Method for reducing the storage requirement and structure of a database

The invention relates to a method for reducing the storage requirement of a database and also an apparatus for carrying out the method.

Skilful modelling of a database is a problem particularly when very large volumes of data are involved. A concrete example is the desire or the necessity to store telephone tariff information or other individually assignable information units for any location-location combinations worldwide or at least for a relatively large country, for example Germany. The storage space available for this should be kept to a minimum. In Germany alone, however, there is already a total of 5604*5604 possible dialling code combinations. A total, therefore, of more than 30 million data records. In this case, each data record comprises at least two long-integer values of 4 bytes just for the dialling code, up to 25 bytes for the place name and at least 4 bytes for the assignment to the individual tariff zones. Each data record thus has at least 4+4+25+4=37 bytes. Consequently, the whole database would have a size of about 1.1 Gbytes, even without taking account of the so-called overhead, (indexing, etc.) of the respectively selected database. Skilful modelling of the database and classic compression technology would allow the storage requirement to be reduced to approximately 100 Mbytes. For certain database systems, which will be specified in more detail below, there is a desire to reduce the storage requirement further.

DE 34 25 844 A1 discloses a method for rapidly retrieving specific data records, produced by a telecommunications switching system or data processing system, in a data memory. The method proposes that when the data records are stored in the data memory, search

terms are simultaneously stored as so-called key words in parallel therewith in a particular memory area.
In the course of further processing or when the data records are printed out, whenever a key word corresponding to the search term is found, the counter position of an address counter for serial searching of the particular memory area is used to address the associated data record sought in a data memory.

Another search method for a data memory is disclosed in DE 39 08 684 A1. In this method, a finite management area, a search area located within the management area, and also search objects are predetermined in an n-dimensional co-ordinates system, the search objects which intersect the search area being determined.

These methods also permit only a slight reduction in the storage requirement.

The invention is therefore based on the object of specifying a method and an apparatus suitable for carrying out the method which enable the storage requirement to be reduced beyond the extent known to date in the case of certain databases.

The inventive method for reducing the storage requirement of a database is characterized in that information units to be stored are assigned to the points of intersection of an at least two-dimensional raster matrix, individual colour values are assigned to the information units located in the raster matrix, and in that the image matrix thus formed is stored as database or as part of a database, preferably after the image matrix has been subjected to image compression.

The raster matrix is spanned by axes and information contents defined either by algorithms or by an assignment table can be assigned in [sic] individual axis values. Suitable colour values used within the raster image matrix are, in particular, the known and defined RGB colour values, thereby enabling classic loss-free compression methods to be employed for a

further considerable reduction in the storage requirement.

An inventive apparatus for carrying out the method according to the invention is characterized by the features of Patent Claim 5 with advantageous refinement possibilities in accordance with further dependent patent claims.

The invention will also be explained in more detail with reference to the appended drawing, using an exemplary embodiment.

The invention is based on the concept that any items of information can be interpreted as colour values within a multi-dimensional, for example a two-dimensional, colour space which defines an image. Complex clusters of information, e.g. databases or extensive tables, are represented as a stringing together of colour values, that is to say as images. The colour values are not strung together in an arbitrary order, but rather in the form of a multi-dimensional matrix or of a co-ordinate system, the axes being assigned their own information content. Quite apart from the reduction associated with this colour value conversion, classic loss-free image compression methods can now be applied to the images produced by the conversion.

Classic databases usually store information in tables, the columns each having their own information content but the rows being used (if at all) in an identifying manner.

On the basis of the invention, combinations of atomic or molecular information units are generated or interpreted in particular as complex colour values. For example, three ASCII characters are combined to form a colour value. In this case, the assignment of the colour values to specific character strings is defined either by an algorithm or by assignment tables which, in turn, can likewise be stored as image information.

With reference to the above example of a telephone tariff information database, the x axis of a two-dimensional colour matrix is assigned the destination locations and the y axis the originating locations, for example. A 5604*5604 matrix is thus constructed. The corresponding tariff zone is entered at the points of intersection. In addition, an index is required [lacuna] which a dialling code is assigned to a matrix row and/or column. The respective tariff information is entered as a colour value into the matrix. The result is then a bitmap with a total of 5604*5604 and a colour depth of in each case 4 bytes. When stored as a bitmap, the storage requirement is still approximately 13 Mbytes. However, since a graphic is now present, classic methods for loss-free compression of graphics can be employed without any difficulty. If the PNG format is taken as a basis, for example, then ultimately a storage requirement of only approximately 650 Kbytes is still necessary for the resulting database, in conjunction with the possibility of faster access to individual data records.

The drawing diagrammatically illustrates the units or the structure of a database with application of the method according to the invention.

Insofar as a multi-axis assignment is taken as a basis, an axis value assignment unit or axis translation unit ATU is provided. This ATU assigns a specific axis value to an item of atomic or molecular information or meta-information by means of an algorithm or an assignment table and also transfers the said axis value back again. Examples of meta-information that may be imagined are a pointer on [sic] a document or, in more complex cases, a "genuine" image.

Images are known to be (as a rule) two-dimensional arrangements ("distributions") of individual colour values. An image can thus also be understood as a co-ordinate system, two- or multi-

dimensional co-ordinate systems in which individual colour values are to be located being assumed in the present context. In the case of more than two-dimensional co-ordinate systems one may also speak of colour spaces.

If a two-dimensional co-ordinate system is assumed, as an example, then the assignment rule underlying the ATU can be represented, with reference to the above example, simply as follows:

"Take the integer value of a dialling code as axis value",

the quantity of all the originating locations then being determined as the information space.

The ATU itself accesses, in turn, a colour value converter unit or colour translation unit CTU, which determines the assignment of specific information items to the axis values, that is to say the assignment of individual colour values to the atomic or molecular information units or meta-information units, and vice versa.

This assignment can again be made by means of algorithms or else by means of assignment tables. In this case, the following fundamental concept is applied:

As is known, colour values can be described as RGB values. In this case, an RGB value represents the component of red, green and blue by means of a respective 8-byte value for example (decimal: 0 to 255). The three individual colour values are converted into a single RGB value, for example

Red = 255, Green = 255, Blue = 255 → RGB = 16777215 or

Red = 65, Green = 43, Blue = 0 → RGB = 11073.

The three individual values are then combined by bit-by-bit combination of the individual colour values to form a 24-byte colour value.

Example: 255 corresponds to 11111111. As a result, the RGB value (255, 255, 255) yields, when represented as a bit value: 11111111 11111111 11111111.

As the example illustrates, specific information items are thus "converted" into colour values, to be precise

- either by means of a defined assignment table for individual bit values,

- or by means of an assignment table for RGB values to items of atomic or molecular information or meta-information,

- or by means of an algorithm or an assignment table of items of atomic or molecular information or meta-information in each case to a specific red, green and blue value.

In other words, the CTU thus assigns a colour value, for example an RGB value, to an item of atomic or molecular information or meta-information. In this case, the application is not restricted to the creation of, for example, 24-byte RGB values, but rather can be applied analogously to any colour values, for example 32-byte RGB values.

This will be illustrated using a simple example for the CTU logic arrangement:

"Compose a colour value from three successive ASCII values in each case".

Using this simple rule, the word "Beispieltext" ["example text"], for instance, could be represented by means of a sequence of colour values composed of four pixels:

Beispieltext → (66,101,105)(115,112,105)(101,108,116)(101,120,116)
→ 6907202,6910067,7629925,7632997.

If the axis value assignment unit ATU is used in conjunction with the colour value converter unit CTU, then certain axes which cannot be used in classic database systems can likewise be utilized, in a simple manner, for storing additional information.

The data or information are then actually stored in images or colour spaces in a storage bank referred to as data cube in the drawing.

As mentioned, in the case of a two-dimensional colour space (image) it is possible to employ classic loss-free compression methods in order to compress and reduce the storage requirement further. In the case of more than two-dimensional colour spaces, the latter can either be simulated by two-dimensional images or it is possible to adapt appropriate loss-free compression algorithms to multi-dimensional colour spaces.

Finally, for communication with the outside world, an interface is present which is designated as database engine and whose main task is the (buffer-) storage and retrieval of information. This interface must additionally fulfil all the essential functions of classic database systems; it therefore differs only in the form of its actual realization.

If its additional function is reduced to what is essential, then the actual difference is that the pixel values required in the context of the invention have to be read or set via the database interface in multi-dimensional colour spaces. As a rule, the elementary processing process appears as follows:

(1) The CTU converts the information units arising into colour values;

(2) the ATU (more than one of which are present, if appropriate) locates the information units (colour values) once again in their respective information spaces, and

(3) the database engine sets or reads at the corresponding location the colour values determined by the CTU in the data memory (data cube).

The particular advantages of the database solution according to the invention are, primarily:

- The storage space requirement is reduced considerably, for example by a factor of 100, that is to say from 100 Mbytes to 650 Kbytes, for example.

- The database access is faster on account of the architecture and the smaller database sizes.

A first reduction in the storage space requirement is already achieved by the use of the colour value converter unit CTU; a further substantial reduction is achieved by the additional use of the axis value assignment unit ATU.

The method according to the invention or the described database structure with reduced storage requirement is primarily suitable wherever very large volumes of data are involved or for volume-dependent transmission costs, for example for the "video on demand", "news on demand" sectors and also for internet transmissions and the like. The additional database modelling effort required in comparison with classic database systems is worthwhile for these areas of application. Applications of interest that are currently in the foreground are worldwide routing database systems, in particular for the internet fax service, as well as the abovementioned application for systems for tariff information and calculation via ticket machines.



Patent Claims

1. Method for reducing the storage requirement of a database, characterized in that

- information units to be stored are assigned to the points of intersection of an at least two-dimensional raster matrix,

- individual colour values are assigned to the information units located in the raster matrix, and in that

- the image matrix thus formed is stored as database or part of a database.

2. Method according to Claim 1, characterized in that complex RGB colour values are used as colour values.

3. Method according to Claim 1 or 2, characterized in that the colour values of the image matrix are subjected to an image compression process prior to their storage as a database or part of a database.

4. Method according to one of Claims 1 to 3, characterized in that algorithms or information contents defined by an assignment table are assigned to the axis values of the raster matrix.

5. Apparatus for carrying out the method according to Claim 1 or 2, characterized by

- a colour value converter unit (CTU), which, by means of an algorithm or an assignment table, allocates atomic or molecular information units or meta-information units to defined points of intersection of a colour matrix as a respective colour value within the colour space predetermined by the colour matrix,

- a storage bank (data cube) for storing the colour values and by

- an interface (database engine) which sets or reads colour values determined via the colour value converter unit.

* C L M S *

6. Apparatus according to Claim 5, characterized by an axis value assignment unit (ATU) which, by means of an algorithm or an assignment table, assigns atomic or molecular information units or meta-information units to predetermined axis values of a co-ordinate system corresponding to the multi-dimensional colour value matrix.

7. Apparatus according to Claim 5 or 6, characterized by a device for loss-free compression of the colour value matrix, which device is assigned to the storage bank.