.5	64.0
Shimizu	172.3	173.5	67.5	67.0
Mean	171.4	172.1	65.4	65.7

[图 16]

	Change in height		Change in weight	
Tanaka	170.5	172.5	65.5	66.0
Yamada		170.2	63.5	64.0
Shimizu	172.3	173.5	67.5	67.0
Mean	171.4	172.1	65.4	65.7

[图 17]

	Change in height		Change in weight	
Tanaka	170.5	172.5	65.5	66.0
Yamada		170.2	63.5	64.0
Shimizu	172.3	173.5	67.5	67.0
Mean	171.4	172.1	65.4	65.7

[图 18]

	Change in height		Change in weight	
Tanaka	170.5	172.5	65.5	66.0
Yamada		170.2	63.5	64.0
Shimizu	172.3	173.5	67.5	67.0
Mean	171.4	172.1	65.4	65.7

[图 19]

	Change in height		Change in weight	
Tanaka	170.5	172.5	65.5	66.0
Yamada		170.2	63.5	64.0
Shimizu	172.3	173.5	67.5	67.0
Mean	171.4	172.1	65.4	65.7

[图 20]

	Change in height		Change in weight	
Tanaka	170.5	172.5	65.5	66.0
Yamada		170.2	63.5	64.0
Shimizu	172.3	173.5	67.5	67.0
Mean	171.4	172.1	65.4	65.7