



[12] 发明专利说明书

专利号 ZL 200310119659.0

[45] 授权公告日 2006 年 11 月 8 日

[11] 授权公告号 CN 1284107C

[22] 申请日 2003.11.27

[21] 申请号 200310119659.0

[30] 优先权

[32] 2002.11.27 [33] GB [31] 0227692.1

[71] 专利权人 索尼英国有限公司

地址 英国瑟里郡

[72] 发明人 J·R·托尔普

审查员 于 平

[74] 专利代理机构 中国专利代理(香港)有限公司
代理人 程天正 王 勇

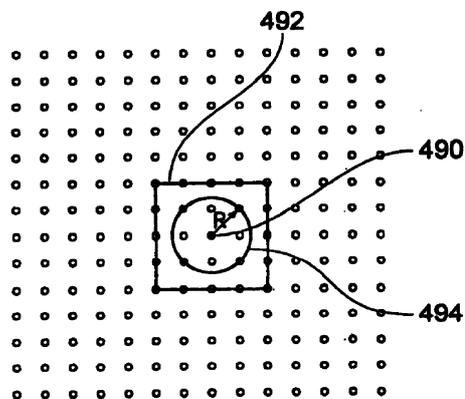
权利要求书 2 页 说明书 21 页 附图 17 页

[54] 发明名称

用于搜索一组信息项的信息检索设备和方法

[57] 摘要

搜索一组信息项的信息检索设备，包括映射处理器、图形用户接口、用户控制器及搜索处理器。映射处理器用于生成代表来自一组信息项中的映射的数据。该映射依据信息项相互间的相似性即相似的信息项映射到阵列中相似的位置就阵列中的位置提供信息项。图形用户接口显示至少一些信息项的描述，用户控制器选择信息项。搜索处理器通过识别与阵列中的位置相对应的信息项来针对用户已选信息项作相关搜索，该位置是与对应于用户选择的信息项的阵列位置相邻的位置。搜索处理器设置成在搜索中识别阵列中的信息项而不是通过针对诸如关键词的表征信息特征搜索信息项而识别信息项，故能实现对感兴趣信息项的搜索且降低复杂性。



1. 一种用于搜索一组信息项的信息检索设备，该设备包括：
一个映射处理器，它用于产生代表来自一组信息项中的信息项的映射的数据，该映射依据该信息项相互间的相似性来提供与一个阵列
5 中位置相关的信息项，相似的信息项映射到阵列中相似的位置；
一个用于显示至少一些信息项的描述的图形用户接口；以及
一个用于选择信息项的用户控制器，以及
一个通过识别与阵列中的位置相对应的信息项来针对用户选择的信息项进行相关搜索的搜索处理器，所述的阵列中位置是与对应于用
10 户选择的信息项的阵列位置相邻的位置。
2. 如权利要求1所述的信息检索设备，其中所述搜索处理器响应一个搜索查询以便根据该搜索查询来搜索该组信息项，并识别与该搜索查询相对应的信息项，并且所述映射处理器响应对该搜索查询的搜索结果来产生由该搜索处理器因对该搜索查询进行搜索而识别的信息
15 项的映射数据。
3. 如权利要求1或2所述的信息检索设备，其中，所述图形用户接口用于将与已被识别的信息项相对应的阵列的至少一些位置的描述显示为显示区域内显示点的 n 维显示阵列。
4. 如权利要求3所述的信息检索设备，其中，维度数 n 为二，阵
20 列中的位置由 x ， y 坐标定义。
5. 如权利要求4所述的信息检索设备，其中，所述搜索处理器用于通过识别与阵列中的位置相对应的信息项来针对用户已选信息项进行相关搜索，所述的位置位于始于与用户选择的信息项相对应的阵列位置的位置半径内。
- 25 6. 如前述任一项权利要求的信息检索设备，其中，用户控制器用于为用户提供根据在相关搜索中通过搜索处理器搜索的信息项的相似性针对感兴趣的阵列位置指定相邻的位置的数量的功能。
7. 一种搜索一组信息项的方法，该方法包括：
生成代表来自一组信息项中的信息项映射的数据，该映射依据信
30 息项相互间的相似性来提供与一个阵列中的位置有关的信息项，相似的信息项映射到阵列中相似的位置；
显示至少一些信息项的描述；

选择一个信息项；以及

通过识别与阵列中的位置相对应的信息项来针对用户选择的信息项进行相关搜索，所述的阵列中的位置是与对应于用户选择的信息项的阵列位置相邻的位置。

5 8. 如权利要求7所述的方法，包括

根据搜索查询来搜索该组信息项；

识别与搜索查询相对应的信息项，并且所述生成映射数据包括：

生成由所述相关搜索因对该搜索查询进行搜索而识别的信息项的映射数据。

10 9. 如权利要求8所述的方法，包括

将与已被识别的信息项相对应的阵列的至少某些位置的描述显示为显示区域内显示点的 n 维显示阵列。

10. 如权利要求9所述的方法，其中维度数 n 为二，阵列中的位置由 x ， y 坐标定义。

15 11. 如权利要求10所述的方法，其中，执行相关搜索包括通过识别与阵列中的位置相对应的信息项来针对用户已选信息项进行相关搜索，所述的阵列中的位置位于始于与用户选择的信息项相对应的阵列位置的位置半径内。

20 12. 如权利要求7至11中任意一项所述的方法，其中，为用户提
供根据在相关搜索中要搜索的信息项的相似性针对感兴趣的阵列位置
指定位置半径的功能。

用于搜索一组信息项的信息检索设备和方法

技术领域

5 本发明涉及信息检索设备及方法。

背景技术

10 目前，许多既定的系统都是利用搜索关键字来定位信息（如：文档、图像、电子邮件、专利、因特网内容和媒体内容，如音频/视频内容）。例如，“Google”™或“Yahoo”™提供的因特网搜索引擎就是利用关键字来实现搜索并根据搜索引擎所感知的相关度来排列搜索结果列表。

15 然而，在一个包含大量内容通常被称作海量内容集合的系统中，制定一个能够给出相对较短的搜索“命中”列表的有效搜索查询是非常困难的。例如，在准备本申请时，在Google上搜索关键字“海量文档集合”提取到了243000个命中结果。如果在晚些时候重复这一搜索，命中结果数字将会增加，因为互联网上所存储的内容的数量通常都随时间而增加。浏览这样的命中结果列表是非常耗费时间的。

通常，海量内容集合不能被很好的利用的原因主要有以下几个方面：

- 20
- 用户不知道相关内容存在；
 - 用户知道相关内容存在但是不知道该内容在哪里；
 - 用户知道内容存在但是不知道它是相关的；
 - 用户知道相关内容存在并知道如何找到它，但是寻找该内容所花费的时间太长。

25 Kohonen 等人在 2000 年 5 月 IEEE 神经网络学报第三期第 11 卷第 574 - 585 页发表的论文“海量文档集合的自组织”中公开了一种使用被称为“自组织映射”（SOM）的技术。该论文作者利用了一种被称为无人管理自学习神经网络的算法，其中，代表每个文档属性的“特征向量”都被映射到一个 SOM 的节点上。

30 在 Kohonen 等人的论文中，第一步是对文档正文进行预处理，而后从每个预处理文档中得出一个特征向量。在一种形式中，它可以是显示大词典中每一个单词出现频率的直方图。直方图中的每个数据值

(也就是字典中每个单词相应的出现频率)都是一个 n 值向量中的值, 其中 n 是字典中候选单词的总数(该论文所举示例为 43222)。为了强调对某个单词所增加的相关度或所有所提高的区分度, 可以对 n 向量值进行加权。

5 而后, n 值向量都被映射到更小维度的向量上(也就是具有 m 值的向量(该文中以 500 为例), m 远小于 n)。这可以通过将该向量同一个由随机数阵列(array)组成的($n \times m$)“投影矩阵”相乘来实现。业已显示出这一技术可用来产生更小维度的向量, 其中, 任何两个已被缩减维度的向量与两个相应的输入向量具有相同的向量点积。这种
10 向量映射过程公开于 Kaski 的载于 1998 年 Proc IJCNN 第 413-418 页题为“用随机映射进行维度缩减: 快速聚类相似计算”的论文中。

被缩减维度的向量通过将每个向量乘以一个“模型”(另一个向量)的方法被映射到 SOM 的节点(或称为神经元)上。这些模型由一个学习过程产生, 该学习过程将这些模型自动地按照其相互间的相似
15 性排列到 SOM 上, 这通常被表述为节点的二维网格。该过程并不简单, 它令 Kohonen 等人为一个不足 7000,000 文档的文档数据库在配备 6 个处理器及 800MB 内存的电脑上花费了六周时间。最后, 形成 SOM 的节点网格被显示出来, 用户可以放大映射区域并选择节点, 这将导致用户接口提供一个网页的连接, 该网页包含连接到该节点的文档。

20 发明内容

有关本发明的各个方面和特点将在附随的权利要求中限定。

根据本发明的一个方面, 提供一种用于搜索一组信息项的信息检索设备。该设备包括一个映射处理器, 以便从一组信息项中产生表示
25 信息项映射的数据。该映射根据信息项相互间的相似性来提供就阵列中的位置被识别的信息项, 相似的信息项映射到阵列中相似的位置上。一个图形用户接口可用于显示至少某些信息项, 并且, 提供一个用户控制器来选择一个已被识别的信息项。一个搜索处理器可以通过下述方法就用户已选择的信息项执行相关搜索: 根据与用户所选择的信息项相对应的阵列位置来识别那些与阵列中相邻的位置相对应的信息项。
30 由于该搜索处理器被设置成在搜索中根据阵列而不是通过搜索信息项的某些诸如关键字之类的字符信息特征来识别信息项, 因此寻找感兴趣的信息项将会因减少了复杂性而更有效。

本发明实施例所体现的一个优点是，如果用户已经从一组信息项中识别出了一个感兴趣的信息项，则具备重大相关性的信息项也将提供给用户。这些相关的信息项是通过根据阵列中这样的位置来识别项目从而提供的，所述位置是从阵列中与感兴趣的信息项相对应的位置中得出的预定数量的位置。

在另一实施例中，搜索处理器可根据搜索查询来搜索信息项并识别与搜索查询相对应的信息项。所述映射处理器可用来产生被搜索处理器因根据搜索查询进行了搜索而识别出的信息项的映射数据。由此，搜索处理器就可以执行初始搜索来识别与一个特定搜索查询相对应的信息项。由于有这样的搜索，故用户可以识别一个感兴趣的信息项。因此，本发明的一个实施例提供了一种能使得用户可以依照一个“相关查找”的选项进行搜索的功能，该功能可识别这样的信息项，这些信息项对应于与感兴趣的信息项相对应的阵列位置的预定数量位置中的阵列位置。为此，所述用户控制器提供启动相关查找搜索的功能。

所述图形用户接口可用来将一些与已被识别的信息项相对应的阵列位置在显示区域内显示为 n 维显示点的显示阵列。最佳的是，为了便于观看并导航，阵列的维度可以是 2。从而，阵列中的位置可由 x , y 坐标来定义。因此，在一些实施例中，搜索处理器可通过识别信息项来执行相关的搜索，该信息项对应于阵列中这样的位置，这些位置位于一个圆内，该圆相对与用户选择的信息项相对应的阵列位置具有一定的位置范围。

为了根据所需要的相关项目的范围启动相关搜索，用户控制器可设置成能向用户提供根据与搜索处理器在相关搜索中要搜索的信息项的相对相似性来限定上述位置半径的功能。

本发明其它相应方面和特征在附属的权利要求中限定。

附图说明

以下，将仅通过举例的方式参考下面的附图，描述本发明的实施方式，附图中：

- 图 1 概略地说明了一个信息存储及检索系统；
- 图 2 是示出了产生自组织映射 (SOM) 的概略流程图；
- 图 3a 和 3b 概略地说明了词频直方图；

- 图 4a 概略地说明了一个原始的特征向量；
图 4b 概略地说明了一个被简化的特征向量；
图 5 概略地说明了一个 SOM；
图 6 概略地说明了一个随机处理过程；
5 图 7 至 9 概略地说明了显示屏幕，它提供了访问 SOM 所表示的信息的用户接口；
图 10 提供了本发明的一个实施例的信息检索设备的概略框图；
图 11 提供了在搜索中被识别的信息项的一种层次结构的示意图；
图 12 提供了屏幕的概略图，该屏幕提供了两个区域，以显示图 11
10 中所示层次的不同级别；
图 13 提供了范例信息项的三种表示信息特征的说明图；
图 14 提供了本发明的一个示例性实施例的用以形成搜索查询的图形用户接口的概略图；
图 15 提供了依照布尔与操作形成一个复合特征向量的概略图示；
15 图 16 说明了两个特征向量依照布尔或操作符以及第三个特征向量依据布尔非操作符的结合；
图 17 概略地说明了被识别的信息项的二维映射的一部分，它示出了依据图 16 的布尔操作符及特征向量的搜索结果；
图 18(a) 和 (b) 提供了示意性的条线图，这两个条线图提供了组成
20 搜索查询的两个视频图像的两个色彩直方图的例子，图 18(c) 提供了通过组合图 18(a) 和 18(b) 的色彩直方图而产生的示意性条线图。
图 19 概略地说明了与图 7 至 9 相对应的显示屏幕，它提供了访问由 SOM 代表的信息的用户接口。
图 20 提供了二维位置阵列的说明图，用该图来说明相关搜索；
25 图 21 提供了一个流程图，它说明了搜索处理器所执行的相关搜索的操作。

具体实施方式

- 图 1 是一个基于通用计算机的存储和检索系统 10 的示意图，所述通用计算机具有一个处理器单元 20 并包括：一个用于程序和数据的磁
30 盘存储器 30；一个连接到诸如以太网或因特网之类的网络 50 上的网络接口卡 40；一个诸如阴极射线管设备 60 之类的显示设备；键盘 70；以及，一个诸如鼠标 80 之类的用户输入设备。系统在程序控制下工作，

程序被存储于所述磁盘存储器 30 并通过例如网络 50、可移动磁盘（未示出）或者预装于磁盘存储器 30 来提供。

上述存储系统在两种通用操作模式下工作。在第一种模式下，一组信息项（以文本信息项为例）被汇集在磁盘存储器 30 上或者汇集在通过网络 50 连接的网络磁盘驱动器上并且被分类和作了索引，以便进行搜索操作。第二种操作模式是对索引和分类数据进行实际搜索。

这些实施例可以适用于多种类型的信息项。一个不排除它的适当信息类型列表包括：专利、视频素材、电子邮件、演示文稿、因特网内容、广播内容、商务报告、音频素材、图形及剪贴画、照片等等，或者是上述任意类型的组合或混合。在本说明书中，将涉及文本信息项。文本信息项可与非文本项相关联或相连接。所以，举例来说，音频和/或视频素材可与一个作为文本信息项的“元数据”相关联，元数据是一个以文本项的形式来定义上述素材的文本信息项。

信息项通常的方式加载到磁盘存储器 30 上。最好将其存储为数据库结构的一部分，这能更容易地对信息项进行检索和索引，但是，这样做并不是必需的。一旦信息和项目被这样存储起来了，用于设置它们以便其搜索的过程即由图 2 概略地示出。

应该注意，被索引的信息项不必存储在本地磁盘驱动器 30 上。信息项可以存储在通过网络 50 连接到系统 10 的一个远程驱动器上。另外，信息也可以以分布的方式存储，例如存储在因特网上的各种站点上。如果信息存储于不同的因特网或网络站点，则使用二级信息存储器在本地存储一个远程信息的“链接”（如一个通用资源指示器 URI），也许要附随一个与该链接联系的相关概要、摘要或元数据。所以，远程存储的信息只有在用户选择了相关的链接（例如从下述结果列表 260 之中）后才会被访问，尽管为了下面的技术描述而将远程存储的信息或摘要/概要/元数据或者链接/URI 看作是“信息项”。

换句话说，“信息项”的正式定义应该是一种这样的项目，特征向量是从该项目中提取并处理的（见下），以便为 SOM 提供一个映射。在结果列表 260（见下）中所示的数据可以是用户寻找的实际信息项（如果它是本地存储并且足够短从而能够方便地显示），也可以是代表和/或指向信息项的数据，例如一个或多个元数据、URI、摘要、一组关键字、一个有代表性的关键标记图像等等。这些都是“列表”操作中所

固有的，“列表”操作虽不总是但通常是指列出表示一组项目的数据。

在另一个例子中，信息项可存储于联网的工作组，例如一个研究集体或合法的厂商。一个混合的方法可以包括一些本地存储的信息项和/或一些在局域网上存储的信息项和/或一些在广域网上存储的信息项。在这种情况下，所述系统对由其他人来定位相似的工作是非常有益的，例如在一个大型多国研究发展组织中，相似的研究工作总是映射到 SOM 中相似的输出节点上（见下）。或者，在策划一个新的电视节目时，本技术可以用于通过检测具有相似的内容的先有节目来检验该节目的独创性。

还应注意，图 1 的系统 10 仅仅是一个利用作过索引的信息项的可能的系统的一个例子。虽然可以想象初始（索引）阶段可由一个相当强大的计算机来实现，最可能用非便携式电脑来实现，但随后访问信息的阶段可以在诸如“个人数字助理”（此术语用于通常适于手持的具备显示器和用户输入设备的数据处理设备）之类的便携式机器、诸如膝上电脑之类的便携式电脑、甚至诸如移动电话之类的设备、视频编辑设备或摄影机上实现。通常，任何具备显示器的设备在实际上都可被用于信息访问阶段的操作。

上述过程并不限于作特定数目的信息项。

以下参照图 2 至 6 说明产生信息项的自组织映射（SOM）的过程。图 2 是说明其后是一 SOM 映射过程的所谓的“特征抽取”过程的概略流程图。

特征抽取是一个将原始数据转换成摘要表示的过程。而后这些摘要表示可用在诸如模式分类、聚类以及识别之类的过程中。所谓的“特征向量”就产生于这一过程之中，“特征向量”是一个文档中所使用的术语的频率的摘要表示。

通过创建特征向量形成可视显示的过程包括：

- 创建术语的“文档数据库字典”
- 基于“文档数据库字典”为每一个文档创建“术语频率直方图”
- 利用随即映射减少“术语频率直方图”的维度
- 创建信息空间的二维可视显示。

更细致地考虑上述步骤，依次打开每一文档（信息项）100。在步骤 110，从文档中删除所有“停止”。停止词是在预先准备的列表中的

非常普通的词，例如，“a”、“the”、“however”、“about”、“and”以及“the”。由于这些词非常普通，故在所有足够长的文档中它们平均出现的频率很可能是相似的。基于这一原因，这些词对于表征一个特定文档的内容不起作用，从而需要被删除。

- 5 在删除停止词之后，剩余单词在步骤 120 中提取词干，这一步骤包括寻找单词变化形式的共同词干。例如单词“thrower”、“throws”，以及“throwing”具有共同的词干“throw”。

保持文档中出现的词干词的“字典”（不包括“停止”词）。在遇到新词时，将其加入到字典中，同时将该词在整个文档集合（一组
10 信息项）中出现次数的连续计数值记载下来。

结果是一个上述集合中所有文档都使用的术语列表以及这些术语出现的频率。出现频率过高或过低的单词将被扣除，这就是说，它们将从字典中被删除并且不参与随后的分析。频率过低的单词可能是拼写错误、编造的，或者与该文档集合所代表的领域不相关。而出现频率
15 过高的单词不适于用来辨别上述集合中的文档。例如，在由与广播有关的文档组成的测试集合中，所有文档中大约三分之一都使用了单词“News”，而在测试集合中仅有大约 2% 的文档使用了单词“football”。因而“football”可被认定为是比“News”更好的用于表征文档内容的单词。相反，单词“fottball”（“football”的错拼）仅
20 在整个文档集合中出现一次，由于出现频率过低因而被删除。这样的单词可被定义为具有比小于平均出现频率的两个标准偏差更低的出现频率，或具有比大于平均出现频率的两个标准偏差更高的出现频率。

然后，在步骤 130 产生一个特征向量。

为此，要为集合中每个文档产生一个术语频率直方图。术语频率
25 直方图的构建是通过计算字典中存在的单词（属于该文档集合）在各个文档中出现的次数的方法来实现的。字典中的多数术语都不会在单个的文档中出现，因此这些术语的频率是 0。图 3a 和 3b 示出了两个不同文档的术语频率直方图的示例。

从本例中可以看出直方图是如何表征文档内容的。通过检查本
30 例，可以看出文档 1 中术语“MPEG”和“Video”的出现次数比文档 2 多，而文档 2 中术语“MetaData”的出现较多。由于相应的单词没有在文档中出现，直方图中许多条目都是 0。

在真实的例子中，实际的术语频率直方图具有比上例大得多的术语数量。一般地说，一个直方图可标绘出超过 50000 个不同术语的频率，赋予直方图大于 50000 的维度。如果需要在构建 SOM 信息空间时使用，这种直方图的维度就需要大幅减少。

5 术语频率直方图中的每个条目都被作为代表该文档的特征向量的相应值。这一过程的结果是一个包含字典为文档集中的每个文档所指定的所有术语频率的 (50000x1) 向量。由于大部分值一般都是 0，并且其他大部分值一般都是比较低的数字例如 1，故该向量可被视为是“稀疏的”。

10 特征向量的大小以及术语频率直方图的维度在步骤 140 中被缩减。就缩减直方图维度过程提出了两种方法。

i) 随机映射 - 使直方图乘以一个随机数矩阵的技术。这是一个容易计算的过程。

15 ii) 潜在语义索引 - 通过查找在文档中具有同时出现的高概率的术语组来缩减直方图的维度的技术。这些单词组可被缩减为单个的参数。这是一个计算费用昂贵的过程。

在本实施例中所选择的缩减术语频率直方图的方法是“随机映射”法，如上面提到的 Kaski 的论文所述。随机映射的成功在于它与随机数矩阵相乘的方法来缩减直方图维度。

20 如上所述，“原始”的特征向量（在图 4a 中概略地示出）一般是一个约有 50000 值大小的稀疏向量。它可被缩减到大概 200 值的大小（如概略图 4b 所示）并仍然保留特征向量相对正交的特征，即保留了诸如与其他同样处理后的特征向量的相对交角（向量点积）之类的关系。这是因为，虽然特定维度的正交向量数量受到限制，但接近正交向量的数目却非常之大。

事实上，由于向量维度的增加，任何给出的随机产生的向量组都接近于相互正交。这一性质意味着将保留向量的相对方向与一个随机数矩阵的乘积。这可通过借助察看向量点积而展示向量在随机映射之前或之后的相似性来得到论证。

30 在实验上可以看出，通过将稀疏向量从 50000 值缩减到 200 值而保留了其相对的相似性。然而，这一映射并非完美，但对达到以简洁的方法表征文档内容的目的来说却足够用了。

一旦形成了用于文档集合的特征向量以便限定集合的信息空间，就在步骤 150 将它们投射进一个二维 SOM 以建立一个语义映射。接下来的部分说明利用通过 Kohonen 自组织映射对特征向量聚类的从而映射为 2 维的过程。仍参照图 5。

5 Kohonen 自组织映射被用来对为每个文档已产生的特征向量进行聚类和组织。

一个自组织映射由在二维阵列或节点网格中的输入节点 170 以及输出节点 180 来构成，所述二维阵列或节点网格如二维平面 185 所示。输入节点的数目与形成映射时所使用的特征向量中的值是相等的。映射中的每一个输出节点都通过加权连接 190 (每个连接加一个权值) 与输入节点相连接。

15 初始时这些权重都被设为一个随机值，而后，通过一个迭代过程来“调整”权重。映射是通过为映射输入结点提供每个特征向量来进行调整的。“最接近”的输出结点是通过计算输入向量与和每个输出结点相关的权重间的欧几里得距离而计算出来的。

由输入向量与和上述节点相关的权重间的最小欧几里得距离所标识的最接近的节点被指定为“胜利者”，并且，该节点的权重通过稍微改变该权重的值而得以调整，由此，它们会移动成“更靠近”输入向量。除了胜出的节点之外，与胜出节点相邻的节点同样也被调整，并移动成略微更靠近输入向量。

20 正是这种不是仅调整单个节点的权重而是调整映射上节点区域的权重的过程，使得该映射一旦被调整就能保留了 2 维节点映射中输入空间的大部分拓扑结构。

25 映射一旦被调整，每个文档可被提供给映射以便看出哪个是最靠近用于所述文档的输入特征向量的输出节点。权重不可能等同于特征向量，而且，一个特征向量与映射上与之最近的节点之间的欧几里得距离被认为是“数量误差”。

30 通过在映射上显示每个文档的特征向量来决定文档的位置，会为每个文档产生一个 x, y 映射位置。这些 x, y 位置在被放进带文档 ID 的查找表时可用于使文档之间的关系可视化。

最后，在步骤 160 加入一个随机处理成分，下面将参考图 6 进行描述。

上述过程中存在的一个潜在问题是，两个同样的或者基本上相同的信息项可能被映射到 SOM 节点阵列中的同一个节点上。这不会导致处理数据的困难，但是却无助于数据在显示屏上的可视化（下述）。具体地说，当数据在显示屏是可视化的时，业已认识到，区分一个特定节点上的单个项目中的多个非常相似的项目是有用的。因而，一个“随机处理”成分被附加到每个信息项所映射到的节点位置中。所述随机处理成分是将节点间隔随机增加 $\pm 1/2$ 。因而，参考图 6，映射过程为其选择了输出节点 200 的信息项附加了一个随机处理成分，从而它实际上可以被映射到图 6 中节点 200 周围虚线 210 区域内的任何映射位置。

因此，信息项可被认为映射到图 6 的节点位置的平面位置上而不是 SOM 过程的输出节点上。

在任何时候，一个新的信息项都可通过上述的步骤（即步骤 110 至 140）被加到 SOM 中并将最终缩减的特征向量应用到“预调整”的 SOM 模型即由映射的自组织预加工而产生的 SOM 模式组。因此，通常映射不会为新加入的信息项而进行重新调整；而是在不修正所有 SOM 模型的情况下使用步骤 150 和 160。每次在新的信息项加入时都对 SOM 进行重新调整的计算代价是昂贵的，并且对那些已经习惯于映射中通常访问的信息项的相对位置的用户也不友好。

然而，存在有进行适当重新调整的时候。例如，如果自 SOM 产生时首次形成以来有新的术语（或许是新的新闻项目或是新的技术领域）都被加入到字典中，则它们可能不会特别好地映射到现存输出节点集合上。这可被检测为所谓“数量误差”有增加，上述“数量误差”是将新接收的信息项映射到现存 SOM 中时检测到的。在本实施例中，将数量误差与阈值误差值作比较。如果大于该阈值，则（a）利用所有原始信息项以及自产生以来增加的任何项目自动调整 SOM；或者（b）提示用户在方便的时候启动重新调整过程。重新调整过程利用所有相关信息项的特征向量并全部再利用步骤 150 和 160。

图 7 概略地说明了在显示器 60 上的显示。该显示示出了一个搜索查询 250、一个结果列表 260 以及一个 SOM 显示区域 270。

在操作上，初始时，显示区域 270 是空。用户向查询区域 250 键入一个关键字搜索查询。用户继而启动搜索，例如通过在键盘 70 上按

输入键或通过使用鼠标 80 选择屏幕“按钮”来启动搜索。利用标准的
关键字搜索技术，将搜索查询区域 250 中的关键字与数据库中的信息
项进行比较。这将产生一个结果列表，每一个结果都显示为列表区域
260 中的一个相应的条目 280。而后，显示区域 270 显示与各结果项目
5 相对应的显示点。

由于用于产生 SOM 画面的分类过程会把 SOM 中相似的信息项聚合
在一起，因此搜索查询的结果会聚集在诸如类 290 之类的类中。在这
里，应该注意，区域 270 中每一个点都与在 SOM 中且与结果列表 260
中的结果相关联的相应条目相对应；并且，在区域 270 中点被显示的
10 位置与那些节点在节点阵列中的阵列位置相对应。

图 8 概略地说明了减少命中（结果列表中的结果）数目的技术。
用户利用鼠标 80 划出一个边界，在本例中为区域 270 所显示的显示点
集合周围的长方型 300。在结果列表区域 260 中，只有那些与边界 300
之内的点相对应的结果才被显示出来。如果这些结果不是所感兴趣的，
15 则用户可以再划定另一个包含不同显示点集合的边界。

应该注意，结果区域 260 针对那些在边界 300 之内为其显示出显
示点并且符合单词搜索区域 250 中搜索标准的结果而显示出列表条
目。边界 300 可以包括其它与驻留在节点阵列中的节点相对应的显示
位置，但是，如果这些不符合搜索标准，那么它们将不会被显示，从
20 而不会构成列表 260 中所示结果的一个子集的一部分。

图 9 说明了本发明的一个实施例。

参考图 9 和步骤 920，自组织映射 SOM 在产生出来时是没有标记的
（不同于 Kohonen 的 SOM）。用户需要标记来进行导航以便浏览映射。
在本发明的实施例中，标记是自动产生的，用以符合用户特定的需要。
25 如参考图 7 和/或图 8 所述，用户产生一个搜索的结果列表。标记根据
上述结果自动动态地产生并且用于标记区域 270 中的显示点群。

交叉类（cross-cluster）联合/辅助关键字搜索

下面将参考图 10、11 和 12 对本发明的一个示例性实施例进行描
述。

30 在图 10 中，一个包含信息项数据库的数据仓库 400 通过数据通讯
网络 410 与搜索处理器 414 以及映射处理器 412 相连。映射处理器与
用户控制器 414 以及显示处理器 416 相连。显示处理器 416 的输出由

图形用户接口 418 来接收，后者与显示器 420 连接。为在显示屏上进行显示，所述显示处理器 416 可用于处理来自映射处理器的数据。

数据仓库 400 可设置成独立于映射处理器 412。相应地，搜索处理器也可以设置成独立于数据仓库 400、映射处理器 412 以及图 10 所示的那些用于显示信息的部分，它们是显示处理器 416、图形用户接口 418 以及显示器 420。另外，映射处理器 412、搜索处理器 404 以及显示处理器 416 可按在诸如图 1 所示的通用计算机上执行的软件模块方来实现。然而，应该认识到，映射处理器、搜索处理器以及显示处理器可以分别产生和设置。

图 10 所示的实施例基本上是作为在图 1 中并结合图 7, 8 和 9 所示出的存储和检索数据处理器来进行操作的。图 7、8 和 9 提供了如何针对搜索查询搜索信息项以及如何显示搜索结果的示意图。相应地，图 10 所示的实施例被设置成能接收一个搜索查询，例如一个来自用户控制器 414 的关键字。搜索处理器 404 响应该关键字进行搜索，以便结合映射处理器识别阵列中 x, y 位置集合，该 x, y 位置与作为搜索结果被而识别的信息项相对应。例如，对于一个 40×40 的节点阵列，在方形二维阵列中有 1600 个位置。如上所述，搜索处理器根据搜索查询来搜索信息项。搜索处理器的搜索将导致符合搜索查询的信息项的 x, y 位置集合被搜索处理器所识别。所述搜索结果的 x, y 位置由映射处理器 412 接收。

在一个实施例中，搜索处理器 404 可设置成能搜索信息项并产生可识别出与搜索查询相对应的信息项的搜索结果。映射处理器 412 可以接收代表搜索结果的数据，该搜索结果能识别出符合搜索查询的信息项。映射处理器而后产生符合被识别的信息项在阵列中的位置的 x, y 坐标。

映射处理器 412 可通过一个 k -means 聚类 (clustering) 过程在第一全局层次上识别信息项的类。 k -means 聚类过程识别阵列中的类和类的位置。 k -means 聚类过程公开于由牛津大学出版社出版的 Christopher M. Bishop 所著《用于模式识别的神经网络》一书的 187 - 188 页。对于 k -means 聚类算法的其它说明公开于网址：

http://cne.gmu.edu/modules/dau/stat/clustgalgs/clust5_bdy.html

如图 11 所示，关键字“show”的搜索结果可以识别出阵列中那些与以单词“show”作为其一部分元数据的信息项相对应的位置。因此，对阵列执行 k-means 聚类算法的结果将例如识别出“quiz”、“game”以及“DIY”三类信息项。这些信息项的类形成了第一层级 h_level1。显示处理器 416 从映射处理器 412 接收符合第一层级 h_level1 的信息项类的数据。为提供代表第一层级 h_level1 的二维显示的数据，显示处理器 416 处理数据的第一层级。由显示处理器 416 所产生的数据提供给图形用户接口 418，以便在显示屏 420 的第一区域 430 上进行显示，如图 12 所示。

在一些实施例中，其它操作可以通过映射处理器 412 来执行，以使用 k-means 算法来改进对类的识别。上述其它操作被称为“k-means 聚类 and 剪枝”。已知的 k-means 聚类过程可标识在的表示相似信息项的搜索结果中识别出信息项的阵列位置。而后执行其它的修剪枝过程，它用于判定结果项目的 x, y 位置的相邻子类是否是同一母类的一部分。如果两个子类中心间的距离小于一个阈值的话，则这两个子类被视为是同一母类的一部分。剪枝以已知的方式重复进行直到聚类稳定下来为止。

映射处理器 412 对在第一层级 h_level1 上识别出的每一个信息项的类进行进一步的分析。为了给用户提供用来分别检查信息项的类并在信息项中更进一步的识别类的功能，映射处理器 412 形成一个更深一层的层级。因此，对于每个信息项类来说，k-means 聚类算法都将对该类执行，以便识别在信息项第一层级之内更深层的类。例如，如图 11 所示，如果对“quiz”类执行 k-means 算法，则三个更深的类将在第二层级 h_level2 被识别。

如对第一层级所述，每一个类都根据一个关键字进行标记。该关键字通过查找最普通的单词来进行识别，所述单词是类中的每一个信息项都具有的并且在与该信息项相关的元数据中出现过。例如在第一层级中，三个类就通过单词“quiz”、“game”和“DIY”来识别。

按与标记第一层级 h_level1 的类相应的方法为第二层级 h_level2 中的每个类识别一个关键字。因此，三个类将被标记为“the chair”、“wipeout”以及“enemy within”。此三类中的每一类都包含不同的智力竞赛 (quiz show) 的情节。

正如所注意到的那样，可以进一步的对每个类进行分析迭代。这可以通过对在第二层级 h_level2 上识别出的每一个类执行 k-means 算法来实现。如图 11 所示，“wipeout”信息类用 k-means 聚类算法进行进一步分析。然而，在第三层级 h_level3，仅展示出了个别的信息项，因而，如图 11 所示，第三层级 h_level3 可识别“wipeout”的个别情节。

映射处理器 412 因而能识别不同层级上的信息项的类。将代表每个层级的数据提供给显示处理器 416。因此，结合图形用户接口 418，例如与第二层级 h_level2 相对应的第二区域可被显示于显示器 420 上。因而，用户可利用缩放控制器放大在第一层级 h_level1 显示的类。缩放控制器可通过用户控制器 414 来控制。因此，放大特定的类可以显示出第二层级 h_level2 上的信息项。另外，用户控制器 414 可被用于在第一区域中选择“当前视窗”区域。因此，针对在“quiz”类中识别出的类说明了的第二级显示，所述“quiz”类是在第一层级 h_level1 内所示的第一级显示中识别出的。

本发明实施例的又一优点是这样一种结构，其中，可以为在第二或其后显示区域中显示的第二或其后层级提供其他类的指示器。指示器可以将用户引导至另外的类，该类指向与在下一层级中观看的类相关联的关键字。因而，在第二显示区域 440 内于低层级上显示的类将具有指向正在观看的类的其它类。例如，在图 12 中，在第一显示区域 430，第一层级显示出“quiz”、“game”以及“DIY”三个类。由于缩放控制器用于放大“quiz”类，故第二显示区域 440 将显示出“quiz”类之下的类，即“the chair”、“enemy within”以及“wipeout”。然而，如第一区域所示，“quiz”类的其它关键字为“DIY”、“horror”以及“game”。因此，提供箭头 444、446 和 448 将用户引导至与第二显示区域内显示的“quiz”类在同一层级上的信息项的类。因此，如果用户希望查阅不同于第一层级的类来发现第二层级中的类，则用户可以使用箭头来导航到第一层级中的其它类。此外，箭头还最好由出现在第一层级的类的关键字标签来标记。在其它实施例中，为了给用户提供类中相关项目数的说明，这一数字显示在与方向指示箭头相关的关键字旁边。当鼠标指针 MP 经过时或定位到指示箭头上时，用户控制器以及显示器设置成能示出这一数字。

另外一些实施例的其它优点是提供一个附加关键字的列表，也即一级类内与二级类相关的关键字。如图 12 所示，对于提供其它第一层级类“horror”的类来说，当鼠标指针 MP 定位到与“horror”相关的箭头上时，将产生与第一类“horror”中的二级类相对应的附加单
5 词。因此，将给用户提供一个与首层类相关联的信息项内容的有效图示，而不需要查阅第二显示区域 440 内的那些类。如图 12 所示，显示区域还可以还包含如 450 所示的控制图标，这些图标用于对第一区域 430 中出现的信息项中进行浏览和导航。

多模式精确搜索

10 本发明的另一示例性实施例将参考图 10 以及图 13 至 17 来加以描述。图 13 提供了一个表征信息特征的类型示意图，所述信息特征连同同一个信息项来进行存储。例如，信息项可以是来自一个电视节目的视频/音频数据的一部分。在本例中，电视节目是足球赛的精彩片断。因此，数据项包括了视频数据 460 以及音频数据。与音频数据相关联
15 的是在方框 462 中示出的音频元数据。音频元数据描述了与视频数据相关联的音频信号的内容和类型。在本例中，音频数据包括“音乐”，“注释”、“噪声”，但也可以包括一个或多个其它类型的表示音频信号类型的元数据。除视频数据和音频数据以外，信息项还可以包括描述视频和音频数据内容或属性的其它元数据。在本例中，元数据在
20 方框 464 中示出，并包括视频节目内容的说明。包含在上述元数据中的单词被用于建立特征向量，根据该特征向量来产生 SOM。然而，在本发明的其它实施例中，在数据仓库 400 中的信息项目集合可以针对是音频元数据 462 的音频元数据或视频数据来进行搜索。为此根据音频数据 460 的帧来产生一个代表性关键标记。

25 该代表性关键标记 RKS 的产生是通过形成每个视频数据帧的色彩直方图来实现的。把所有的或被选择的视频帧色彩直方图结合起来，而后使之标准化以产生一个合成的色彩直方图，典型的形式如图 13 中的条线图 466 所示。将该合成色彩直方图与每一视频帧的色彩直方图相比较。通过将每一视频帧的色彩直方图中每一列的距离与合成直方
30 图中相应的柱相加来确定每一视频帧的色彩直方图与合成色彩直方图之间的距离。具有与合成色彩直方图之间距离最小的色彩直方图的代表性关键标记 RKS 被选择出来。对于足球比赛节目，相应产生的代表

性关键标记很有可能是足球赛过程中某部分的视频图像，如图 13 的代表性关键标记 RKS 所示。

在其他实施例中，可用以下任何一种方法为来自视频帧的每一个信息项产生 RKS：

5 ●用户可以选择与整个信息项的内容相对应的被认为最有代表性的一帧。由于用户能确保选择了主观上代表了一个信息项的视频帧，故这一方法可以改进可靠性。然而它却要花费更多的时间。

 ●用户可以选择某个信息项中的第一帧或随机的一帧。这对于选择合适的 RKS 是缺乏可靠性的。

10 ●可以想像出基于图像帧的内容来处理视频帧并选择一个 RKS 其他方法。

本发明的实施例可以提供了基于已选表征信息特征来产生精确搜索的功能。在一个实施例中，搜索处理器 142 可依据元数据，视频图像或音频数据之一，搜索在第一搜索中被识别的信息项。在另外的实
15 施例中，搜索可仅通过元数据或视频数据或音频数据或它任意的组合来进行。为了简化搜索查询的格式，图 10 所示的显示设备 420 可以包括一个如图 14 所示的图形用户接口 418 所提供的图形显示。

在图 14 中，在显示区域 472 中第一行 470 提供给用户选择根据元数据进行信息查询的功能。因此，如果来自某个信息项的图像代表关键标记的被置于屏幕中的这一行，则与该信息项相关的元数据（如图
20 13 所示）将被加入到搜索查询中。因此，来自不同信息项的一个或多个代表性关键标记将被引入到类型元数据的表征信息特征的搜索查询中。相应地，在第二行 474 中，用户选择的视频帧将被引入，以构成搜索查询的一部分。例如，用户可以浏览视频数据的特定项目并选择
25 感兴趣的帧。而后用户可以将该图像帧放入行 474，以构成搜索查询的一部分。用户可以引入一个或多个视频帧。

用户也可以选择根据信息项中的音频数据来进行搜索的该信息项。因此，显示区域 476 中的第三行为用户提供了引入信息项的代表图像以便在音频数据行进行识别的功能，因而，搜索查询会包括与搜索
30 查询中该信息项相对应的音频数据。

除选择根据表征信息特征的类型进行搜索的信息项外，本发明的实施例还提供了根据所选择的信息项之间的布尔操作符来实现搜索的

功能。如图 14 所示，根据图示的前两列 478，480 之间的“与”操作符来搜索业已针对元数据搜索而选择出的信息项。然而，在第一元数据和搜索查询中的第一视频图像项目之间的搜索查询是通过“或”操作符来连接的。就视频数据进行搜索的这两个项是通过“与”操作符来连接的。根据音频数据进行搜索的信息项将在搜索查询中按照“非”操作符进行搜索。

在建立好搜索查询之后，搜索处理器 404 将根据搜索查询搜索由关键字搜索识别的信息项，所述搜索查询是依据用户的选择而产生的并如图 14 所示。搜索处理器依据所选择的表征信息特征的类型以不同的方式搜索的信息项，如下段所述：

对于就如元数据之类的表征信息特征进行搜索的例子，对于任何信息项来说，从元数据产生的信息项的特征向量可被用于识别二维阵列中与该特征向量相对应的一个点。因而，阵列中在该被识别位置的预定距离之内的信息项可被作为搜索查询的结果返回。然而，如果在元数据搜索行中选择了不只一个信息项，则搜索查询就必须按着能根据所选择的布尔操作符来搜索这些项目的方式加以构建。

对于“与”布尔操作符的例子，如图 15 所示，将每个信息项的特征向量结合来以形成一个合成向量。为此，与元数据中每一个单词相关的值都被相加起来并规格化成能形成合成特征向量。因此，如图 15 所示，与用户选择的在行 470 和列 478 至 480 所示的具有代表性关键标记的元数据相关联的两个特征向量 A, B 和元数据搜索查询行 470 将结合在一起形成特征向量 C。搜索处理器接收特征向量 C 并将其与 SOM 做比较。在与合成特征向量 C 相对应的在阵列中的最近位置被识别出来之后，阵列中来自在该阵列中的被识别位置的预定数量的位置内的信息项将作为搜索结果返回。

对于相应元数据搜索的“或”布尔操作符的例子，对于第一特征向量 A 和第二特征向量 B 来说，阵列中用于这些特征向量的相应位置将被识别。同样，搜索查询将返回阵列中每个被识别的点的预定数量位置内的所有信息项。这一点由图 16 和 17 示出。在图 17 中，二维阵列中与特征向量 A 相对应以及与特征向量 B 相对应的位置将被识别。如图 17 所示，阵列中位于 A 和 B 的阵列位置预定半径之内的位置将在被识别时作为搜索查询的结果返回。然而，如果另一个特征向量 C 在

搜索查询中被识别并且对特征向量指定了“非”布尔操作符，则阵列中与该特征向量 C 相对应的位置又将被识别。因此，始于 C 的预定半径的阵列位置中的信息项也将被识别。然而，由于有“非”操作符，在特征向量 A、B 和 C 的阵列位置的半径之间被识别的那些任何的相互包含的阵列位置都被排除在搜索结果之外。因而，搜索处理器将返回由与阵列中 A 或 B 但不是 C 所产生的位置所对应的信息项。

对第二行来说，在与是搜索的表征特征的视频图像数据相对应的搜索查询中，搜索处理器可用于针对与已选择的用户视频图像相对应的代表性关键标记来搜索视频数据。为此，与用户已选视频图像相关联的色彩直方图将与同信息项相关联的每一个代表性关键标记的色彩直方图相比较。计算每一个信息项的代表性关键标记的色彩直方图与用户指定的视频图像的色彩直方图之间的距离。这可以通过计算代表该图像色彩成分的各列之间的距离并将对每一列的距离求和来实现。与信息项相对应的并且在用户选择的视频图像的直方图与和该阵列位置相对应的代表性关键标记的直方图之间距离最小的阵列位置将被识别。查询的结果将返回具有来自被识别阵列位置的预定数量位置中的阵列位置信息项。

对于布尔操作符，通过将两个已被选择和指定用于“与”操作符的图像的色彩直方图结合起来而形成一个色彩直方图。形成合成色彩直方图的过程如图 18 所示。行 474 中所提供的第一和第二用户选择的图像的直方图和如图 14 所示的显示区域内的视频图像搜索查询的列 478, 480 将通过平均该直方图每一列的值来进行合成。由此，图 18a 和 18b 所示的两个色彩直方图合成起来以形成在图 18c 中形成的色彩直方图。针对将被搜索的信息项的代表性关键标记来搜索该色彩直方图。

对于音频数据的例子，搜索处理器可以根据与所选信息项相关的音频元数据形成一个特征向量。例如，音频元数据可以识别在音频信号、语音数据中存在的谐波或者识别是否在由音频元数据所代表的音频信号中存在有音乐。另外，元数据可以识别音频信号中是否有特定的说话者如 Tony Blair 或特定的解说员如 Tony Motson。因此，特征向量也可以由所选音频数据产生，所述音频数据可针对具体与音频数据相关的其他特征向量来进行搜索。按着上述方法相应的方式，布尔

操作符可用于将一个以上的音频元数据类型的搜索合并起来。对于“与”操作符的例子，音频元数据项可被合并起来产生一个合成元数据项。寻找具有与合成项距离最近的特征向量的相应信息项将识别一个信息项。而后，当指定“或”操作符时，搜索处理器可以再现阵列中用于两个元数据项目的预定数量位置范围内的信息项。“非”布尔操作符将从搜索查询结果中排除所返回的具有匹配音频数据的信息项。

本发明的实施例提供了从已被识别的信息项中进行精确搜索。然而，应该认识到，在其他实施例中，由图 14 所示的显示所形成的搜索查询以及该搜索查询符针对元数据、视频图像数据和音频数据的应用可用来搜索数据仓库 400 中的整个信息集合。

相关搜索

如上所述，根据本发明的一个示例性的实施例，依照利用如图 14 所示的图形用户接口构建的搜索查询的信息项可通过识别与搜索查询所识别的特定阵列位置相邻的项目来进行搜索。然而，在其他示例性实施例中，可对为任何理由而识别的信息项进行相关搜索。但是，一般地说，根据特定关键字的搜索会产生一组被识别的信息项。从这些信息项中，用户可决定其中是否有一个是其特别感兴趣的。而后，相关搜索可以依据 SOM 提供与该信息有某种相关性的项目。这可以通过识别与阵列位置相对应的信息项来实现，所述阵列位置位于例如始于与所感兴趣的信息项相对应的阵列位置的预定半径内。

图 19 提供了搜索处理器 404 如何执行一个“相关查找”搜索的示意图。用户可认定一个特定信息项是其所感兴趣的。例如，图 19 再现了图 7、8 和 9 所示的图形用户接口。假定早先搜索的结果展示出了阵列中如前用黑点所述的被识别位置的结构，并且用户找到了所感兴趣的与阵列中的位置 490 相对应的特定信息项。为进行相关搜索，用户需要将鼠标指针 MP 定位于感兴趣的位置 490 上并通过菜单选项启动一个相关搜索，所述菜单选项例如可以自动出现。在启动相关搜索时，搜索处理器 404 识别那些与感兴趣的位置 490 相邻的预定数量的位置内的阵列位置相对应的信息项。例如，搜索处理器 404 可识别与正方形 492 内的阵列位置相对应的信息项，该正方形是由沿 x, y 方向加减两个位置而形成的。另外，搜索处理器 404 可以识别与一个圆之内的

阵列位置相对应的信息项，该圆具有始于所选感兴趣的信息项的阵列位置 490 的由对角线上的一个位置构成的预定半径 R。

在识别了与相关阵列位置相对应的信息项后，每个被识别信息项的表征信息特征可在显示区域 260 中被显示，如图 19 所示。

- 5 在某些实施例中，依据相关搜索赖以实现的相关敏感性，用户可以通过用户控制器改变被识别的相关信息项所对应的阵列位置数量。因此，在相关搜索中识别的预定相邻位置的数量会有所变化。这可以改变圆 494 的半径 R 或者正方形 492 的大小来实现。

- 10 基于阵列执行相关搜索而不是通过针对某种表征信息特征搜索信息项（例如关键字搜索）来执行搜索，可以提供搜索感兴趣的信息项的功能，该功能可以减少搜索关键字的计算复杂度。利用阵列的相关搜索操作可以通过 SOM 的属性而被简化，这就能确定在阵列中的相似位置处是相似的信息项。因而，具有阵列中和与感兴趣的信息项相对应的位置相邻的位置的信息项将与该信息项相关。因此，检索与这些
15 相邻位置相对应的信息项会展示出对最有可能与用户搜索要求最相一致的项目的重点搜索。

相关搜索的概要流程图

图 21 示出了概括进行相关搜索时搜索处理器的操作的流程图。相关搜索过程的步骤概括如下：

- 20 S. 2 虽然操作中的第一步可以根据用户指定的搜索查询产生代表来自信息项集合的信息项映射的数据，但是，如果从用户已识别的信息项开始来执行相关搜索，则步骤 S. 2 至 S. 10 可以被省略。映射依据信息项相互间的相似性即相似的信息项映射到阵列中相似的位置来提供与阵列中位置相关的信息项。

- 25 S. 4 根据搜索中被识别的 x , y 阵列位置或通过映射处理器而将信息项被映射到二维阵列中的位置上。

S. 8 映射数据根据阵列中的阵列位置的 x , y 位置产生，以便进行显示。

- 30 S. 10 依照映射数据至少一些信息项的代表将作为二维阵列而被显示。

S. 12 用户选择其感兴趣的一个信息项。

S. 14 用户可以指定执行相关搜索的条件。用户可以识别始于感兴

趣的信息项的阵列位置的邻居数量或者半径。

S. 16 如果用户没有指定相关搜索的特定要求，则搜索处理器自动识别预定相邻位置的数目并返回与这些位置相对应的信息项。

S. 18 如果用户指定了相关搜索的特定要求，则搜索处理器根据用户的要求识别相邻位置并返回与这些位置相应的信息项。

在不脱离本发明的范围的情况下，可以对上述实施例进行各种改进。后附权利要求中限定了本发明的各个方面和特征。

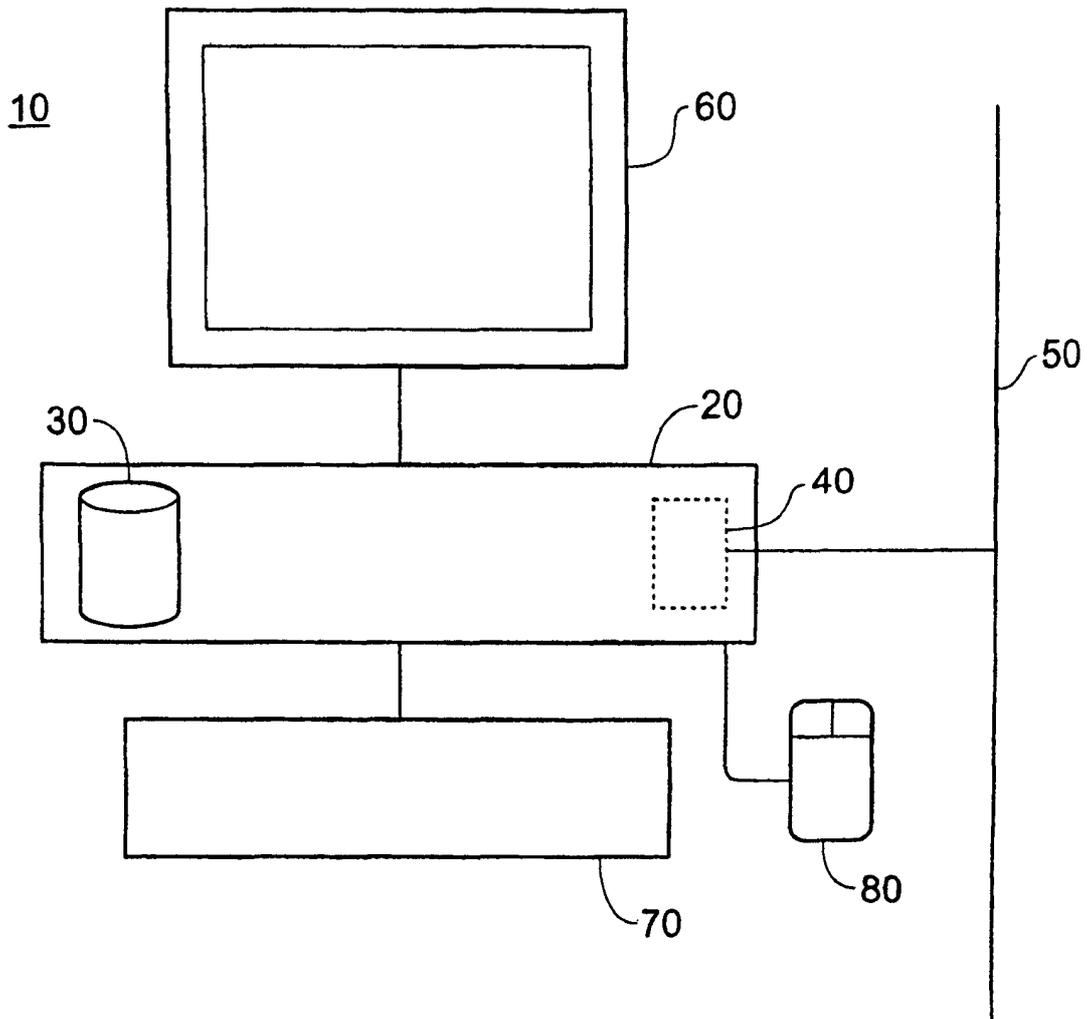


图 1

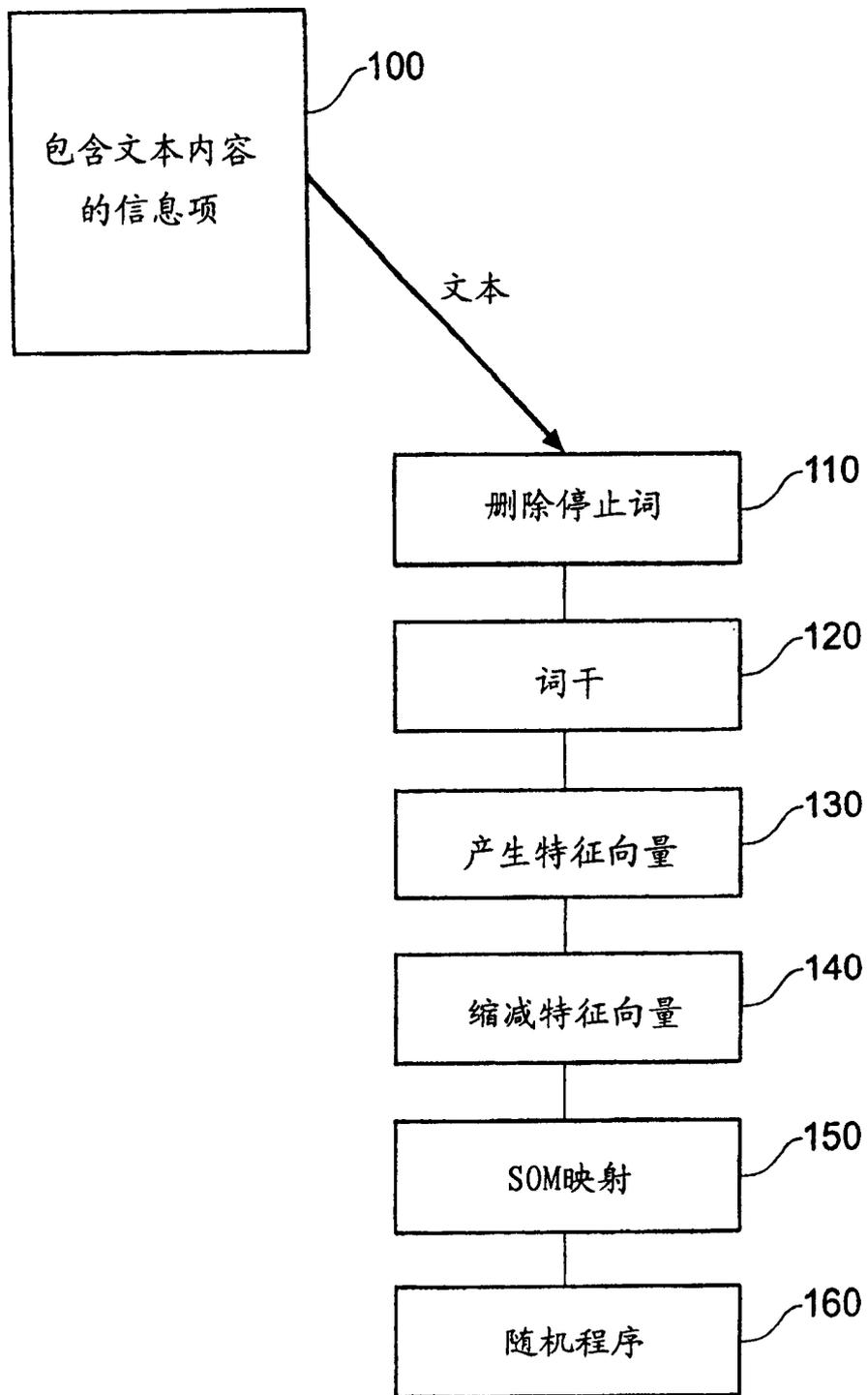
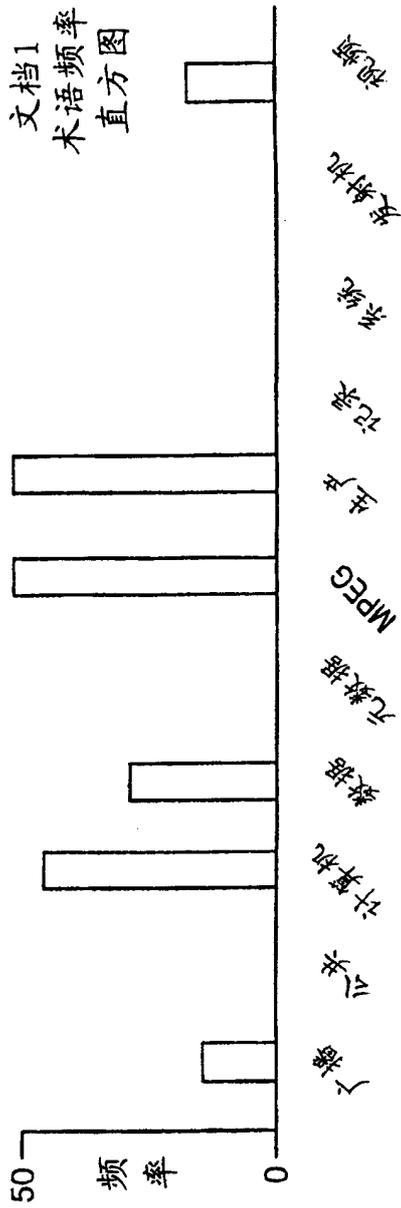
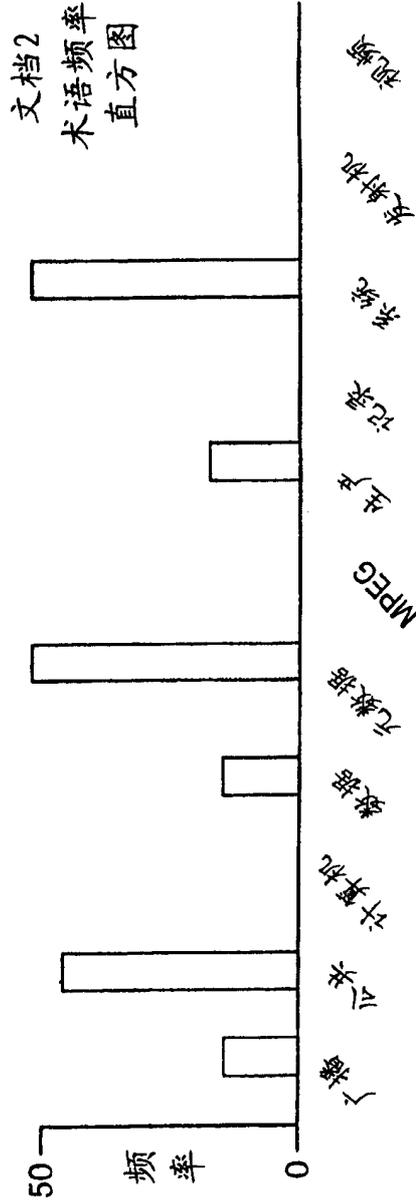


图 2



术语 图 3a



术语 图 3b

0
0
0
0
1
0
0
0
0
0
0
2
.
.
.
0
0
0
0
3

图 4a

0
0
3
0
1
0
7
0
0
0
2

图 4b

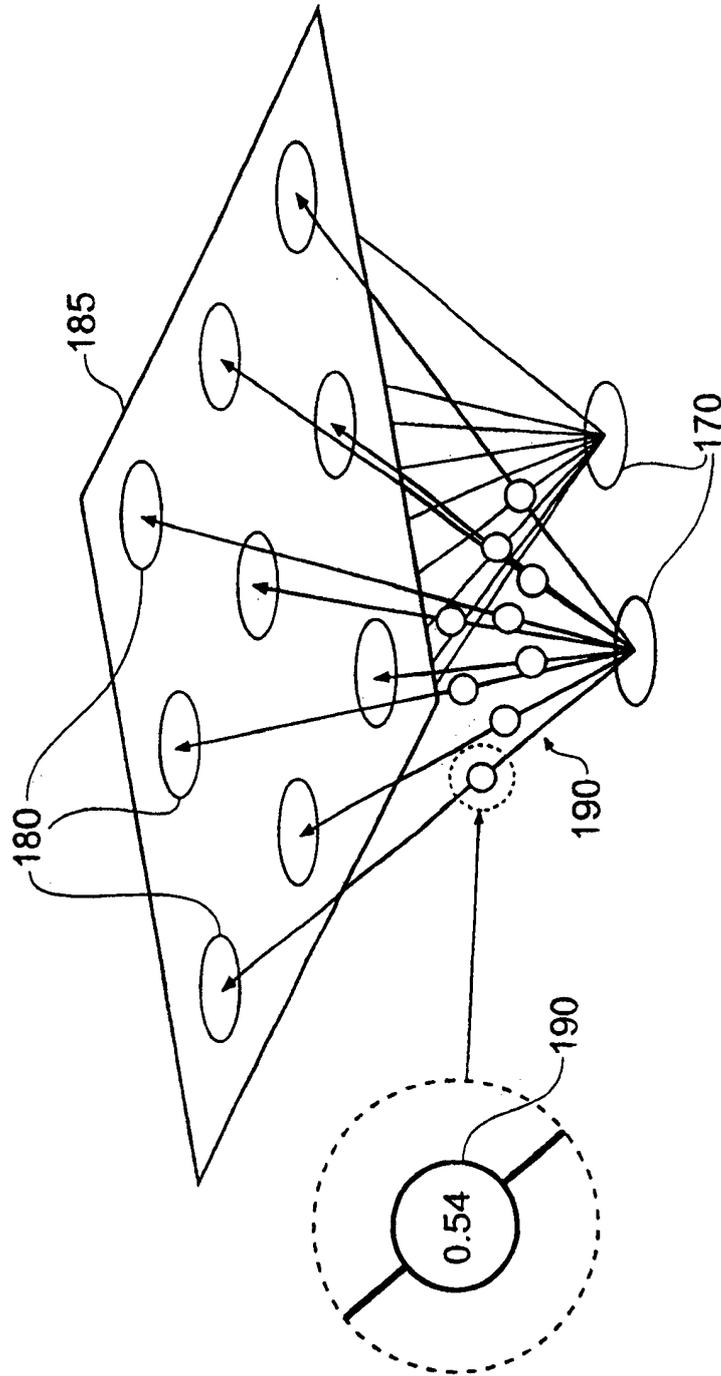


图 5

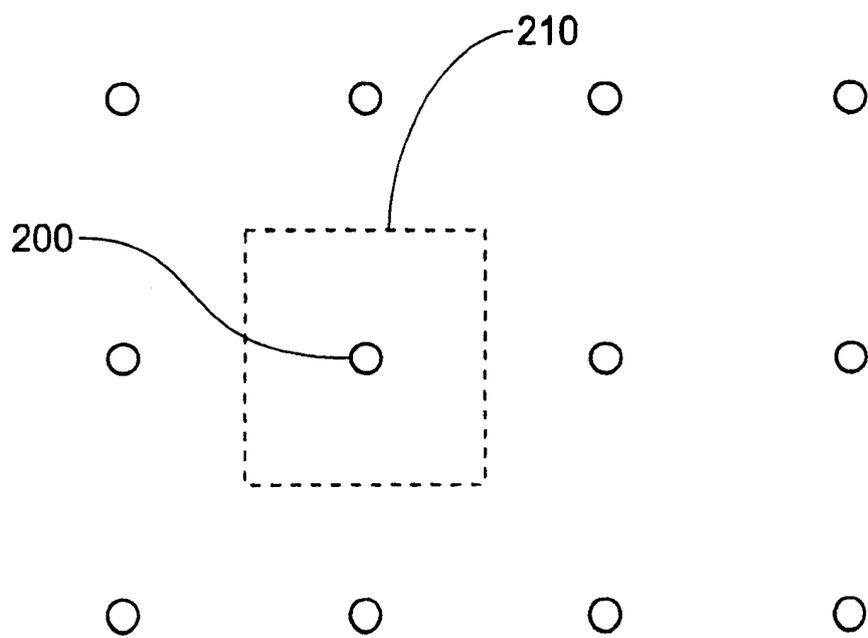


图 6

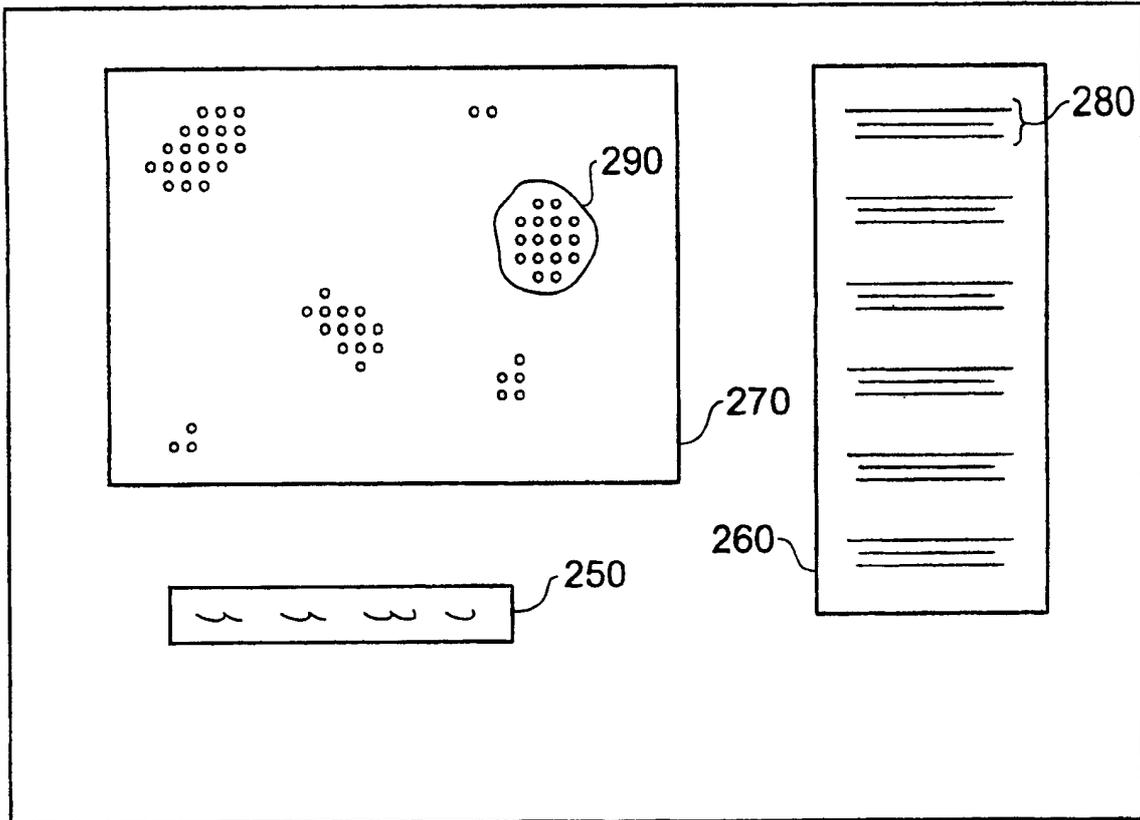


图 7

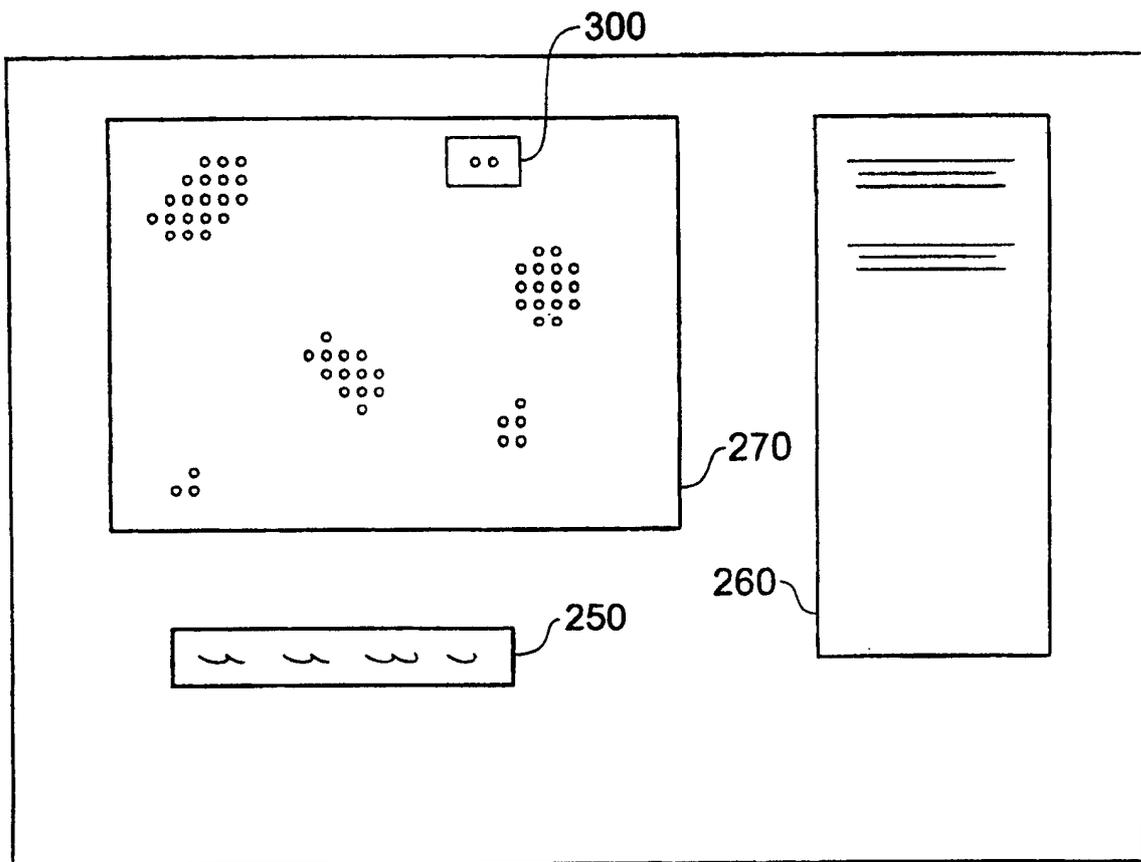


图 8

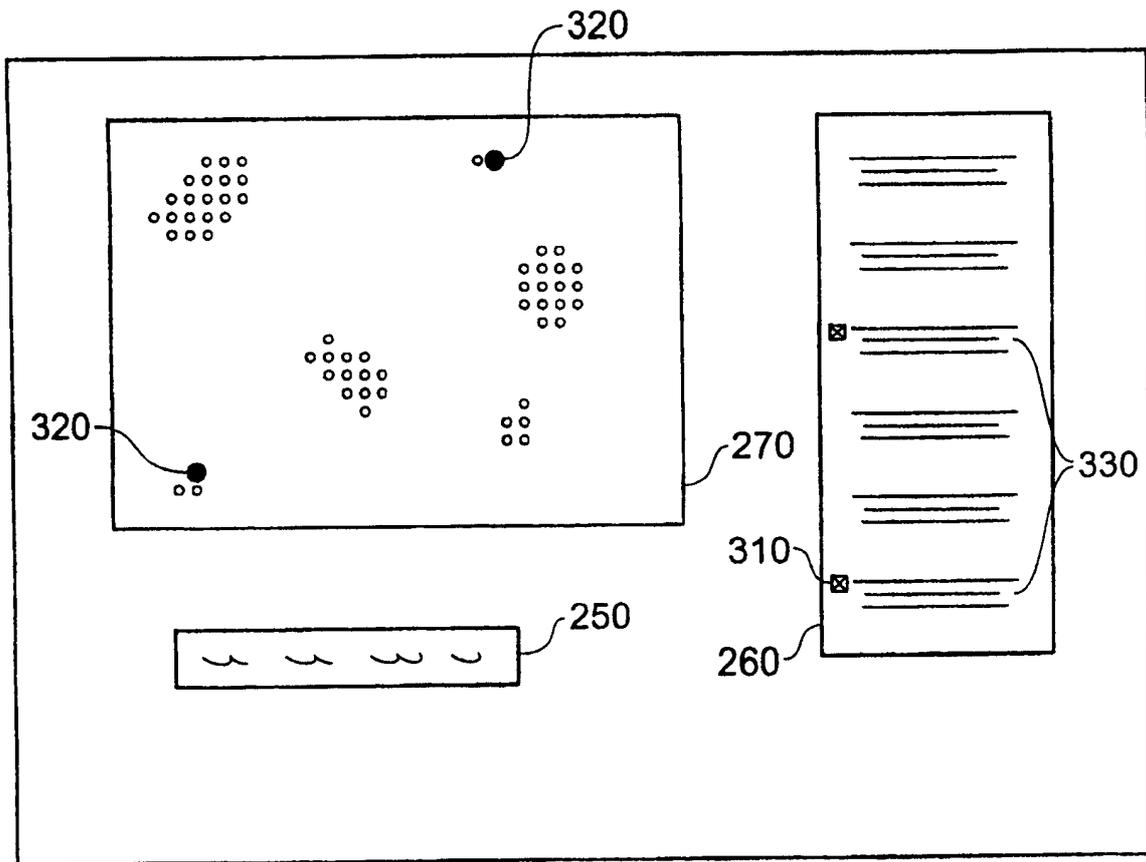


图 9

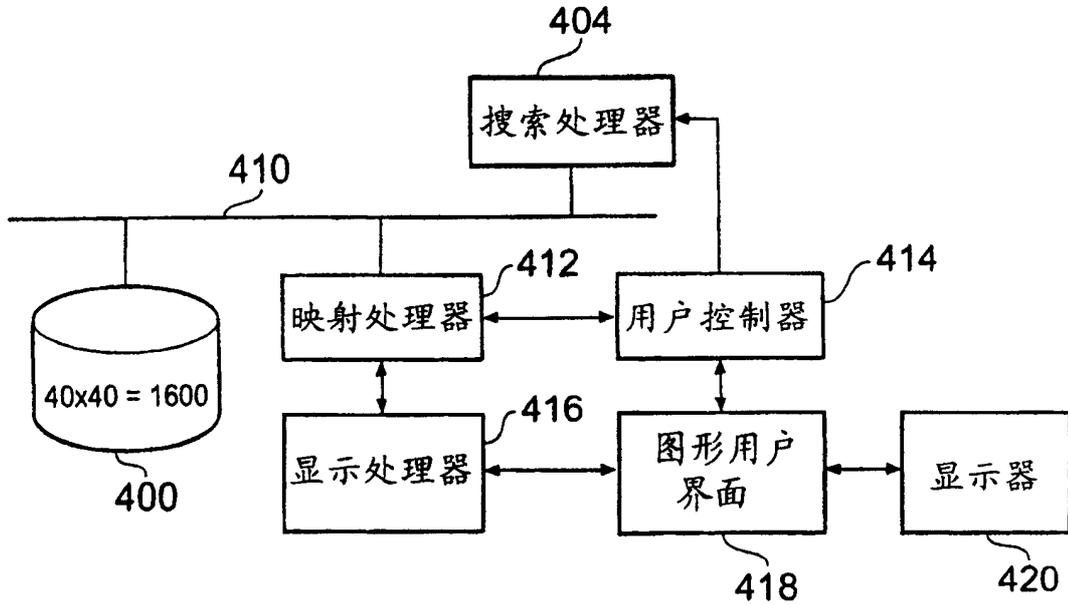


图 10

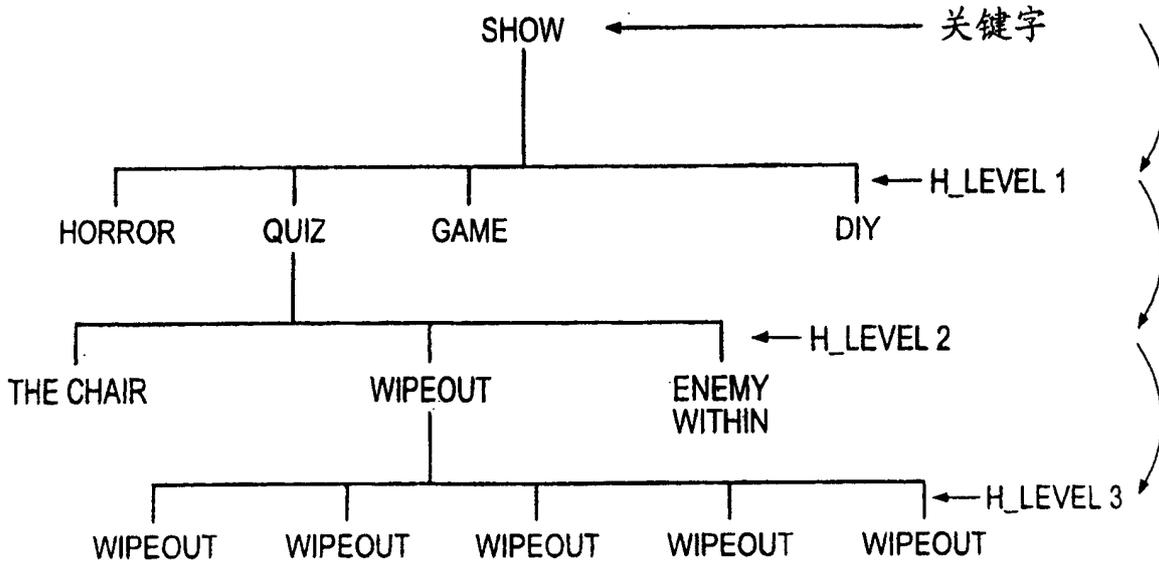


图 11

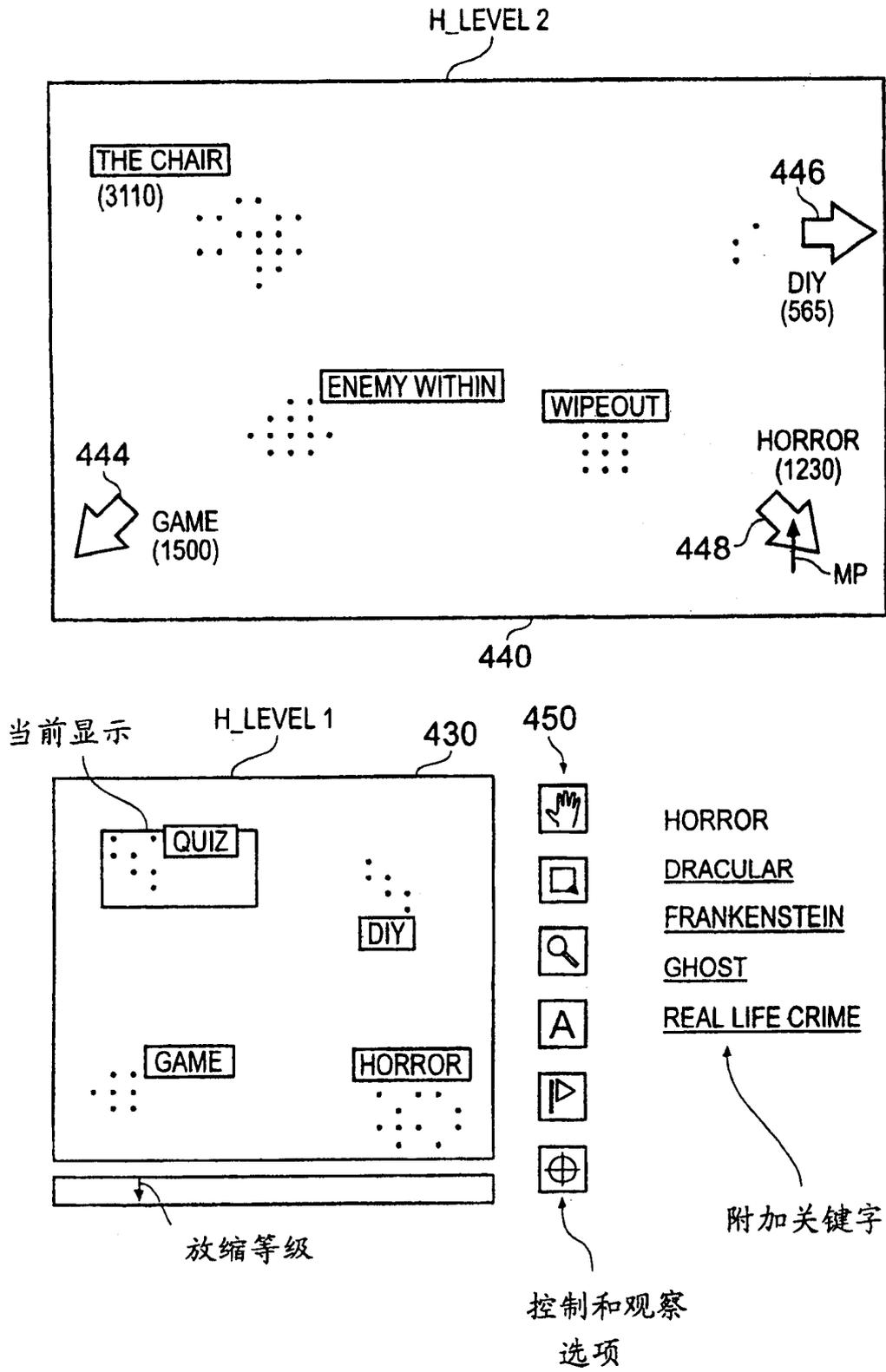


图 12

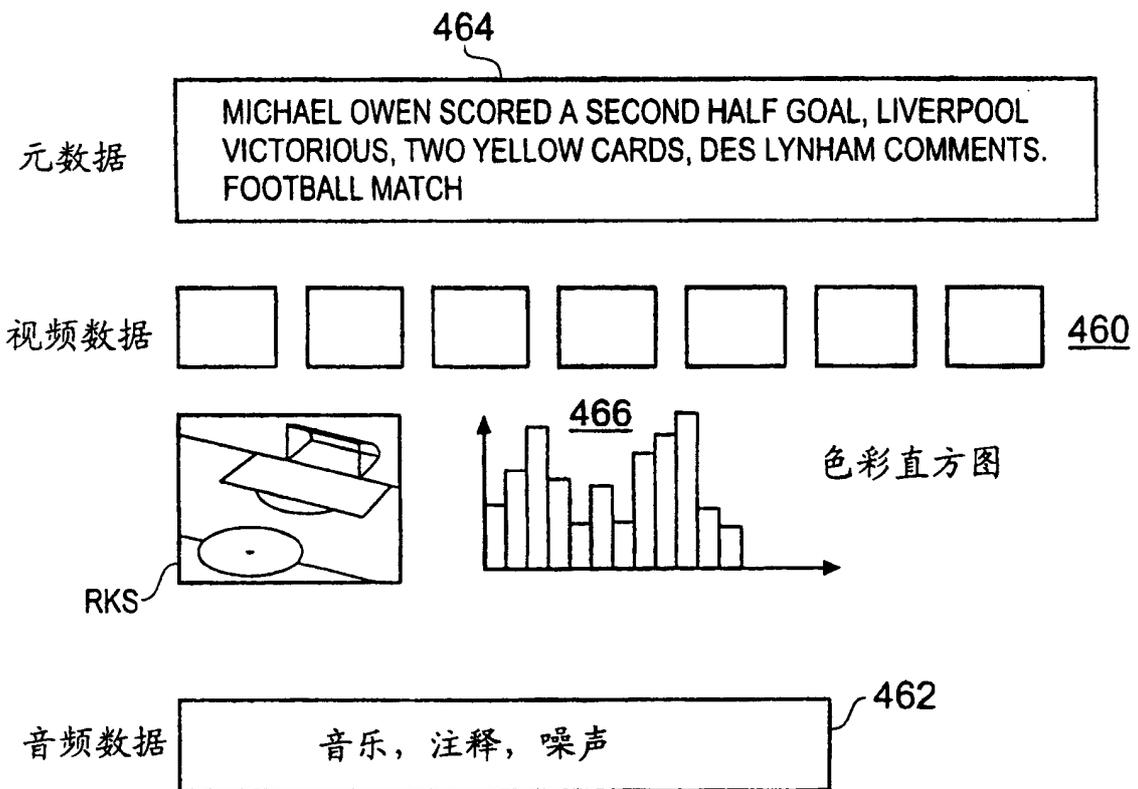


图 13

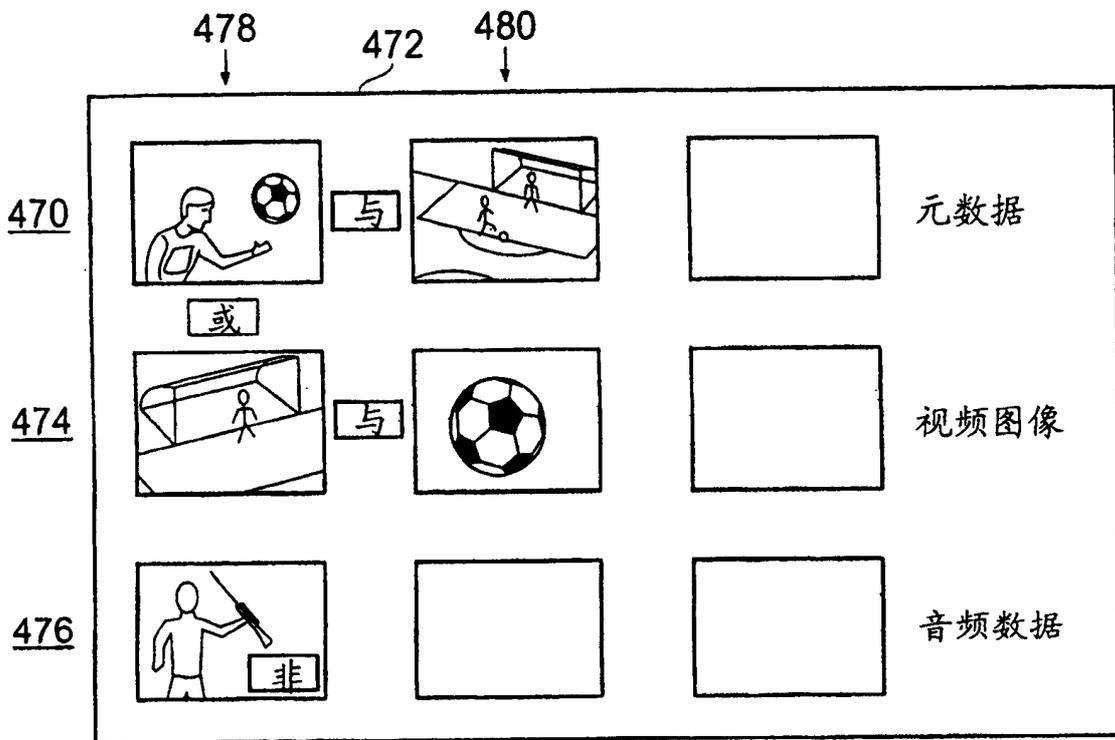


图 14

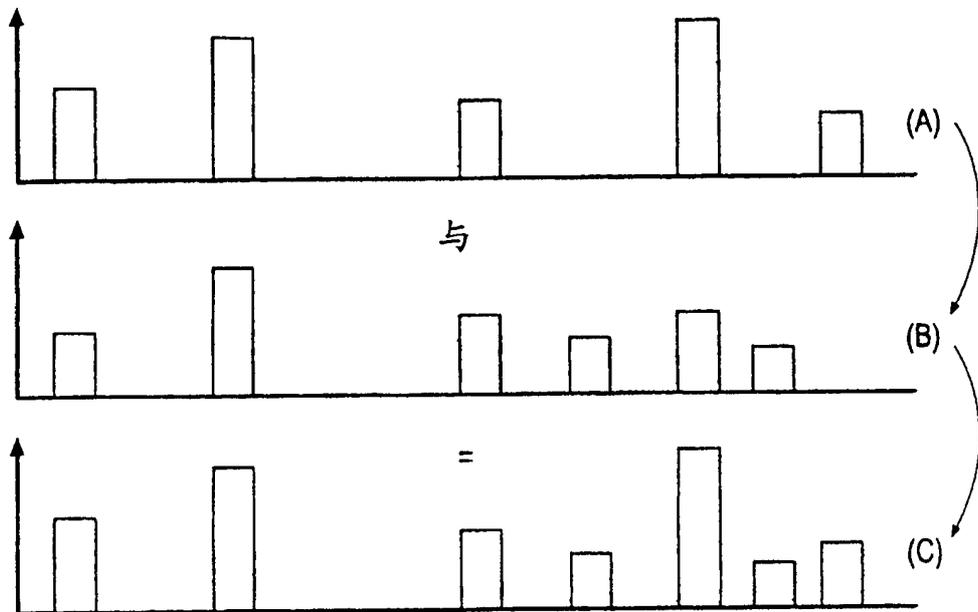


图 15

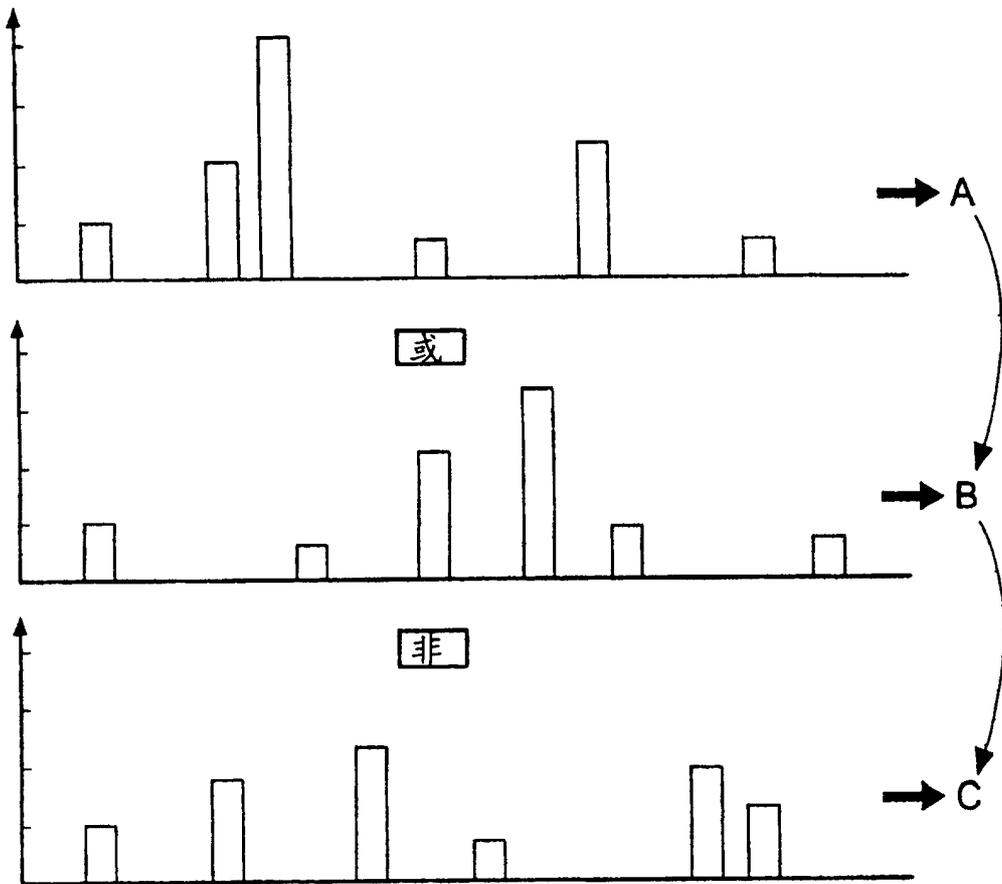


图 16

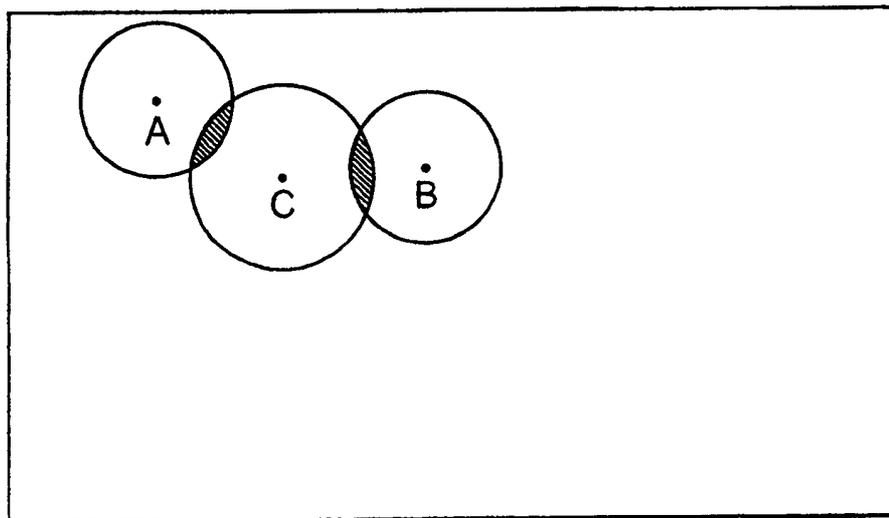


图 17

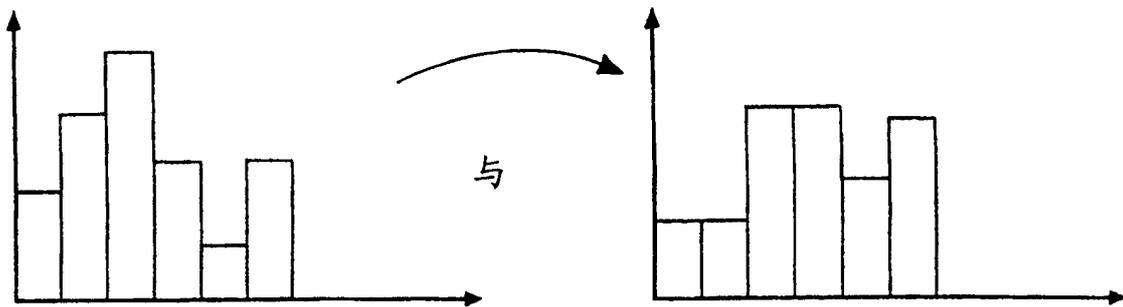


图 18 (a)

图 18 (b)

=

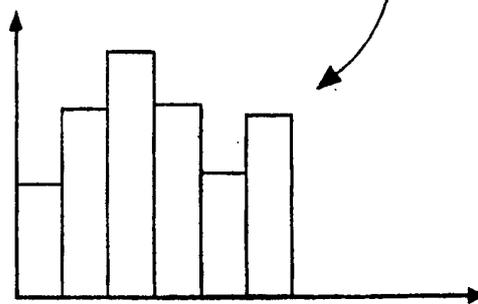


图 18 (c)

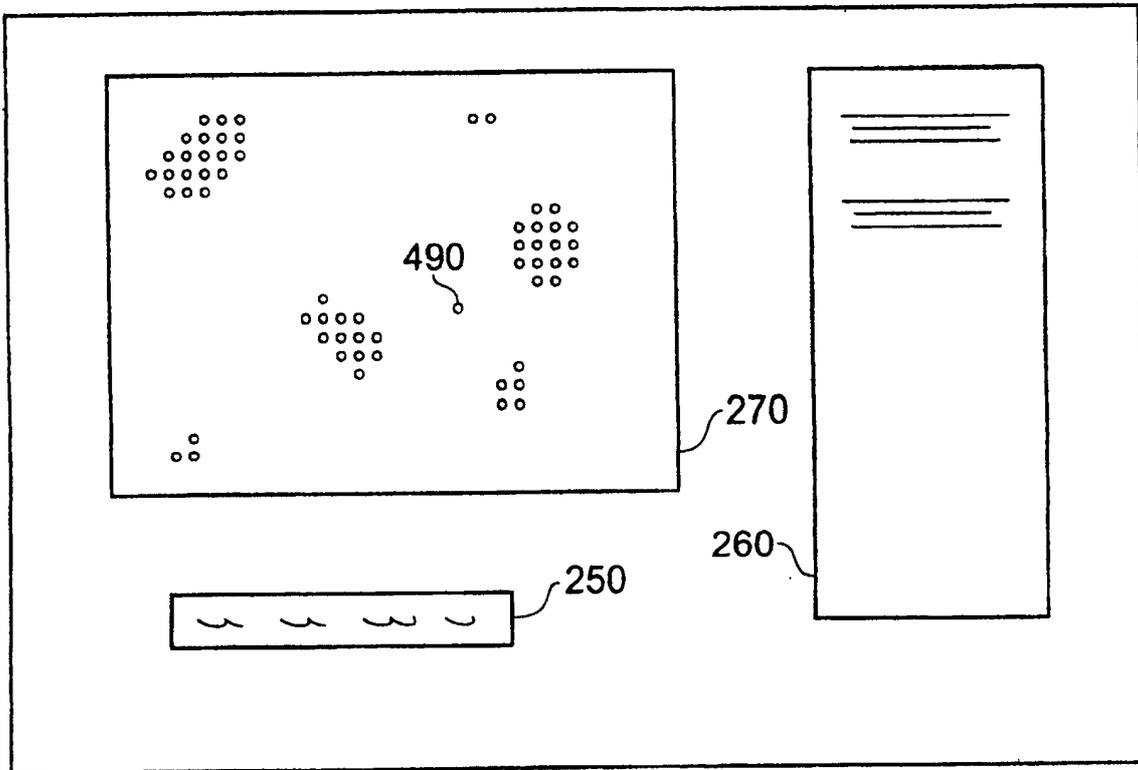


图 19

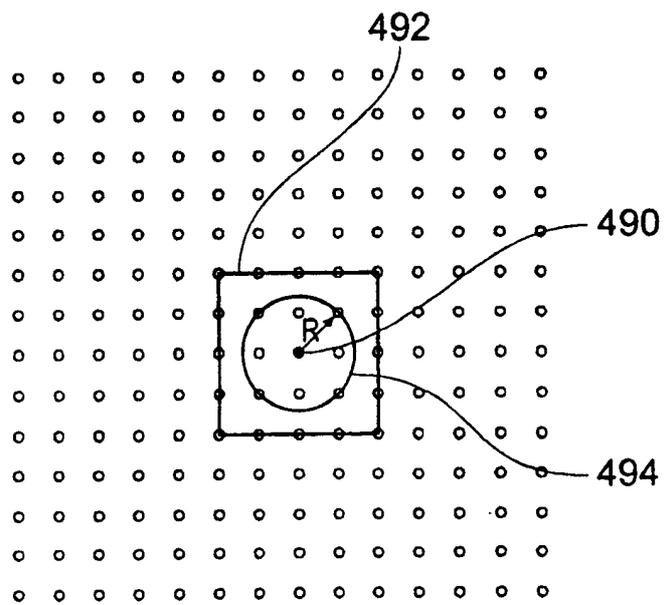


图 20

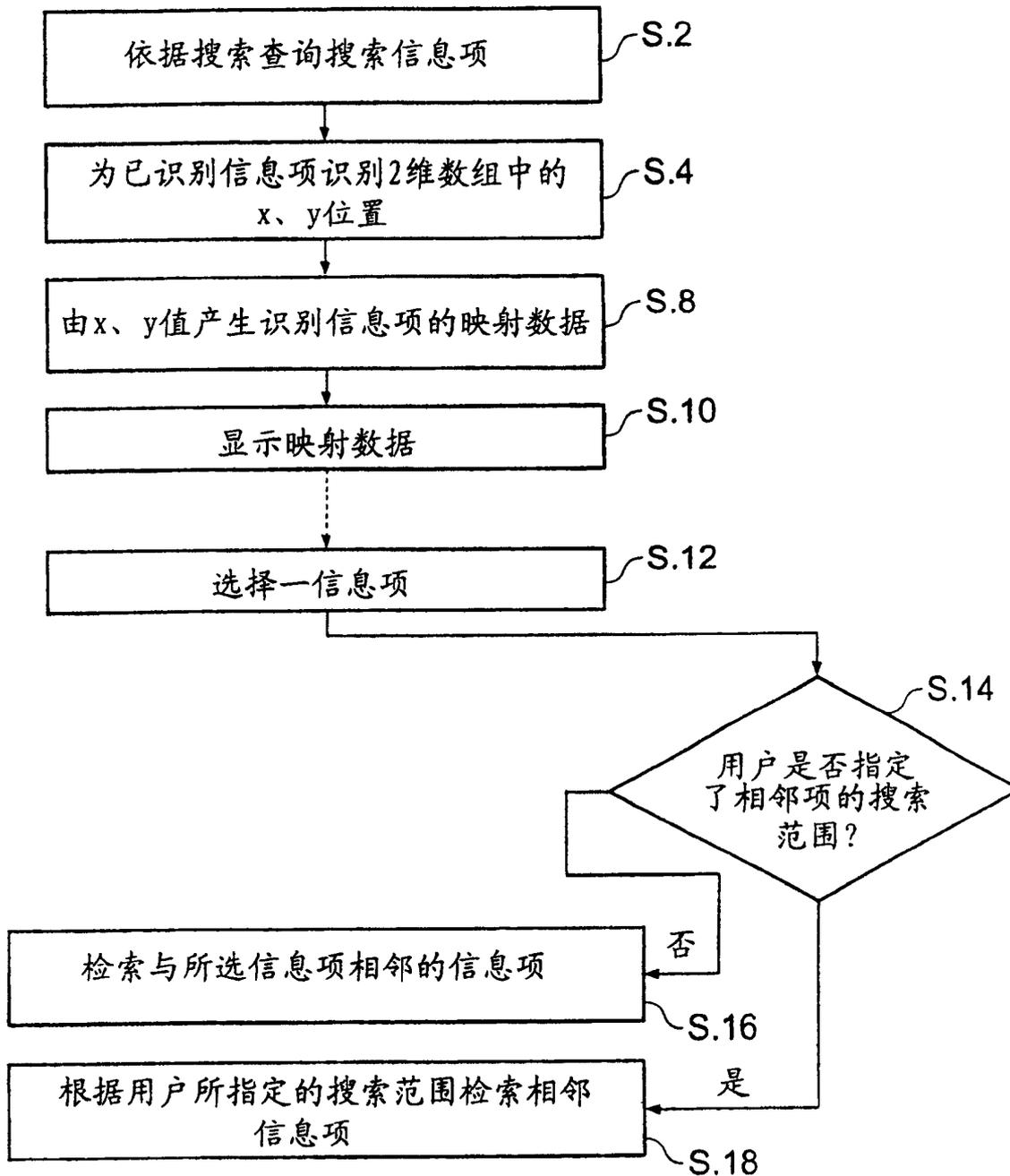


图 21