OPIC
Office de la propriété intellectuelle du Canada

CIPO
Canadian Intellectual Property Office

(12) (19) (CA) **Brevet-Patent**

(11) (21) (C) **2,130,033**

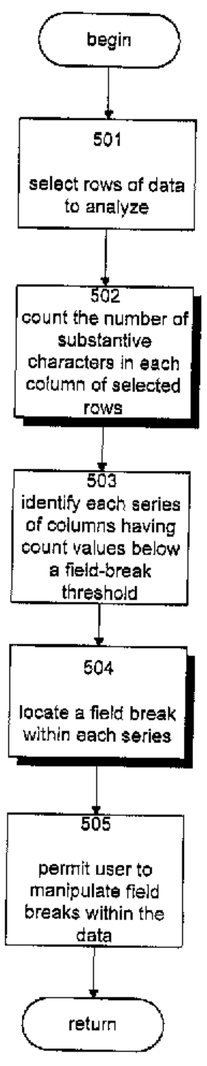
(22) 1994/08/12
(43) 1995/03/16
(45) 2001/04/10

(72) Comer, Ross W., US
(72) Solomon, Cynthia Ann, US
(73) MICROSOFT CORPORATION, US

(51) Int.Cl.[5] G06F 15/403
(30) 1993/09/15 (08/122,891) US

(54) **METHODE ET DISPOSITIF POUR LOCALISER LES INTERRUPTIONS DANS LES CHAMPS DE DONNEES**

(54) **METHOD AND SYSTEM FOR LOCATING FIELD BREAKS WITHIN INPUT DATA**



(57) A method and system for locating field breaks within input data is provided. In a preferred embodiment, a software facility locates field breaks within input data. The facility first receives input data made of a large number of rows of text. The facility then selects a group of the received rows of text to analyze. The facility identifies field breaks within the input data by analyzing the selected group of rows of text. The facility preferably analyzes the selected rows by counting the number of substantive characters in each column of the selected rows. The facility then uses the count to determine which columns have less than a field-break threshold of substantive characters. The facility locates one field break within each series of columns that have less than the field-break threshold of substantive characters.

METHOD AND SYSTEM FOR LOCATING FIELD BREAKS WITHIN INPUT DATA

Abstract of the Disclosure

A method and system for locating field breaks within input data is provided. In a preferred embodiment, a software facility locates field breaks within input data. The facility first receives input data made of a large number of rows of text. The facility then selects a group of the received rows of text to analyze. The facility identifies field breaks within the input data by analyzing the selected group of rows of text. The facility preferably analyzes the selected rows by counting the number of substantive characters in each column of the selected rows. The facility then uses the count to determine which columns have less than a field-break threshold of substantive characters. The facility locates one field break within each series of columns that have less than the field-break threshold of substantive characters.

Description

METHOD AND SYSTEM FOR LOCATING FIELD BREAKS WITHIN INPUT DATA

Technical Field

The invention relates generally to a method and system for receiving input data, and more specifically, to a method and system for locating field breaks within input data.

Background of the Invention

Computers are often used to store and process data. Computer users usually use an application program to store a particular kind of data, or to process data in a particular way. A particular type of application program, called a database application, permits users to store and process a body of related data in the form of a table. A table is composed of records, each containing data about a particular thing, such as a mechanical part needed to construct a go-cart. Each record corresponds to one "line" of data, and a line is terminated by an end-of-line character, such as a return or line feed character. Each record is divided into a set of fields, or particular pieces of data. The set of fields is the same for every record in the table. Example fields for part records include part number, part description, and part cost. A record in such table would contain the part number, part description, and part cost for a particular part. Fields usually also have a uniform length, measured in number of characters, in every record.

Figure 1 is a diagram showing the contents of an example table. The table 100 contains twenty records 101-120, each specifying data for each of seven fields 131-137. The fields are a stock number field 131, a description field 132, a quantity required field 133, a responsible mechanic field 134, an item cost field 135, a total cost field 136, and a date required field 137. In order to construct a go-cart, two axles, stock number 663661, which cost $215 each, are needed by Sue by December 16, 1993, at a total cost of $430.00. Accordingly, the stock number field for record 113 contains the stock number "663661", the description field contains the description "axles", the quantity required field contains the quantity "2", the responsible mechanic field contains the name "Sue", the item cost field contains the item cost "$215.00", the total cost field contains the total cost "$430.00", and the date required field

contains the date "12/16/93". These field contents would typically have been stored in the table after being inputted by a user. A user typically positions an insert cursor to a particular field and types the data the field is to contain, then instructs the database application to move the input cursor to another field using a next field key, such as the tab key.

Some database applications store table data in fixed-width field format, in which the contents of every field of each record are stored together in a string of text, and the fields are delineated by absolute position within the string. Text includes alphabetic, numeric and other characters. Figure 2 is a diagram of the information shown in the table Figure 1 stored as text strings. The figure shows character position numbers 201, also called column numbers, above the text strings. The first seven characters of each string contains the contents of the stock number field, the eighth through twenty-ninth character contains the contents of the description field, etc. Field width information is often stored separately from the table data, and is therefore sometimes unavailable to other application programs, such as other database applications or spreadsheet applications, when reading the table data.

When a spreadsheet or database application reads a text file containing a table in fixed-width field format for which field width information is not available, it must position field breaks within the table data. Typically, spreadsheet and database applications use one of two methods to position field breaks in such data. In a first manual method, the application program solicits the necessary information from the user, either by interactively inputting the width of each field, or by inputting a control string containing information about every field, including width. This method has the shortcoming that it requires the user to provide width information that is likely both outside of the user's knowledge and difficult to determine manually.

A second single row analysis method automatically analyzes the text in a single record, usually the first, or topmost record. This usually involves positioning a field break at every location at which the analyzed record contains a non-substantive character, such as a space, that follows a substantive character, such as a letter. This method has the shortcoming of often incorrectly positioning field breaks, since a single record is seldom representative of the entire table. Figure 3 is an example of how the single row analysis method of positioning field breaks would position field breaks in the strings of text shown in Figure 2. The diagram shows field breaks 301-304. It can be readily seen that, because of idiosyncrasies of the first record, field

breaks 302, 303, and 304 have all been positioned improperly: each is positioned within a field, rather than between two fields.

Summary of the Invention

It is an object of the invention to provide a method and system in a computer system for locating field breaks within input data.

It is another object of the invention to provide a method and system in a computer system for identifying fields within input data.

It is a further object of the invention to provide a method and system in a computer system for positioning field breaks within input data.

These and other objects, which will become apparent as the invention is more fully described below, are provided by a method and system for locating field breaks within input data. In a preferred embodiment, a software facility locates field breaks within input data. The facility first receives input data made of a large number of rows of text. The facility then selects a group of the received rows of text to analyze. The facility identifies field breaks within the input data by analyzing the selected group of rows of text. The facility preferably analyzes the selected rows by counting the number of substantive characters in each column of the selected rows. The facility then uses the count to determine which columns have less than a field-break threshold of substantive characters. The facility locates one field break within each series of columns that have less than the field-break threshold of substantive characters.

Brief Description of the Drawings

Figure 1 is a diagram showing the contents of an example table.

Figure 2 is a diagram of the information shown in the table Figure 1 stored as text strings.

Figure 3 is an example of how the single row analysis method of positioning field breaks would position field breaks in the strings of text shown in Figure 2.

Figure 4 is a high-level block diagram of the general-purpose computer system upon which the facility preferably operates.

Figure 5 is an overview flow diagram of the facility.

Figure 6 is a flow diagram depicting the step of counting the number of substantive characters in each column of data in greater detail.

Figure 7 is a flow diagram depicting the step of locating a field break within each series of columns having count values below a field-break threshold in greater detail.

Figure 8 is a count table showing the values of the Count[] array after the data shown in Figure 2 is counted.

Figure 9A is a display image of the preview window that shows the field breaks automatically located by the facility.

Figure 9B is a display image of the same preview window after the user has used the user interface portion of the facility to correct the position of a field break.

Figure 9C is a display image of the same preview window after the user has used the user interface portion of the facility to insert a missing field break.

Figure 10 is a diagram of a table stored as text strings that contains header information.

Detailed Description of the Invention

A method and system for locating field breaks within input data is provided. (Locating field breaks is also known as identifying or positioning field breaks.) In a preferred embodiment, a software facility (facility) acquires table data in fixed-width field format (input data), then locates field breaks within the input data by analyzing a group of rows of the input data. The facility selects a group of rows to analyze and counts the number of substantive, or non-space, characters in each character position, or column, of the rows in the group. The facility then uses the count to determine which columns have less than a field-break threshold of substantive characters. The facility locates one field break within each series of columns that have less than the field-break threshold of substantive characters. The facility preferably locates each established field break by first selecting an adjacent group of columns in the series that each has the smallest substantive character count in the series. This group of columns is called a contiguous minimum subset. The facility then tests each column that is adjacent to an end of the minimum subset to see whether text has been justified to that column. If just one of these columns is a justification point, then the facility locates the field break between it and the minimum subset. If neither or both of these columns is a justification point, then the facility locates the column break in the middle of

the minimum subset. The facility also includes a user interface for adding, deleting, and repositioning field breaks.

Figure 4 is a high-level block diagram of the general-purpose computer system upon which the facility preferably operates. The computer system 401 contains a central processing unit (CPU) 402, a computer memory (memory) 403, and input/output devices 404. Among the input/output devices is a storage device 403, such as a hard disk drive, a keyboard 406, a pointing device 407, such as a mouse and a display device, such as a monitor. The computer programs of the facility reside in the memory and execute on the CPU. The user preferably receives visual information from the facility through the monitor and interacts with the user interface portion of the facility using the keyboard or the mouse.

Figure 5 is an overview flow diagram of the facility. In step 501, the facility selects a group of the rows that make up the input data to analyze. While the facility selects as large a group of rows as is practicable, it is generally not considered necessary to select more than a few hundred rows. Significant improvements in prior art techniques can be seen even by selecting only a few rows.

In step 502, the facility counts the number of substantive characters in each column of the selected group of rows, storing the counts for all of the columns in a Count[] array. A substantive character is one that is likely part of the data, such as a letter or a numeral. On the other hand, non-substantive characters, also called separator characters, are those that usually are not part of the data, such as spaces, commas, and tab and carriage return characters.

In step 503, the facility identifies columns having count values below a field-break threshold contiguous columns meeting this requirement are identified as a series (which may include only one column) and a field break is later located in each series. The field-break threshold is a percentage of the total number of selected rows, whose value is chosen with reference to the amount of field separation expected. The larger the average number of non-substantive characters between fields expected, the lower the field-break threshold may be set, thereby reducing the chance of erroneously locating a field break. Conversely, the smaller the average number of non-substantive characters between fields, the higher the field-break threshold must be set, thereby increasing the chance of erroneously locating a field break. As an example, for ledger-type data like the data shown in Figure 2, a field-break

threshold of 12.5% is preferably chosen. In a preferred embodiment, the facility permits the user to adjust the field-break threshold. For example, a user may know from experience that the facility often inserts extraneous field breaks in input data from a particular source. When processing input data from that source in the future, the user would decrease the field break threshold, thereby reducing the chance of erroneously locating a field break. Each series is preferably stored as a beginning column number and an ending column number. The facility preferably discards any series that begins in the first column or ends in the last column.

In step 504, the facility locates a column break within each series identified in step 503 by selecting a contiguous minimum subset of the series, then testing for justification points, or columns to which text in each row of text of the input data is aligned, at each end of the series.

In step 505, the facility displays the data in a preview window with the field breaks shown by vertical lines, and permits the user to manipulate the field breaks. The user can reposition a field break by dragging it to a new position, and can delete a field break by double-clicking on it. The user can establish a new field break by clicking on a location in the data where a field break does not exist. A user can also use scroll bars to scroll the table data in both dimensions to observe the affect on his or her changes on all the fields and rows.

Figure 6 is a flow diagram depicting the counting step 502 in greater detail. In steps 601-609, the facility initializes a Count[] array indexed by column number, then sets each element equal to the number of selected rows having a substantive character in the corresponding column. In step 601, the facility initializes the Count[] array by setting each of its elements equal to zero. In step 602, the facility sets a RowIndex counter equal to the number of the first selected row. In step 603, the facility sets a ColIndex counter equal to the first column number, preferably 1. In step 604, if the current row and a column of the input data (Text[RowIndex, ColIndex]) contains a substantive character, then the facility increments the count for the current column (Count[ColIndex]) in step 605, else the facility continues at step 606.

In step 607, if the current column counter (ColIndex) is less than the last column number, then the facility increments the current column counter and continues at step 604 to check the next column in the same row, else the facility continues at step 608. In step 608, if the current row counter (RowIndex) is less than the last selected row number, then the facility

increments the current row counter and continues at step 603 to check the first column of the next row, else the steps shown in Figure 6 conclude.

Figure 7 is a flow diagram depicting the locating step 504 in greater detail. In steps 701-704, the facility locates a column break within each series identified in step 503 by selecting a contiguous minimum subset of the series, then testing for justification points at each end of the minimum subset. These steps are executed for each series identified in step 503. In step 701, the facility selects a contiguous minimum subset is preferably stored as a beginning column value and an ending column value.

In steps 702-704, the facility locates a column break either just beyond one of the ends of these contiguous minimum subsets selected in step 503, or as near to the center of the selected contiguous minimum subset as possible. The contiguous minimum subset corresponds roughly to empty columns between two fields. If a single one of the two columns is justified on the side of the blank columns, then it is most effective to locate the field break at that end of the contiguous minimum subset. On the other hand, if neither field is justified to the side of the blank columns, or they both are, then it is most effective to position the field break in the center of the contiguous minimum subset. In step 702, if exactly one of the two columns beyond an end of the contiguous minimum subset is a justification point, then the facility continues at step 703 to locate the field break between the single justification point and the contiguous minimum subset, else the facility continues at step 704 to locate the field break in the center of the contiguous minimum subset. The facility determines whether a column is a justification point by comparing its count to a justification threshold. The justification threshold, like the field-break threshold, is a percentage of the total number of selected rows, whose value is chosen with reference to the type of data expected. A justification threshold of 50% is preferably chosen for most data. In a preferred embodiment, the facility permits the user to adjust the justification threshold. For example, the user may know from experience that, when processing input from a particular source, the facility tends to erroneously detect justification points in the input data. When receiving input data in the future from the same source, the user would reduce the justification threshold to reduce the likelihood of erroneously recognizing justification points. The facility then repeats steps 701-704 for any remaining series.

The following example illustrates the operation of the facility. Figure 8 is a count table showing the values of the Count[] array after step 502

is executed for the data shown in Figure 2. The lines of count table 800 each contain a column number and a count value. The count value for each line indicates the number of substantive (here, non-space) characters appearing in the numbered column. For example, the count table shows that column 10 (of Figure 2) has a count value of 18, which is equal to the number of substantive characters in column 10. Because the sample data has 20 rows, with a 12.5% field-break threshold, the nominal field-break threshold is 2.5. The count values are compared to the field break threshold in step 503, and contiguous columns having count values below the field break threshold are grouped into series. The first series in the sample data are from column 30 to column 31, from column 33 to column 33, from column 46 to column 50, and from column 58 to column 58. As an example of selecting a contiguous minimum subset in the series that begins with column 46 and ends with column 50, in step 504, the facility selects the contiguous minimum subset of column 46 to column 49, which has a count value of 0.

The contiguous minimum subset for column 46 to column 49 is preceded by column 45, which has a count of 20, and followed by column 50, which has a count of 2. The justification threshold for the sample data is 10 (assuming a justification threshold of 50%), which is exceeded by the count of column 45, but not by the count of column 50. Because a facility recognizes a single justification point, it executes step 703 to locate the field break between the minimum subset (column 49) and the justification point (column 50).

Figure 9A is a display image of the preview window that shows the field breaks automatically located by the facility. The facility displays the preview window 900 as the visual interface for the preview mode supported by the facility in step 505. The preview window contains a portion of the table data 901, and the position of each field break is shown in the data as vertical bars 902-906. The user can use scroll bars 907 and 908 to scroll the table data in both dimensions to observe the effect of his or her changes on all of the fields and rows.

It can be seen in Figure 9A that the facility has mispositioned field break 903. Figure 9B is a display image of the same preview window after the user has used the user interface portion of the facility to correct the position of field break 903. The user uses the mouse to position the mouse cursor over field break 903 in its initial position, presses a button on the mouse, uses the mouse to move the mouse cursor, as well as the field break, to its destination position, and releases the button. The facility interprets these

actions by the user as a field break move command and moves the field break between columns 31 and 32 to between columns 29 and 30.

It can be seen in Figure 9B that the facility has also failed to locate a field break between the responsible mechanic field and the item cost field. Figure 9C is a display image of the same preview window after the user has used the user interface portion of the facility to insert a missing field break between these fields. The user uses the mouse to position the mouse cursor between columns 38 and 39 and presses a button on the mouse. The facility interprets these actions by the user as a field break add command and locates a field break between columns 38 and 39.

The facility has further features that enhance its effectiveness. In a preferred embodiment, the facility detects when it has located a field break inside a field organized in one of several common data formats that incorporates non-substantive characters, such as spaces. As an example, some data formats specify that parts of the data are separated by spaces, such as the representation "Aug 3 1993" for August 3, 1993. The facility preferably maintains a table of such formats, which include various time ("12:30 PM") and currency ("3,500,00 fr.) formats, as well as date formats, and compares them to consecutive pairs of fields in the selected rows of input data. When the facility locates a pair of fields that together comply with the format, it removes the field break separating them.

When the input data contains a large number of rows, in step 501, the facility preferably skips, or omits, rows containing header information, which is often formatted differently than actual data. Figure 10 is a diagram of a table stored as text strings that contains header information. In the data 1000 shown in Figure 10, first row 1001 contains header information and should be omitted. In a preferred embodiment, the facility skips any rows at the beginning of the input data that contain obviously different formatting than the main records of the input data. In an alternate embodiment, the facility merely skips the first several rows of data, using the heuristic that most header information usually occurs in the first few rows of data.

It can be difficult to identify a field break between a first field that is left justified and a second field that is right justified, because, while the fields are often separated by non-substantive characters in individual rows, the counts for the columns on either side of the intended field break often average out to nearly the same value. The field break between the responsible mechanic field 134 and the item cost field 135 in the sample data is an

example of this situation. It can be seen from the count table that the last column of the responsible mechanic field (38) contains seven substantive characters, while the first column of the item cost field (39) contains ten substantive characters. Since neither one of these values is below the nominal field-break threshold of 2.5 characters, the facility would not normally identify a series between these fields, and would therefore normally not locate a field break between them. The facility is preferably able to generate a multiple-column count to address this difficulty. When the facility recognizes it, the facility increments the multiple-column count for a given column once for each row in which both the current column and the next column both contain a substantive character. The multiple-column count is then analyzed in the same way as the single-column count described above. Use of a multiple column count is effective to recognize a field break between the responsible mechanic field and the item cost field.

It can also be difficult to identify a field break between two fields that are separated by few or no non-substantive characters. However, in a preferred embodiment, the facility is able to identify a field break between a first field containing mostly letters and a second field containing mostly numerals. In this embodiment, the facility produces two separate counts: one in which only letters are considered substantive characters, and one in which only numerals are considered substantive characters. The facility then compares the two counts to identify columns at which one count drops off quickly and the other increases abruptly, and inserts additional field breaks at this column.

While this invention has been shown and described with reference to preferred embodiments, it will be understood by those skilled in the art that various changes or modifications in form and detail may be made without departing from the scope of the invention. For example, an aesthetically different user interface portion could be used to receive field break manipulation commands from the user. Further, the count values could be analyzed differently. For example, more than one count value at a time could be compared to relative thresholds. Also, more sophisticated forms of statistical analysis could be applied to the count values.

Claims

We claim:

1. A method in a computer system for locating field breaks within input data, the method comprising the steps of:

receiving input data comprising a multiplicity of rows of data;

selecting a plurality of the multiplicity of rows of data; and

identifying field breaks within the input data by analyzing the selected rows of data.

2. The method of claim 1 wherein the multiplicity of rows of data have ordered columns, each containing a character that may be either substantive or non-substantive, and wherein the selecting step selects every row of data of the multiplicity of rows of text, and wherein the identifying step includes the steps of:

counting the number of substantive characters in each columns of the selected rows of data, wherein a count value is produced for each column, and wherein the count values are ordered to correspond to the order of the columns; and

identifying field breaks within the input data based upon the count values for one or more of the columns.

3. The method of claim 1 wherein the selected rows of data have ordered columns, each containing a character that may be either substantive or non-substantive, and wherein the identifying step includes the steps of:

counting the number of substantive characters in each column of the selected rows of data, wherein a count value is produced for each column ; and

identifying field breaks within the input data based upon the count values for one or more of the columns.

4. The method of claim 3 wherein the plurality of rows of data is ordered, such that one of the rows of data is the top row of data, wherein the selecting step includes the step of omitting a contiguous group of one or more rows of data that includes the top row of data.

5. The method of claim 3 wherein the count values are ordered to correspond to the order of the columns, and wherein the rows of data include header

rows that contain header information for describing the contents of other rows of data, and wherein the selecting step includes the step of omitting the header rows.

6. The method of claim 3 wherein the step of identifying field breaks within the input data based upon the count values includes the steps of:

identifying one or more series of columns for which the count values are below a field-break threshold; and

locating a field break within each series of columns identified.

7. The method of claim 6 wherein the step of identifying one or more series of columns for which the count values are below a field-break threshold includes the step of omitting a series of columns for which the count values that are below the field-break threshold that appears in the order of the columns before the first column for which the count value is greater than the field-break threshold.

8. The method of claim 6 wherein the input data may contain one or more justification points, each of which is a column within the input data to which data in each row of data of the input data is aligned, and wherein the locating step includes the steps of, for each series of columns:

selecting a contiguous minimum subset of the series of columns, such that the contiguous minimum subset of the series of columns have counts that are generally not larger than any of the series of columns not within the contiguous minimum subset of the series of series of columns;

determining whether zero, one, or two of the two count values that are immediately adjacent to the contiguous minimum subset of the series of columns are justification points within the input data;

if one of the columns that is immediately contiguous to the contiguous minimum subset of the series of columns is a justification point, locating the field break between the column that is a justification point and the contiguous minimum subset of the series of columns; and

if zero or two of the count values that are immediately contiguous to the contiguous minimum subset of the series of columns are justification points, locating the field break within the contiguous minimum subset of the series off columns.

9. The method of claim 8 wherein the step of, if zero or two of the count values that are immediately contiguous to the contiguous minimum subset of

the series of columns are justification points, locating the field break within the contiguous minimum subset of the series of columns locates the field break as near to the center of the contiguous minimum subset of the series of columns as is possible.

10. The method of claim 8 wherein the step of determining whether zero, one, or two of the two count values that are immediately adjacent to the contiguous minimum subset of the series of columns are justification points includes the steps of:

determining whether the count value for the column immediately preceding the first column of the contiguous minimum subset of the series of columns exceeds a justification threshold; and

determining whether the count value for the column immediately following the last column of the contiguous minimum subset of the series of columns exceeds the justification threshold.

11. The method of claim 6 wherein the input data contains data organized in one of a predefined multiplicity of data formats, a plurality of the data formats specifying organizing the data using non-substantive characters, further including the steps of, after the step of locating a field break:

recognizing that a field break has been located within a field containing data organized in a data format specifying organizing the data using non-substantive characters; and

removing the field break recognized in the recognizing step.

12. The method of claim 1 wherein the selected rows of data have ordered columns, each containing a character that may either be non-substantive or substantive, and wherein the step of identifying field breaks within the input data by analyzing the selected rows of data includes the steps of:

for each column, counting the number of selected rows in which there is a substantive character in both the column and in an adjacent column; and

identifying field breaks within the input data based upon the count values produced by the counting step.

13. The method of claim 1 wherein the plurality of rows of data is ordered, such that one of the rows of data is the top row of data, wherein the

selecting step includes the step of omitting a contiguous group of one or more rows of data that includes the top row of data.

14. The method of claim 1 wherein the rows of data include header rows that contain header information for describing the contents of other rows of data, and wherein the selecting step includes the step of omitting the header rows.

15. A method in a computer system for identifying fields within input data, the method comprising the steps of:

receiving input data comprising a multiplicity of rows of text, the rows of text having ordered columns, each containing a character that may be either substantive or non-substantive;

selecting a plurality of the multiplicity of rows of text;

counting the number of substantive characters contained in a first predefined group of characters in each column of the selected rows of text, such that a first character group count value is produced for each column, and such that the first character group values are ordered to correspond to the order of the columns;

counting the number of substantive characters contained in a second predefined group of characters in each column of the selected rows of text, such that a second character group count value is produced for each column, and such that the second character group count values are ordered to correspond to the order of the columns; and

identifying fields within the input data based upon the first character group count values and second character group count values produced by the counting steps.

16. The method of claim 15 wherein the first predefined group of characters and the second predefined group of characters are mutually exclusive, and wherein the step of counting the number of substantive characters contained in a first predefined group of characters in each column of the selected rows of text counts only characters not counted by the step of counting the number of substantive characters contained in a second predefined group of characters in each column of the selected rows of text.

17. The method of claim 16 wherein some characters are letters, and wherein the first predefined group of characters contains letters and the second predefined group of characters contains non-letters, and wherein the step of

counting the number of substantive characters contained in a second predefined group of characters in each column of the selected rows of text counts only non-letters.

18. The method of claim 16 wherein some characters are numerals, and wherein the first predefined group of characters contains numerals and the second predefined group of characters contains non-numerals, and wherein the step of counting the number of substantive characters contained in a second predefined group of characters in each column of the selected rows of text counts only non-numerals.

19. A method in a computer system for positioning field breaks within input data, the method comprising the steps of:

receiving input data comprising a multiplicity of rows of text;

selecting a plurality of the multiplicity of rows of text;

identifying field breaks within the input data by analyzing the selected rows of text; and

manipulating the field breaks positioned in the identifying step in response to receiving input from a user.

20. The method of claim 19 wherein the manipulating step includes the steps of:

receiving a relocate instruction from the user; and

relocating a field break identified in the identifying step in response to receiving the relocate instruction from the user.

21. The method of claim 20 wherein the step of receiving a relocate instruction from the user includes the step of receiving an indication of which field break to relocate and an indication of the position to which to relocate the field break, and wherein the relocating step relocates the field break in accordance with the indications.

22. The method of claim 19 wherein the manipulating step includes the steps of:

receiving a delete instruction from the user; and

deleting a field break identified in the identifying step in response to receiving the delete instruction from the user.

23. The method of claim 22 wherein the step of receiving a delete instruction from the user includes the step of receiving an indication of which field break to delete, and wherein the deleting step deletes the field break in accordance with the indication.

24. The method of claim 19 wherein the manipulating step includes the steps of:

receiving an add instruction from the user; and

adding a field break to those identified in the identifying step in response to receiving the add instruction from the user.

25. The method of claim 24 wherein the step of receiving an add instruction from the user includes the step of receiving an indication of the position at which a field break is to be added, and wherein the adding step adds the field break at the position indicated by the indication.

26. A method in a computer system for acquiring and interpreting rows of data, the method comprising the steps of:

inputting a multiplicity of rows of data in which a field break separating a first field of data in the rows from a second field of data in the rows is implicit;

identifying the field break in the multiplicity of rows of data by analyzing a plurality of the rows of data; and

displaying a row of data so that the first field of data of the displayed row is separated from the second field of data of the displayed row.

27. The method of claim 26, further including the steps of, after the displaying step:

receiving from a user a move field break command specifying a new position for the field break;

in response to the move field break command, changing the position of the field break to the position specified in the move field break command; and

redisplaying the displayed row of data so that the first field of data is visually separated from the second field of data.

28. A computer-based field-break locator for identifying fields within input data, the input data having rows of fields, each field having characters, comprising:

a receiver for receiving the input data;

a selector for selecting a plurality of rows of the received input data; and

an analyzer for identifying the fields within the input data based on the selected rows.

29. The field break locator of claim 28 wherein the rows of the input data being logically divided into one-character columns, each character being either substantive or non-substantive, and wherein the analyzer comprises:

a counter for determining the number of non-substantive characters in each column of the selected rows of the input data;

a counter memory for storing the number of non-substantive characters in each column of the selected rows of the input data; and

a field break positioner for positioning a field break at a column proximate to a column having a relative minimum of non-substantive characters with respect to surrounding columns.

30. The field break locator of claim 29 wherein the field break positioner includes a data justification detector for detecting whether data in columns adjacent to the column having the smallest number of non-substantive characters is justified to the columns adjacent to the column having the smallest number of non-substantive characters, and wherein the field break positioner positions the field break based upon the results realized by the detector.

31. The field break locator of claim 28 wherein one or more rows of data contain header information for identifying portions of the remainder of the rows of data, and wherein the selector includes a discriminator for detecting and omitting one or more rows of data containing header information.

Oyen Wiggs
Green & Mutala
PATENT AGENTS

100

101 —	004234	control cable (inches)	115	Cindy	$.50	$5.75	12/6/93
102 —	121212	antilock brake units	4	Ann	$375.00	$1500.00	1/7/94
103 —	125402	spoiler	1	Roger	$50.00	$500.00	2/1/94
104 —	200009	65 hp. engine	1	Al	$450.00	$450.00	12/27/93
105 —	221122	heavy duty bumpers	2	Ty	$120.00	$240.00	12/2/94
106 —	234102	titanium-cone speakers	4	Ann	$200.00	$800.00	1/7/94
107 —	284782	body	1	Ann	$225.00	$225.00	11/28/94
108 —	345353	traction control senso	4	Al	$720.00	$2880.00	1/16/94
109 —	412344	tractor transmission	1	Ty	$275.00	$275.00	1/30/94
110 —	534523	steering carriage	1	David	$75.00	$75.00	11/30/93
111 —	560000	seat belt	2	Cindy	$25.00	$50.00	12/16/93
112 —	663661	axles	2	Sue	$215.00	$430.00	12/16/93
113 —	747474	steering wheel	1	Roger	$30.00	$30.00	2/3/94
114 —	746744	CD changer	1	Al	$550.00	$550.00	2/15/94
115 —	787686	mag wheels	4	Gina	$40.00	$160.00	1/12/94
116 —	846262	bucket seat	1	Gina	$75.00	$75.00	2/1/94
117 —	876543	pedals	2	Roger	$15.00	$30.00	12/26/93
118 —	934456	valve stems	4	Cindy	$.10	$.40	2/7/94
119 —	982528	frame	1	Ann	$200.00	$200.00	11/28/94
120 —	999999	racing slicks	4	Al	$175.00	$700.00	12/9/93
	131	132	133	134	135	136	137

Prior Art

FIG. 1

Oyen Wiggs Green & Mutala
PATENT AGENTS

2130033

```
00000000011111111112222222222333333333344444444445555555555666666666
12345678901234567890123456789012345678901234567890123456789012345678 } 201

004234 control cable (inches)115 Cindy   $.05       $5.75   12/6/93
121212 antilock brake units    4 Ann  $375.00    $1500.00   1/7/94
125402 spoiler                 1 Roger $50.00     $500.00   2/1/94
200009 65 hp. engine           1 Al   $450.00     $450.00  12/27/93
221122 heavy duty bumpers      2 Ty   $120.00     $240.00   12/2/94
234102 titanium-cone speakers  4 Ann  $200.00     $800.00   1/7/94
284782 body                    1 Ann  $225.00     $225.00  11/28/94
345353 traction control senso  4 Al   $720.00    $2880.00   1/16/94
412344 tractor transmission    1 Ty   $275.00     $275.00   1/30/94
534523 steering carriage       1 David $75.00      $75.00  11/30/93
560000 seat belt               2 Cindy $25.00      $50.00  12/16/93
663661 axles                   2 Sue  $215.00     $430.00  12/16/93
747474 steering wheel          1 Roger $30.00      $30.00   2/3/94
746744 CD changer              1 Al   $550.00     $550.00   2/15/94
787686 mag wheels              4 Gina  $40.00     $160.00   1/12/94
846262 bucket seat             1 Gina  $75.00      $75.00   2/1/94
876543 pedals                  2 Roger $15.00      $30.00  12/26/93
934456 valve stems             4 Cindy   $.10       $.40   2/7/94
982528 frame                   1 Ann  $200.00     $200.00  11/28/94
999999 racing slicks           4 Al   $175.00     $700.00   12/9/93
```

FIG. 2

Prior Art

Owen Wiggs
Green & Mutala
PATENT AGENTS

2130033

0000000	00111111111122222222223333333333344				44444444555		5555555		6666666
1234567	89012345678901234567890123456789 01				23456789012		3456789		0123456
004234	control cable (inches)	115	Cindy		$.05		$5.75		12/6/93
121212	antilock brake units	4	Ann	$37	5.00	$15	00.00		1/7/94
125402	spoiler	1	Roger	$5	0.00	$5	00.00		2/1/94
200009	65 hp. engine	1	Al	$45	0.00	$4	50.00	1	2/27/93
221122	heavy duty bumpers	2	Ty	$12	0.00	$2	40.00		12/2/94
234102	titanium-cone speakers	4	Ann	$20	0.00	$8	00.00		1/7/94
284782	body	1	Ann	$22	5.00	$2	25.00	1	1/28/94
345353	traction control senso	4	Al	$72	0.00	$28	80.00		1/16/94
412344	tractor transmission	1	Ty	$27	5.00	$2	75.00		1/30/94
534523	steering carriage	1	David	$7	5.00	$	75.00	1	1/30/93
560000	seat belt	2	Cindy	$2	5.00	$	50.00	1	2/16/93
663661	axles	2	Sue	$21	5.00	$4	30.00	1	2/16/93
747474	steering wheel	1	Roger	$3	0.00	$	30.00		2/3/94
746744	CD changer	1	Al	$55	0.00	$5	50.00		2/15/94
787686	mag wheels	4	Gina	$4	0.00	$1	60.00		1/12/94
846262	bucket seat	1	Gina	$7	5.00	$	75.00		2/1/94
876543	pedals	2	Roger	$1	5.00	$	30.00	1	2/26/93
934456	valve stems	4	Cindy		$.10		$.40		2/7/94
982528	frame	1	Ann	$20	0.00	$2	00.00	1	1/28/94
999999	racing slicks	4	Al	$17	5.00	$7	00.00		12/9/93
301				302		303		304	

Prior Art

FIG. 3

Oyen Wiggs
Green & Mutala
PATENT AGENTS

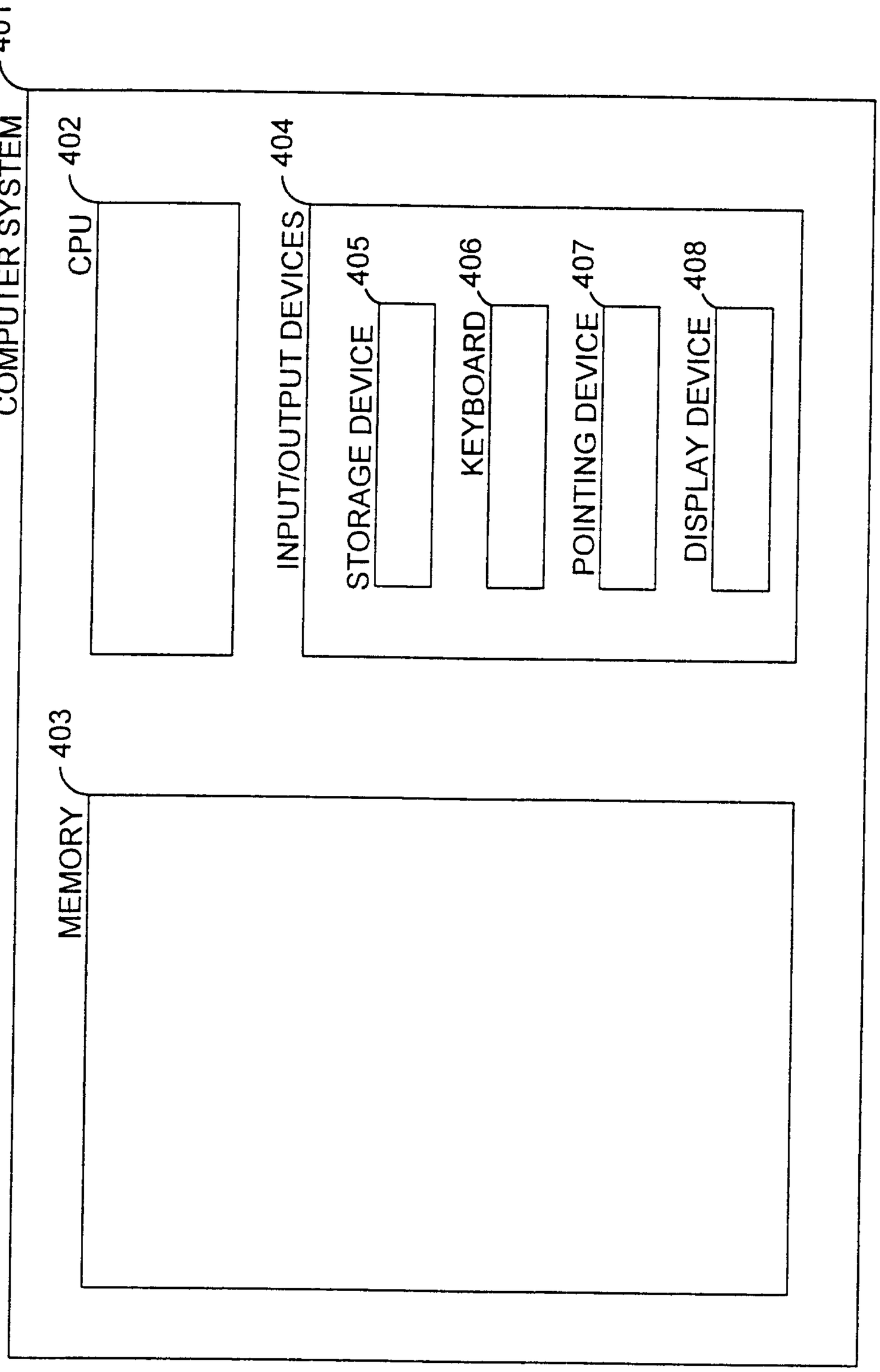


FIG. 4

Oyen Wiggs
Green & Mutala
PATENT AGENTS

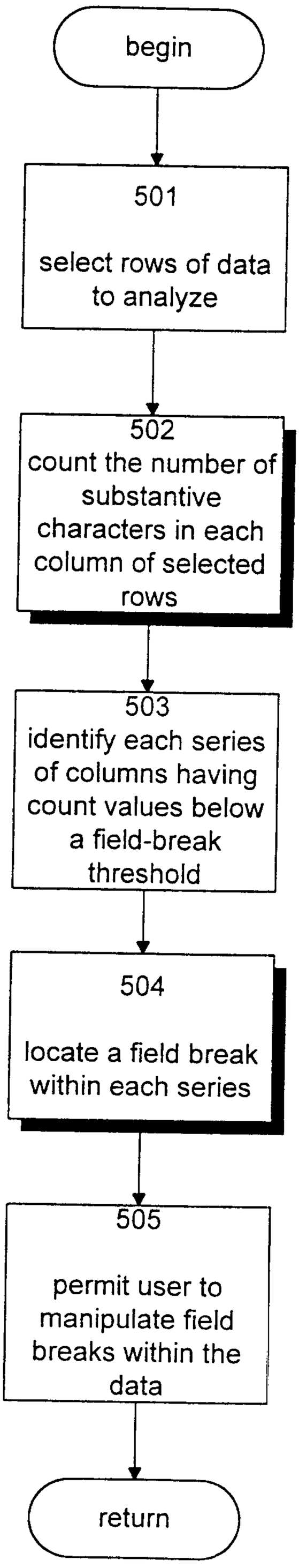


FIG. 5

Oyen Wiggs Green & Mutala
PATENT AGENTS

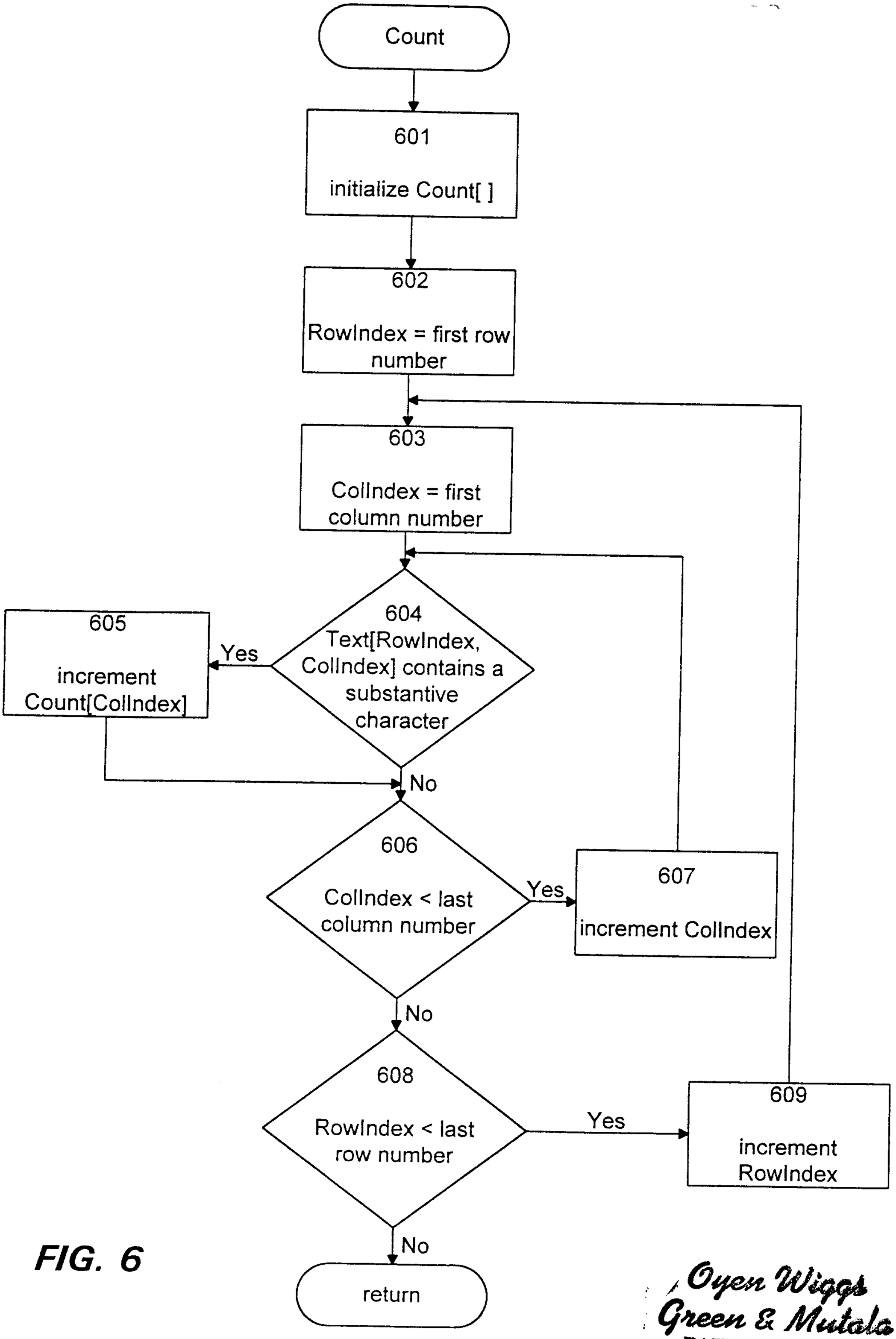


FIG. 6

Oyen Wiggs Green & Mutala
PATENT AGENTS

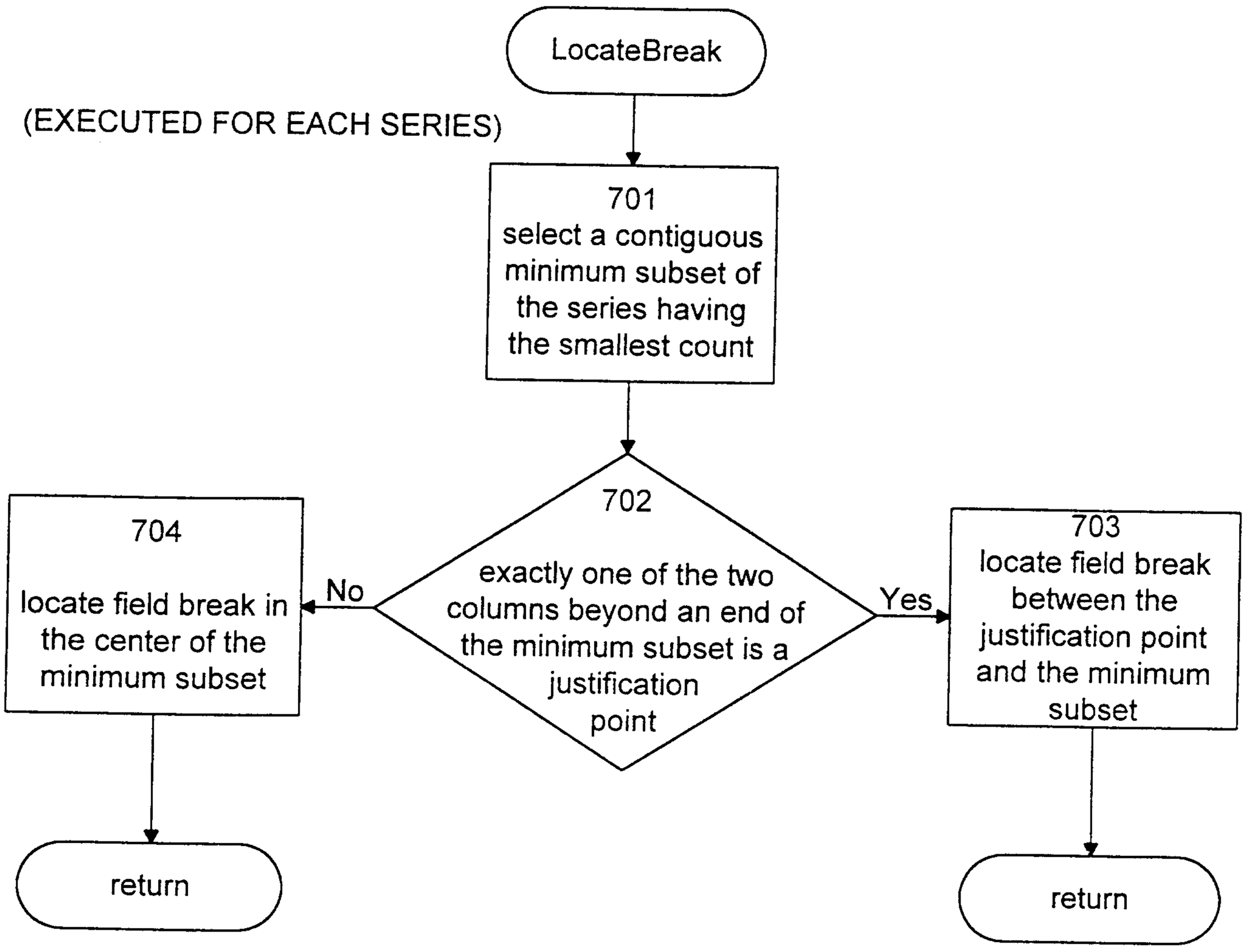


FIG. 7

Oyen Wiggs
Green & Mutala
PATENT AGENTS

800

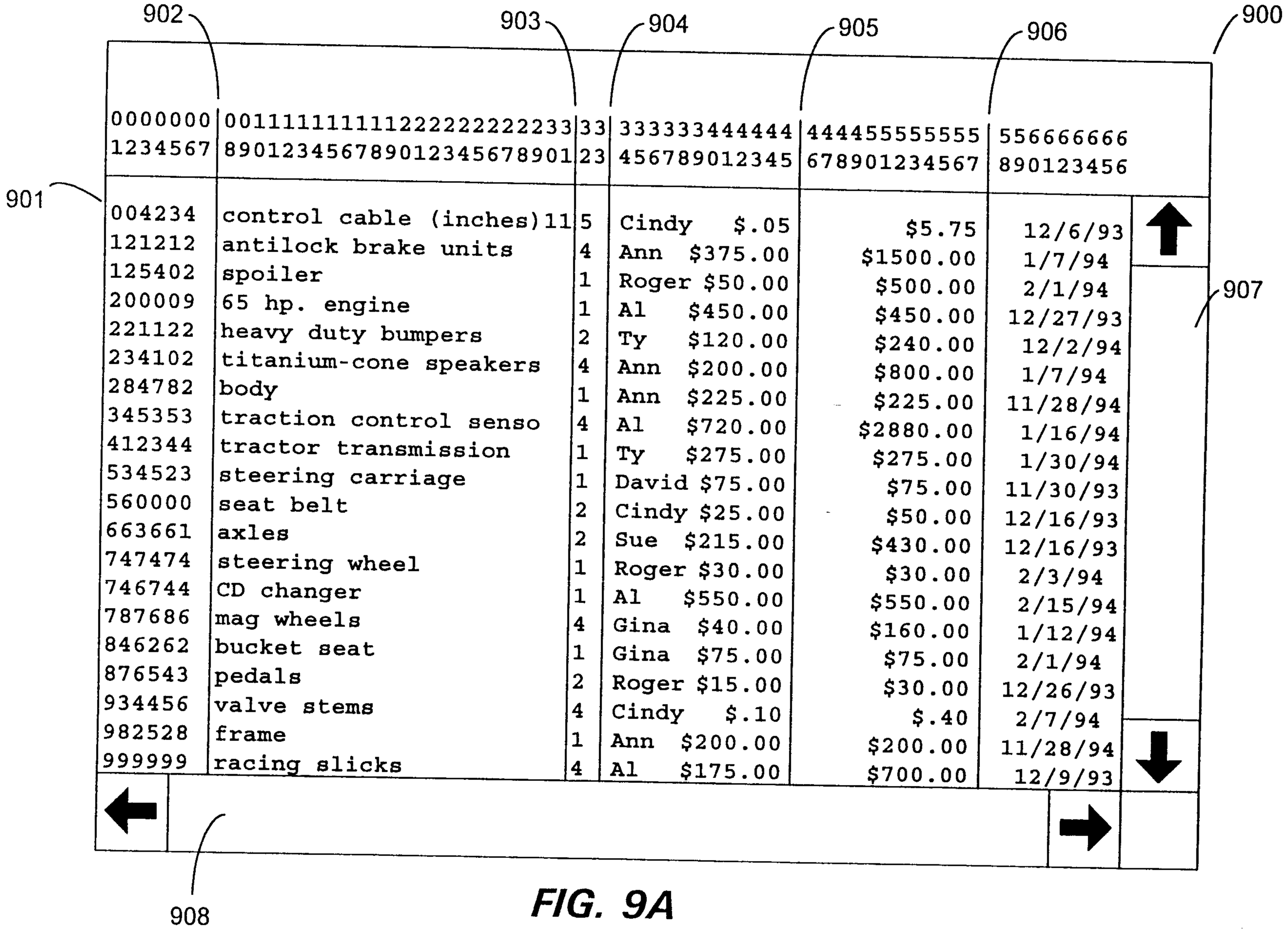
COUNT.XLS

column	count
1	20
2	20
3	20
4	20
5	20
6	20
7	0
8	20
9	20
10	18
11	19
12	18
13	15
14	13
15	13
16	11
17	14
18	11
19	10
20	10
21	6
22	6
23	7
24	6
25	6
26	5
27	5
28	3
29	3
30	1
31	1
32	20
33	0
34	20
35	20
36	14
37	9
38	7
39	10
40	18

column	count
41	18
42	20
43	20
44	20
45	20
46	0
47	0
48	0
49	0
50	2
51	13
52	18
53	19
54	20
55	20
56	20
57	20
58	0
59	7
60	20
61	20
62	20
63	20
64	20
65	20
66	14

FIG. 8

Oyen Wiggs
Green & Mutala
PATENT AGENTS

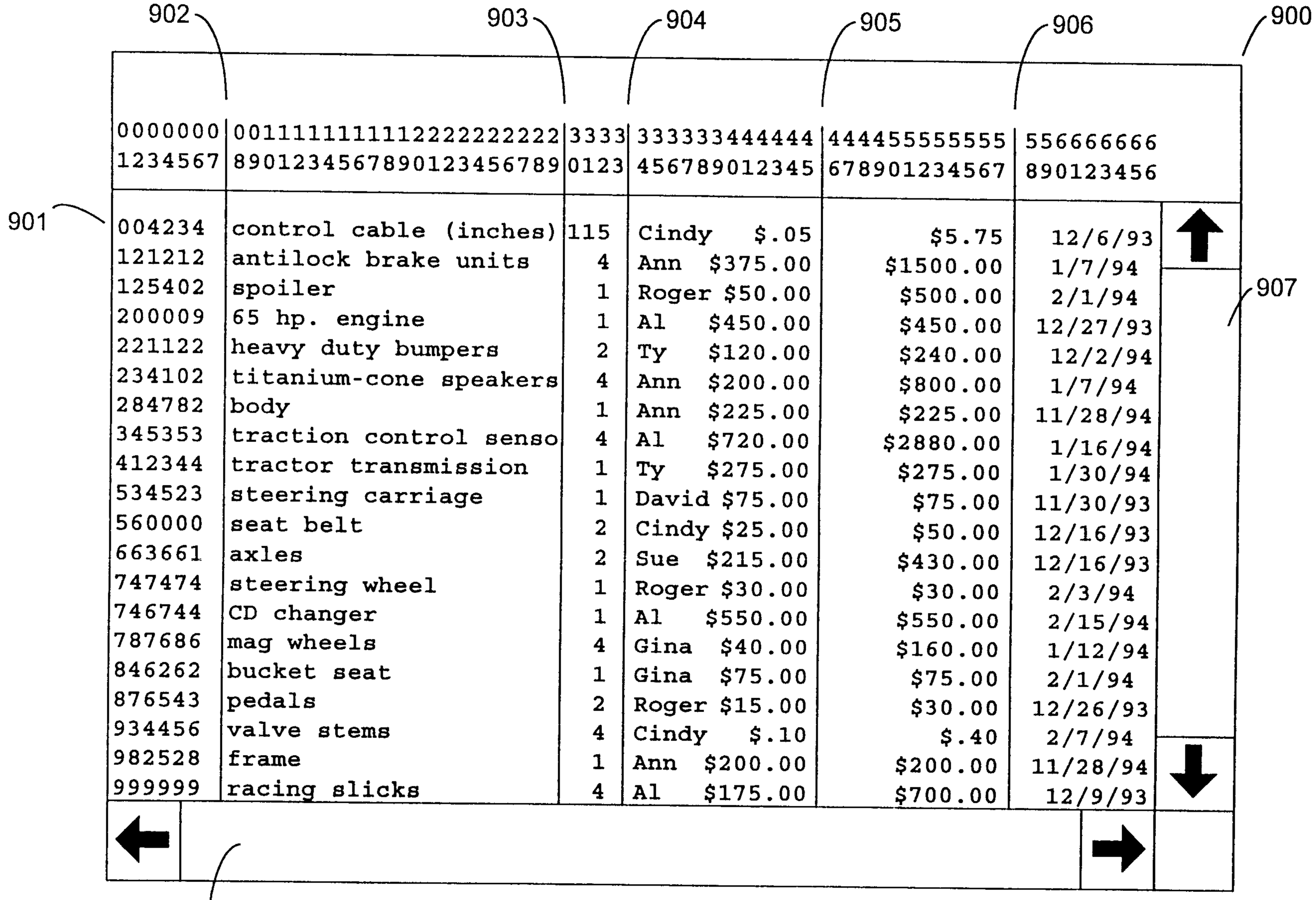


0000000 1234567	00111111111122222222233 89012345678901234567890 1	33 23	333333444444 456789012345	444455555555 678901234567	556666666 890123456
004234	control cable (inches)	11	5 Cindy $.05	$5.75	12/6/93
121212	antilock brake units		4 Ann $375.00	$1500.00	1/7/94
125402	spoiler		1 Roger $50.00	$500.00	2/1/94
200009	65 hp. engine		1 Al $450.00	$450.00	12/27/93
221122	heavy duty bumpers		2 Ty $120.00	$240.00	12/2/94
234102	titanium-cone speakers		4 Ann $200.00	$800.00	1/7/94
284782	body		1 Ann $225.00	$225.00	11/28/94
345353	traction control senso		4 Al $720.00	$2880.00	1/16/94
412344	tractor transmission		1 Ty $275.00	$275.00	1/30/94
534523	steering carriage		1 David $75.00	$75.00	11/30/93
560000	seat belt		2 Cindy $25.00	$50.00	12/16/93
663661	axles		2 Sue $215.00	$430.00	12/16/93
747474	steering wheel		1 Roger $30.00	$30.00	2/3/94
746744	CD changer		1 Al $550.00	$550.00	2/15/94
787686	mag wheels		4 Gina $40.00	$160.00	1/12/94
846262	bucket seat		1 Gina $75.00	$75.00	2/1/94
876543	pedals		2 Roger $15.00	$30.00	12/26/93
934456	valve stems		4 Cindy $.10	$.40	2/7/94
982528	frame		1 Ann $200.00	$200.00	11/28/94
999999	racing slicks		4 Al $175.00	$700.00	12/9/93

FIG. 9A

Owen, Wickersham & Erickson
PATENT AGENTS

2130033



902	903	904	905	906	900
0000000	00111111111112222222222	3333	333333444444	444455555555	556666666
1234567	89012345678901234567890	0123	456789012345	678901234567	890123456

901

004234	control cable (inches)	115	Cindy $.05	$5.75	12/6/93
121212	antilock brake units	4	Ann $375.00	$1500.00	1/7/94
125402	spoiler	1	Roger $50.00	$500.00	2/1/94
200009	65 hp. engine	1	Al $450.00	$450.00	12/27/93
221122	heavy duty bumpers	2	Ty $120.00	$240.00	12/2/94
234102	titanium-cone speakers	4	Ann $200.00	$800.00	1/7/94
284782	body	1	Ann $225.00	$225.00	11/28/94
345353	traction control senso	4	Al $720.00	$2880.00	1/16/94
412344	tractor transmission	1	Ty $275.00	$275.00	1/30/94
534523	steering carriage	1	David $75.00	$75.00	11/30/93
560000	seat belt	2	Cindy $25.00	$50.00	12/16/93
663661	axles	2	Sue $215.00	$430.00	12/16/93
747474	steering wheel	1	Roger $30.00	$30.00	2/3/94
746744	CD changer	1	Al $550.00	$550.00	2/15/94
787686	mag wheels	4	Gina $40.00	$160.00	1/12/94
846262	bucket seat	1	Gina $75.00	$75.00	2/1/94
876543	pedals	2	Roger $15.00	$30.00	12/26/93
934456	valve stems	4	Cindy $.10	$.40	2/7/94
982528	frame	1	Ann $200.00	$200.00	11/28/94
999999	racing slicks	4	Al $175.00	$700.00	12/9/93

907

908

FIG. 9B

Oyen Wiggs
Green & Mutala
PATENT AGENTS

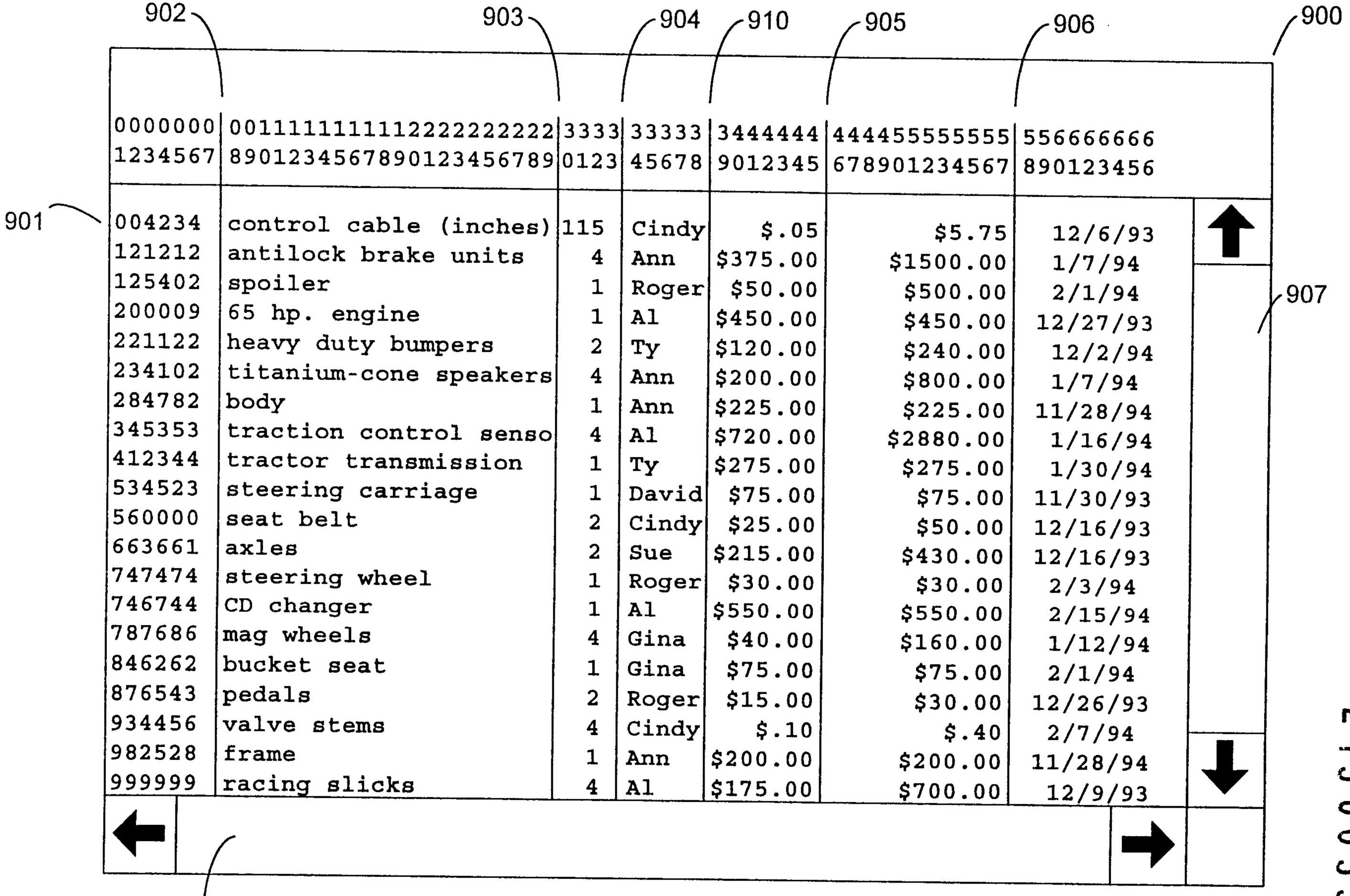


FIG. 9C

Cyen Wiggs Green & Mutala
PATENT AGENTS

```
0000000001111111111222222222233333333334444444444555555555566666666
1234567890123456789012345678901234567890123456789012345678901234567890123456
```

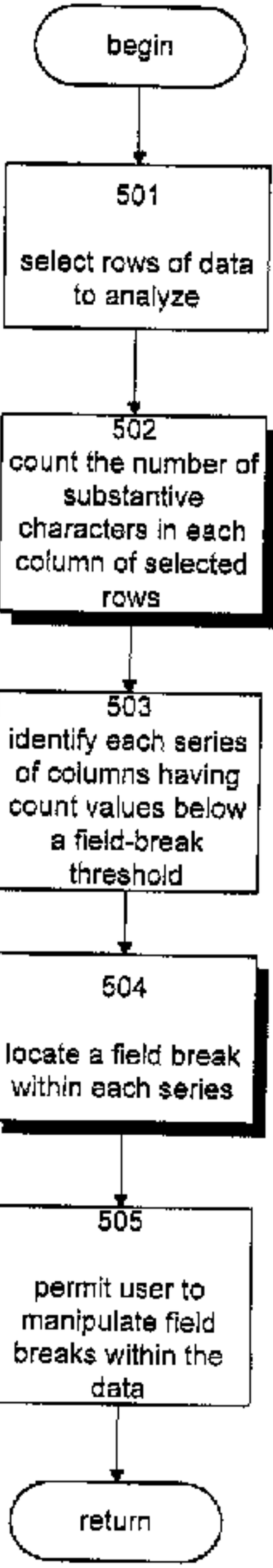
1001 —

1000

STK #	DESCRIPTION	QTY	MECH	ITM CST	TOTAL COST	DATE REQD
004234	control cable (inches)	115	Cindy	$.05	$5.75	12/6/93
121212	antilock brake units	4	Ann	$375.00	$1500.00	1/7/94
125402	spoiler	1	Roger	$50.00	$500.00	2/1/94
200009	65 hp. engine	1	Al	$450.00	$450.00	12/27/93
221122	heavy duty bumpers	2	Ty	$120.00	$240.00	12/2/94
234102	titanium-cone speakers	4	Ann	$200.00	$800.00	1/7/94
284782	body	1	Ann	$225.00	$225.00	11/28/94
345353	traction control senso	4	Al	$720.00	$2880.00	1/16/94
412344	tractor transmission	1	Ty	$275.00	$275.00	1/30/94
534523	steering carriage	1	David	$75.00	$75.00	11/30/93
560000	seat belt	2	Cindy	$25.00	$50.00	12/16/93
663661	axles	2	Sue	$215.00	$430.00	12/16/93
747474	steering wheel	1	Roger	$30.00	$30.00	2/3/94
746744	CD changer	1	Al	$550.00	$550.00	2/15/94
787686	mag wheels	4	Gina	$40.00	$160.00	1/12/94
846262	bucket seat	1	Gina	$75.00	$75.00	2/1/94
876543	pedals	2	Roger	$15.00	$30.00	12/26/93
934456	valve stems	4	Cindy	$.10	$.40	2/7/94
982528	frame	1	Ann	$200.00	$200.00	11/28/94
999999	racing slicks	4	Al	$175.00	$700.00	12/9/93

FIG. 10

Oyen Wiggs Green & Mutala
PATENT AGENTS


begin
501
select rows of data to analyze
502
count the number of substantive characters in each column of selected rows
503
identify each series of columns having count values below a field-break threshold
504
locate a field break within each series
505
permit user to manipulate field breaks within the data
return