

(72) VAN TONDER, MARIUS, ZA

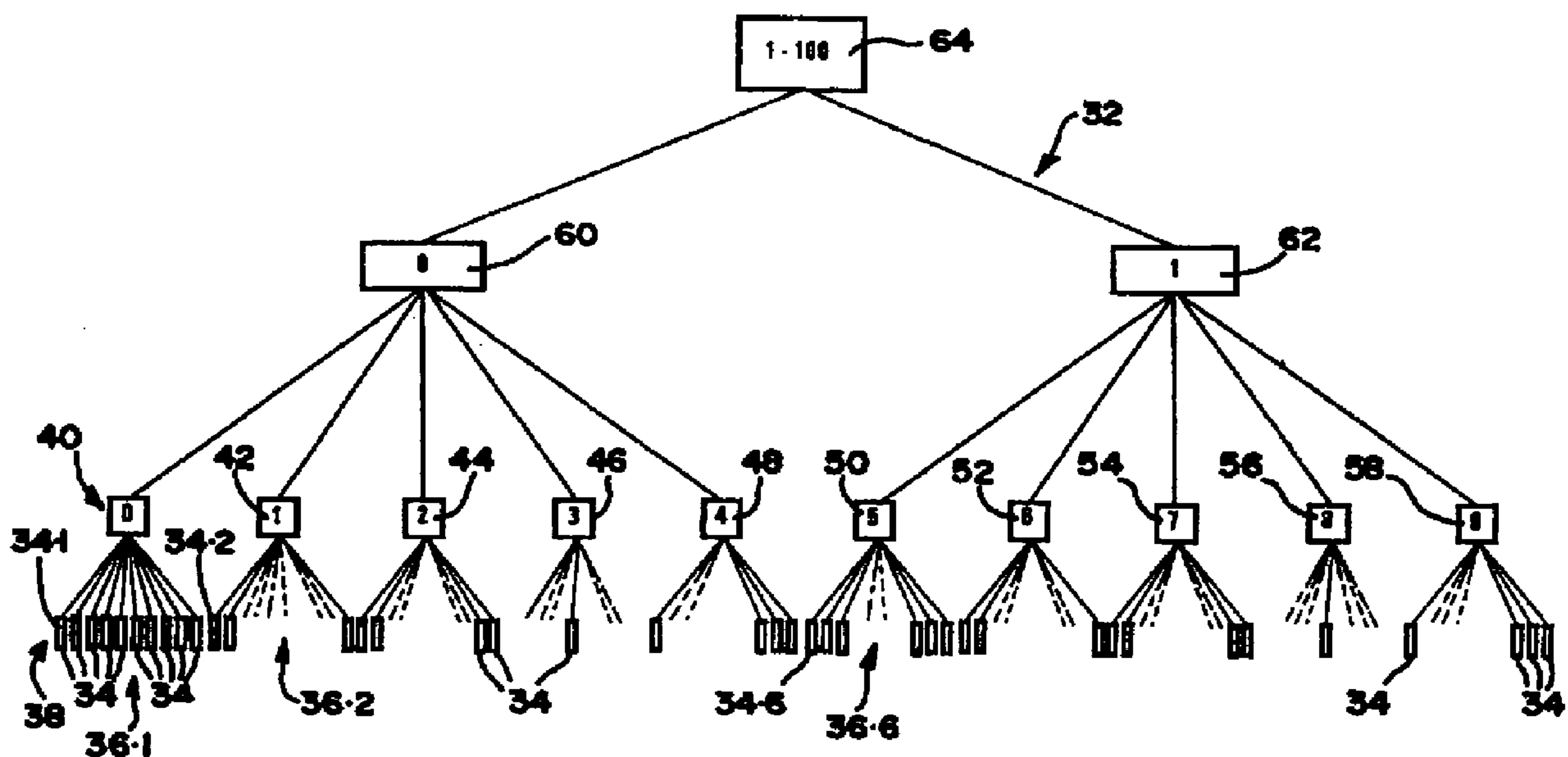
(71) TOP INFO OUTSOURCING SERVICES (PROPRIETARY) LIMITED,  
ZA

(51) Int.Cl.<sup>6</sup> G06F 17/30

(30) 1997/11/03 (97/9873) ZA

(54) **EXTRACTION ET STOCKAGE DE DONNEES A L'AIDE  
D'IDENTIFICATEURS UNIQUES**

(54) **DATA STORAGE AND RETRIEVAL USING UNIQUE  
IDENTIFIERS**



(57) La présente invention concerne un agencement de données dans une base de données. Cet agencement comprend un nombre sélectionné d'implantations (34) de fichiers, chaque implantation (34) comprenant un document pourvu d'un numéro (34-1) d'identification primaire unique. L'agencement comprend, en outre, une pluralité de groupes (36) d'implantation de fichiers, chaque groupe (36) comprenant un nombre spécifique de fichiers NFG et un numéro (40) d'identification secondaire unique défini par la valeur absolue du premier numéro d'identification primaire unique d'une implantation de fichiers dudit groupe partagé par des fichiers NFG. Le nombre d'implantation de fichiers est un nombre présélectionné qui correspond à un nombre de documents qui peuvent être stockés dans au moins une section particulière de la base de données.

(57) An arrangement of data in a database is provided. The arrangement includes a selected number of file locations (34), each file location (34) including a document which includes a unique primary identification number (34-1). The arrangement further includes a plurality of groups (36) of the file locations, each group (36) including a specific number of locations NFG for files in the group and including a unique secondary identification number (40) which is defined by the absolute value of the first unique primary identification number of a file location in said group of file locations divided by NFG files. The number of a file locations is a preselected number of file locations which corresponds to a number of documents which are capable of being stored in at least a particular section of the database.





PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

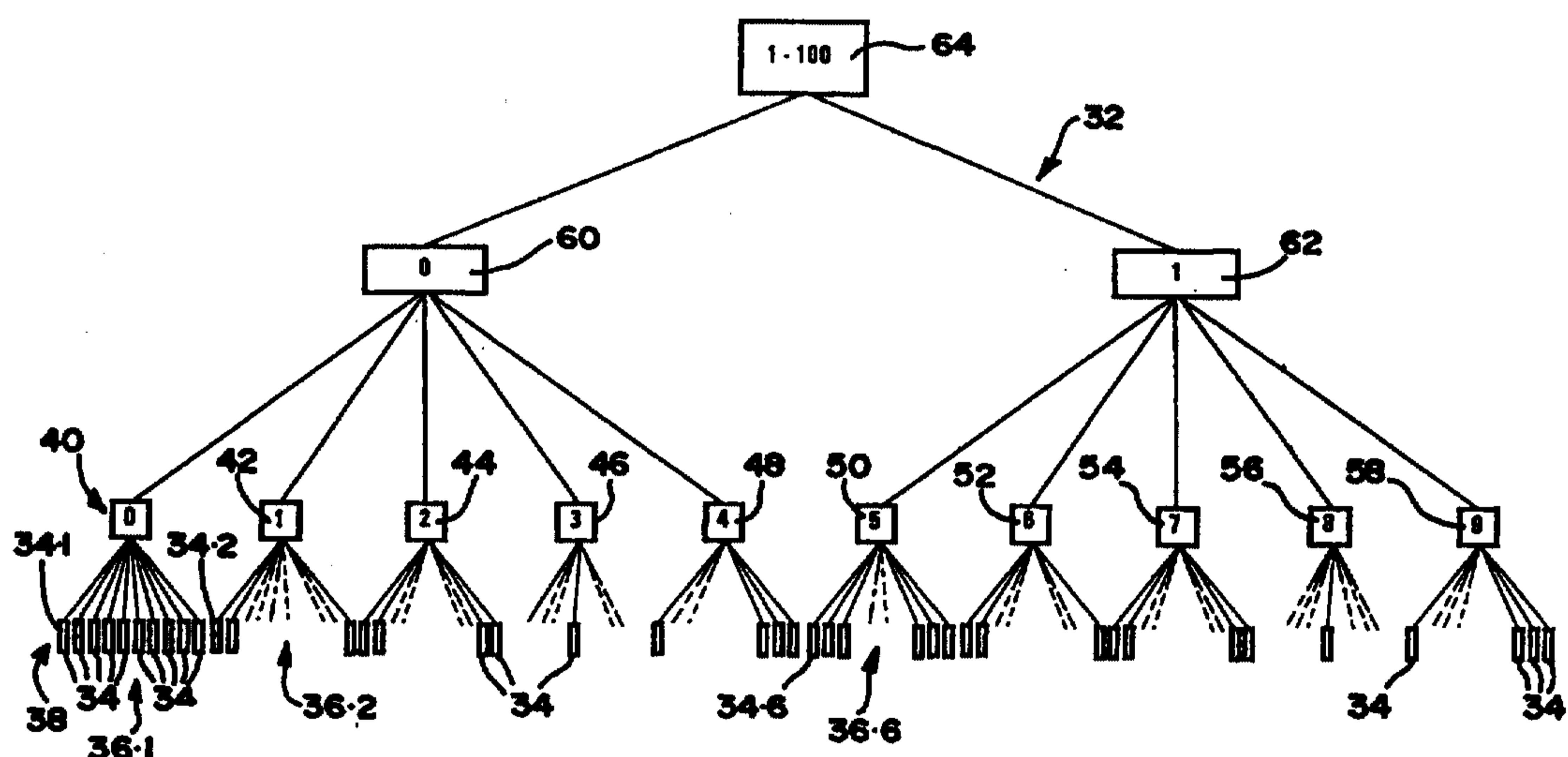
INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification <sup>6</sup> : <b>G06F 17/30</b></p>	<p><b>A1</b></p>	<p>(11) International Publication Number: <b>WO 99/23583</b> (43) International Publication Date: 14 May 1999 (14.05.99)</p>
--	------------------	--

(21) International Application Number: PCT/US98/23125  
 (22) International Filing Date: 30 October 1998 (30.10.98)  
 (30) Priority Data:  
 97/9873 3 November 1997 (03.11.97) ZA  
 (71) Applicant (for all designated States except US): TOP INFO  
 OUTSOURCING SERVICES (PROPRIETARY) LIMITED  
 [ZA/ZA]; 39 Selati Street, Ashlea Gardens, Pretoria 0081  
 (ZA).  
 (71) Applicant (for MW only): HANDELMAN, Joseph, H. [US/US];  
 26 West 61st Street, New York, NY 10023 (US).  
 (72) Inventor; and  
 (75) Inventor/Applicant (for US only): VAN TONDER, Marius  
 [ZA/ZA]; 25 Jaqueline Avenue, The Reeds, Pretoria 0158  
 (ZA).  
 (74) Agents: HANDELMAN, Joseph, H.; Ladas & Parry, 26 West  
 61st Street, New York, NY 10023 (US) et al.

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR,  
 BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD,  
 GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR,  
 KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN,  
 MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,  
 SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW,  
 ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW),  
 Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
 European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,  
 GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF,  
 BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN,  
 TD, TG).  
**Published**  
 With international search report.  
 Before the expiration of the time limit for amending the  
 claims and to be republished in the event of the receipt of  
 amendments.

(54) Title: DATA STORAGE AND RETRIEVAL USING UNIQUE IDENTIFIERS



(57) Abstract  
 An arrangement of data in a database is provided. The arrangement includes a selected number of file locations (34), each file location (34) including a document which includes a unique primary identification number (34-1). The arrangement further includes a plurality of groups (36) of the file locations, each group (36) including a specific number of locations NFG for files in the group and including a unique secondary identification number (40) which is defined by the absolute value of the first unique primary identification number of a file location in said group of file locations divided by NFG files. The number of a file locations is a preselected number of file locations which corresponds to a number of documents which are capable of being stored in at least a particular section of the database.

## DATA STORAGE AND RETRIEVAL USING UNIQUE IDENTIFIERS

FIELD OF THE INVENTION

THIS INVENTION relates to data storage and retrieval. It relates in particular to a method of storing a plurality of documents in a database and to a method of retrieving data from a database. Further, it relates to an arrangement  
5 of data in a database.

DESCRIPTION OF THE PRIOR ART

The storage of data in the form of digital images of documents has played an increasing role in recent years with improvements in computer technology. A typical application of the storage of documents in the form of  
10 digital images is in the medical field e.g. the storage of a claim document submitted by a doctor to a medical aid fund. When such claims are received by the medical aid fund, a digital image of the original claim document is stored in a storage medium e.g. a CD ROM or the like. Selected information is also manually read from the claim document and entered into an independent storage  
15 medium thereby creating an abridgement of the original document. Retrieval of the information from the independent storage medium is generally fairly rapid. However, in the event of full details on the claim being required, the digital image of the original document is usually required. This normally entails obtaining a document reference from the abridgement and retrieving the original document by  
20 means of a independent computing or data management system using conventional search techniques. The result is that, due to the substantial size of the database in which the digital images are stored, the retrieval of the digital image of the original document may take an unacceptably long period of time.

It is an object of this invention to offer a solution to this problem. It is however to be appreciated that not only medical records, but also insurance records, or any other bulk record systems, which are conventionally stored in large electronic databases are to be borne in mind for the purposes of this specification.

## 5 SUMMARY OF THE INVENTION

According to the invention, there is provided an arrangement of data in a database, the arrangement including

a selected number of file locations, each file location including a document which includes a unique primary identification number and the file location being  
10 identified from the unique primary identification number; and

a plurality of groups of the file locations, each group including NFG files in the group and including a unique secondary identification number which is defined by the absolute value of the first unique primary identification number of a file location in said group of file locations divided by NFG files.

15 Accordingly, the path to a selected document may be derived from the unique primary identification number.

The number of file locations may be a preselected number of file locations which corresponds to a number of documents which are capable of being stored in at least a particular section of the database.

20 The file locations may include digital images of the documents which are captured by means of a conventional scanner.

The file locations may be defined by a preselected number of base directories, each group of the file locations including NFG base directories and each base directory being designated by a primary identification number. The

number of base directories in each group of file locations may be less than about 1000, preferably less than about 250.

Each secondary identification number may be associated with a directory used in a conventional computer system, the directory being designated  
5 by the secondary identification number and each base directory being defined by a sub-directory of said directory. The primary identification number is typically the document number and the documents are preferably sequentially numbered.

Thus, each unique secondary identification number may be associated with, typically being the name or label of, a directory of a conventional  
10 directory/sub-directory arrangement used in conventional computer systems and each unique primary identification number may be associated with, typically being the name or label of, a sub-directory of said directory. Accordingly, each file location may be a sub-directory in which a digital image of the document is stored and which is labelled or named with the unique primary identification number  
15 associated with the document. The unique primary identification number is typically the document number.

The database is typically arranged in a hierarchical or so-called "root" structure of directories and in which the file locations are each defined by a sub-directory at a base level  $L_B$  in the hierarchical structure. Each group of directories  
20 at one level above the base level (level  $L_{B+1}$ ) in the structure may include  $NFGL_{B+1}$  sub-directories each of which has a secondary identification number designated by the absolute value of the document number of the first document in the group divided by  $NFGL_{B+1}$ .

Each level  $L_{B+n}$  may include a plurality of groups of directories, each  
25 group of directories including  $NFGL_{B+n}$  directories at an immediately lower level  $L_{B+n-1}$ . Each group of directories at level  $L_{B+n}$  may include a unique secondary identification number which is defined by the absolute value of the unique

4

secondary identification number of the first sub-directory in the group of directories at level  $L_{B+n-1}$  divided by  $NFGL_{B+n}$ .

The number of groups of directories at level  $L_{B+n}$  is typically between about 2 and about 10 times the number of groups of directories at level  $L_{B+n-1}$ . Preferably, an even number of groups of directories  $NFGL_{B+n}$  is provided at each level  $L_{B+n}$ .

Further in accordance with the invention, there is provided a method of storing a plurality of documents in a database which includes a selected number of file locations, each file location including a document which includes a unique primary identification number; and a plurality of groups of the file locations, each group including NFG files in the group and including a unique secondary identification number which is defined by the absolute value of the first unique primary identification number of a file location in a particular group of file locations divided by NFG files, the method including

15 identifying the unique primary identification number of each document to be stored in the database;

determining the secondary identification number by taking the absolute value of the primary identification number of the document to be stored and dividing it by NFG; and

20 storing the document in a file location in the form of a directory which is identified from the unique primary and secondary identification numbers.

The database may be arranged in a hierarchical directory structure. Accordingly, the method may include iteratively identifying an associated directory of a group of directories at one level higher (level  $L_{B+n+1}$ ) by dividing the unique secondary identification number of the directory at level  $L_{B+n}$  by the number of directories  $NFGL_{B+n+1}$  in the group.

The method include scanning an original copy of the document to obtain a digital image thereof, and storing the digital image of the document in the file location.

Still further in accordance with the invention, there is provided a  
5 method of identifying a path to one of a plurality of file locations in a database  
which includes a selected number of file locations, each file location including a  
document which includes a unique primary identification number and the file  
location being identified from the unique primary identification number; and a  
plurality of groups of the file locations, each group including NFG files in the  
10 group and including a unique secondary identification number which is defined by  
the absolute value of the first unique primary identification number of a file  
location in a particular group of file locations divided by NFG files, the method  
including

identifying the unique primary identification number of the document to be  
15 retrieved from the database; and

dividing the unique primary identification number by the number of file  
locations NFG in the group and taking the absolute value of the result to obtain  
the unique secondary identification number of the group in which the document  
lies thereby to identify the path to the selected document.

20 The database may be arranged in a hierarchical directory structure.  
Accordingly, the method may include iteratively identifying an associated directory  
of a group of directories at one level higher (level  $L_{B+n+1}$ ) by dividing the unique  
secondary identification number of the directory at level  $L_{B+n}$  by the number of  
directories  $NFGL_{B+n+1}$  in the group.

25 Further in accordance with the invention, there is provided a method  
of retrieving data from a database which includes a selected number of file  
locations, each file location including a document which includes a unique primary  
identification number and the file location being identified from the unique primary  
identification number; and a plurality of groups of the file locations, each group

including NFG files in the group and including a unique secondary identification number which is defined by the absolute value of the first unique primary identification number of a file location in a particular group of file locations divided by NFG files, the method including

- 5 identifying the unique primary identification number of the document to be retrieved from the database, and dividing the number of file locations NFG in the group by the unique primary identification number, and taking the absolute value of the result to obtain the unique secondary identification number of the group in which the document lies thereby to identify the path to the selected document;
- 10 and
- reading the data stored in the directory via the path.

The unique primary identification number may be associated with a name of a legal entity e.g. a natural person, a business, or the like. The method may include searching for the name of the legal entity in a conventional manner

15 and retrieving the unique primary identification number thereby to identify a name and path of the directory in which the document has been stored.

Further in accordance with the invention, there is provided a data management installation which includes

- reading means for reading data from data storage means which includes a
- 20 digital image of a plurality of documents arranged in an hierarchical structure in a database which includes a selected number of file locations, each file location including a document which includes a unique primary identification number and the file location being identified from the unique primary identification number, and a plurality of groups of the file locations, each group including NFG files in the
- 25 group and including a unique secondary identification number which is defined by the absolute value of the first unique primary identification number of a file location in a particular group of file locations divided by NFG files;
- input means for receiving the primary identification number of a document to be retrieved from the storage means; and

processing means arranged to identify a path in the hierarchical structure to the file location in which the document has been stored, the path being derived from a unique primary identification number of the document.

The installation may include interface means for interfacing the  
5 installation to a conventional data management installation which selectively accesses abridgements of documents e.g. abridgements of medical aid claims or the like which have been stored in the form of a digital image.

The installation may be arranged to receive a document number from the abridgement, the document number being translated into a unique primary  
10 identification number thereby to permit a digital image of the entire original document to be retrieved.

The storage means may be a plurality of CD ROMs which define the database, the reading means being a so-called "CD jukebox".

#### BRIEF DESCRIPTION OF THE DRAWINGS

15 The invention is now described, by way of example, with reference to the accompanying diagrammatic drawings.

#### DESCRIPTION OF THE PREFERRED EMBODIMENT

In the drawings,

Figure 1 shows a schematic diagram of a data management installation in  
20 accordance with the invention;

Figure 2 shows a schematic diagram of an arrangement of data in a database, also in accordance with the invention, of the installation of Figure 1; and

Figure 3 shows a sub-section of a directory structure which is arranged in a similar fashion to that of Figure 2.

Referring to the drawings, reference numeral 10 generally indicates a data management installation in accordance with the invention. The installation 5 10 includes a data capturing sub-section 12, a data storage sub-section 14, and a data retrieval sub-section 16. The installation 10 is configured or arranged to store a digital image of each of a substantial number of documents for subsequent retrieval from a database arrangement as described in more detail below.

The data capturing sub-system 12 includes a conventional digital 10 scanner 18 which scans a substantial number of documents 20 and feeds a digital image of each document 20 into storage means 22 as shown by arrow 24. Once the documents 20 have been scanned, they are physically stored in a warehouse or discarded as indicated by arrow 26.

The documents 20 are numbered sequentially and each number 15 defines a unique primary identification number which is associated with a particular document 20. The installation 10 includes interface means as indicated by arrow 28, for interfacing the installation 10 to a conventional computer system 30 which is configured to access abridgements of the documents 20 in its storage means (not shown). In particular, the documents 20 are typically claim forms 20 received by a medical aid company from various medical practitioners. Conventionally, selected information from each document 20 is manually entered into the conventional computer system 30 to create an abridgement of the document 20 for subsequent retrieval by an operator. However, in certain 25 circumstances, merely accessing the abridgement is insufficient adequately to attend to any query pertaining to the document, and, accordingly, the operator would then in a conventional system physically retrieve the document 20 to obtain comprehensive information on a particular transaction or claim.

However, unlike conventional systems, the data management installation 10 stores each document 20 in a unique fashion in an arrangement of data in a database in the storage means 22. In particular, the database is arranged in an hierarchial or so-called "root" structure 32 (see Figure 2). The structure 32 of the arrangement includes 100 directories or file locations only a few of which are referenced in the drawings by reference numeral 34. The file locations or directories 34 are divided into a number of groups of directories 36.1, 36.2, 36.3 and so on, each comprising 10 directories 34 or number of files in the group (NFG). Each directory 34 has as its title or name the unique primary identification number of a document 20 intended to be stored therein. Thus, each document 20 is stored in a specific location to facilitate subsequent retrieval thereof.

The directories 34 are located at a base level  $L_B$  in the hierarchial structure 32 as indicated by arrow 38. The group of directories 36.1 is associated with a directory 40 at a level  $L_{B+1}$ , which is one level higher in the hierarchial structure 32. In a similar fashion, further directories 42 to 58 are provided at level  $L_{B+1}$  each of which are associated with 10 file locations (NFG) or directories 34 at level  $L_B$ , each file location or directory 34 bearing the name or label of the unique identification number of the document 20 to be stored therein. Further, as in the case of the directories 34 which are grouped into groups of directories 36, the directories 40 to 58 at level  $L_{B+1}$  are grouped into two groups of directories 60, 62, at a level  $L_{B+2}$ , each group having 5 (NFG $L_{B+2}$ ) sub-directories. Further, in a similar fashion, the groups of directories 60, 62 are grouped or branch out from a further directory 64 which bears a label "1-100" and which is thus representative of the range of documents 20 having unique primary identification numbers between 1 and 100 which are associated with the directory. The directory 64 has 2 (NFG $L_{B+3}$ ) directories in its group.

The various names of the directories 64, 60, 62, 40, to 58, and 34 are in the form of reference numerals which are allocated in a specific fashion. In particular, the name of the directory 40 defines a unique secondary

identification number which is defined by the absolute value of the first unique primary identification number 34.1 in the group of directories 36.1 divided by the total number of directories or file locations in NFG in the group of directories 36.1. For example, as the first document is stored in file location 34.1 it bears  
5 a unique primary identification number 0 and the name of the directory 40 is then defined by the absolute value of 0 divided by 10 which is 0 as show in Figure 2. In the case of the directory 42, its name or label is defined by the absolute value of the first unique primary identification number 34.2 in a second group of directories 36.2 divided by the number of files or directories in NFG in the  
10 particular group, i.e. the absolute value of 10 divided by 10 which is equal to 1. In a similar fashion, the unique secondary identification numbers which define the names of the directories 44 to 58 are determined.

In a similar fashion, the label or name of the group of directories 60 is defined by the absolute value of the unique secondary identification number "0"  
15 which is the name of the first group of directories 40 at an immediately lower level  $L_{B+n-1}$ , divided by the number of groups of directories at an immediately lower level, i.e. 5 thus providing a result of 0 as shown in Figure 2. In a similar fashion, the name of the group of directories 62 is derived by the first unique secondary identification number which is the file name of the group of directories  
20 50, i.e. 5 divided by 5 ( $NFGL_{B+2}$ ) which equals 1.

It is to be appreciated that, in other embodiments of the invention, the hierarchial structure may comprise a plurality of different levels. The number of different levels depends upon the number of documents which are to be stored in the hierarchy. Further, the fewer the number of levels, i.e. the flatter the  
25 hierarchial structure is, the more simple the path is to the particular directory in which the document is stored and thus retrieval times may be reduced in comparison to a very pointed hierarchial structure in which a number of levels are included.

When a substantial number of documents are to be stored, a plurality of hierarchial structures (one of which is shown in Figure 3) which are independent of each other may be used. Preferably, due to software and hardware limitations of certain computing systems, the number of file locations or directories 34 at the base level  $L_B$  in the hierarchial structure 38 is typically less than about 1000 and, more preferably, less than about 250. The hierarchial structure 38 may thus include a plurality of levels extending above base level  $L_B$ , each level including a group of directories at a level  $L_{B+n}$  having  $NFGL_{B+n}$  directories in the group. Thus, from a top down point of view, each directory in a group of directories branches out or extends into  $NFGL_{B+n}$  groups of directories at an immediately lower level  $L_{B+n-1}$ . The name or secondary identification number of each group of sub-directories is then determined in a similar fashion as described above. When arranging the database, the number of groups of directories at level  $L_{B+n}$  are typically between about 2 and about 10 times the number of groups of directories at level  $L_{B+n}$ . Thus, it is evident that the number of levels  $L_{B+n}$  is dependent upon the number of documents at the base level  $L_B$  in the hierarchial structure.

Once the hierarchial structure 38 has been established and the data has been arranged in the database as described above, a digital image of each document 20 is stored on a plurality of compact discs 70 as shown in Figure 1. The compact discs 70 may form part of a library of information on various transactions or claims which have been submitted to the medical aid via the various doctors. Certain of the compact discs may be loaded in a CD jukebox 72 to provide a near line facility and other compact discs may be loaded in a CD tower 74 to provide an on-line facility as shown by arrows 76, 78 respectively. In other embodiments of the invention, the database is stored on a magnetic media 80.

In order to retrieve a digital image of a specific document 20 from the database, the installation 10 includes computing means 82 (see Figure 1) which is arranged to generate a variety of user friendly screens to assist in

instructing the computing means 82 to perform various retrieval functions. The computing means 82 is programmed in such a fashion so that an indexed field window 84 prompts a user to enter a client name 86 via a keyboard (not shown). The computing means 82, in a conventional fashion, then retrieves the unique  
5 primary identification number 88 which is associated with the client name 86. The unique primary identification number is then fed to a unique key of documents screen 90 which has a search prompt 92 which may be activated with a mouse to initiate retrieval of a selected document from the database.

In order to facilitate retrieval of the digital image of the selected  
10 document from the database, the path to the particular file location or directory 34 in which the document has been stored is derived directly from the unique primary identification number which defines the name of the file location or directory 34 in which the document 20 has been stored. In order to determine this path, the relevant directories in the groups of directories, at the various levels  
15  $L_{B+n}$  in the hierarchial structure 32 must be determined. As described above, the name of the actual directory 34 in which the document 20 has been stored is determined as indicated above and the particular directory in the group of directories is then determined by taking the absolute value of the unique primary identification number or document number divided by the number of file locations  
20 or directories 34 NFG at base level  $L_B$ . Once the name of the particular directory at each intermediate level in the hierarchial structure 32 has been determined, the path to the relevant document may be reconstructed and thus retrieval time may be reduced. For example, to identify the path to document number 10 in the hierarchial structure 32, the unique primary identification number i.e. "10" is  
25 divided by the number of the number of groups of directories NFG at the base level  $L_B$  i.e. "10" and the absolute value thereof is taken, i.e. directory 42 labelled "1" is identified at level  $L_{B+1}$ . In a similar fashion, if document 15 is to be retrieved, the absolute value is taken of the unique primary identification number or document number 15 divided by the number of files in the group of directories  
30 36, i.e. the result is the absolute value of 1.5 which is 1. Likewise a particular

sub-directory at level  $L_{B+n}$  is determined and thus the path to the document may be determined.

The inventors believe that the invention, as illustrated, provides a data management system 10 which has enhanced retrieval characteristics of a document from a database as the documents are stored in a particular fashion and the actual path to the directory in which the document has been stored is derived from a unique primary identification number allocated to the document.

CLAIMS

1. An arrangement of data in a database, the arrangement including a selected number of file locations, each file location including a document which includes a unique primary identification number and the file location being  
5 identified from the unique primary identification number; and  
a plurality of groups of the file locations, each group including NFG files in the group and including a unique secondary identification number which is defined by the absolute value of the first unique primary identification number of a file location in said group of file locations divided by NFG files.
- 10 2. An arrangement of data as claimed in Claim 1, in which the number of file locations is a preselected number of file locations which corresponds to a number of documents which are capable of being stored in at least a particular section of the database.
3. An arrangement of data as claimed in Claim 2, in which the file  
15 locations include digital images of the documents which are captured by means of a conventional scanner.
4. An arrangement of data as claimed in Claim 1, in which the file locations are defined by a preselected number of base directories, each group of the file locations including NFG base directories and each base directory being  
20 designated by a primary identification number.
5. An arrangement of data as claimed in Claim 4, in which the number of base directories in each group of file locations is less than 1000.
6. An arrangement of data as claimed in Claim 5, in which each secondary identification number is associated with a directory used in a  
25 conventional computer system, the directory being designated by the secondary

identification number and each base directory being defined by a sub-directory of said directory.

7. An arrangement of data as claimed in Claim 6, in which the primary identification number is the document number and the documents are sequentially  
5 numbered.

8. An arrangement of data as claimed in Claim 7, in which the database is arranged in a hierarchical structure of directories and in which the file locations are each defined by a sub-directory at a base level  $L_B$  in the hierarchical structure, each group of directories at one level above the base level (level  $L_{B+1}$ ) in the  
10 structure including  $NFGL_{B+1}$  sub-directories each of which has a secondary identification number designated by the absolute value of the document number of the first document in the group divided by  $NFGL_{B+1}$ .

9. An arrangement of data as claimed in Claim 8, in which each level  $L_{B+n}$  includes a plurality of groups of directories, each group of directories  
15 including  $NFGL_{B+n}$  directories at an immediately lower level  $L_{B+n-1}$ , each group of directories at level  $L_{B+n}$  including a unique secondary identification number which is defined by the absolute value of the unique secondary identification number of the first sub-directory in the group of directories at level  $L_{B+n-1}$  divided by  $NFGL_{B+n}$ .

20 10. An arrangement of data as claimed in Claim 9, in which the number of groups of directories at level  $L_{B+n}$  is between 2 and 10 times the number of groups of directories at level  $L_{B+n-1}$ .

11. An arrangement of data as claimed in Claim 10, in which an even number of groups of directories  $NFGL_{B+n}$  is provided at each level  $L_{B+n}$ .

25 12. A method of storing a plurality of documents in a database which includes a selected number of file locations, each file location including a

document which includes a unique primary identification number; and a plurality of groups of the file locations, each group including NFG files in the group and including a unique secondary identification number which is defined by the absolute value of the first unique primary identification number of a file location

5 in a particular group of file locations divided by NFG files, the method including identifying the unique primary identification number of each document to be stored in the database;

determining the secondary identification number by taking the absolute value of the primary identification number of the document to be stored and

10 dividing it by NFG; and

storing the document in a file location in the form of a directory which is identified from the unique primary and secondary identification numbers.

13. A method as claimed in Claim 12, in which the database is arranged in a hierarchical directory structure, and the method includes iteratively identifying

15 an associated directory of a group of directories at one level higher (level  $L_{B+n+1}$ ) by dividing the unique secondary identification number of the directory at level  $L_{B+n}$  by the number of directories  $NFGL_{B+n+1}$  in the group.

14. A method as claimed in Claim 13, which includes scanning an original copy of the document to obtain a digital image thereof, and storing the

20 digital image of the document in the file location.

15. A method of identifying a path to one of a plurality of file locations in a database which includes a selected number of file locations, each file location including a document which includes a unique primary identification number and the file location being identified from the unique primary identification number;

25 and a plurality of groups of the file locations, each group including NFG files in the group and including a unique secondary identification number which is defined by the absolute value of the first unique primary identification number of a file location in a particular group of file locations divided by NFG files, the method including

identifying the unique primary identification number of the document to be retrieved from the database; and

dividing the unique primary identification number by the number of file locations NFG in the group and taking the absolute value of the result to obtain  
5 the unique secondary identification number of the group in which the document lies thereby to identify the path to the selected document.

16. A method as claimed in Claim 15, in which the database is arranged in a hierarchical directory structure, and the method includes iteratively identifying an associated directory of a group of directories at one level higher (level  $L_{B+n+1}$ )  
10 by dividing the unique secondary identification number of the directory at level  $L_{B+n}$  by the number of directories  $NFGL_{B+n+1}$  in the group.

17. A method of retrieving data from a database which includes a selected number of file locations, each file location including a document which includes a unique primary identification number and the file location being  
15 identified from the unique primary identification number; and a plurality of groups of the file locations, each group including NFG files in the group and including a unique secondary identification number which is defined by the absolute value of the first unique primary identification number of a file location in a particular group of file locations divided by NFG files, the method including  
20 identifying the unique primary identification number of the document to be retrieved from the database, and dividing the number of file locations NFG in the group by the unique primary identification number, and taking the absolute value of the result to obtain the unique secondary identification number of the group in which the document lies thereby to identify the path to the selected document;  
25 and

reading the data stored in the directory via the path.

18. A method as claimed in Claim 17, in which the unique primary identification number is associated with a name of a legal entity, the method including searching for the name of the legal entity in a conventional manner and

retrieving the unique primary identification number thereby to identify a name and path of the directory in which the document has been stored.

19. A data management installation which includes
- reading means for reading data from data storage means which includes a  
5 digital image of a plurality of documents arranged in an hierarchical structure in  
a database of data which includes a selected number of file locations, each file  
location including a document which includes a unique primary identification  
number and the file location being identified from the unique primary identification  
number, and a plurality of groups of the file locations, each group including NFG  
10 files in the group and including a unique secondary identification number which  
is defined by the absolute value of the first unique primary identification number  
of a file location in a particular group of file locations divided by NFG files;  
input means for receiving the primary identification number of a document  
to be retrieved from the storage means; and  
15 processing means arranged to identify a path in the hierarchical structure  
to the file location in which the document has been stored, the path being derived  
from a unique primary identification number of the document.

20. An installation as claimed in Claim 19, which includes interface  
means for interfacing the installation to a conventional data management  
20 installation which selectively accesses abridgements of documents.

21. An installation as claimed in Claim 20, which is arranged to receive  
a document number from the abridgement, the document number being translated  
into a unique primary identification number thereby to permit a digital image of the  
entire original document to be retrieved.

- 25 22. An installation as claimed in Claim 19, in which the storage means  
is a plurality of CD ROMs which define the database, the reading means being a  
so-called "CD jukebox".

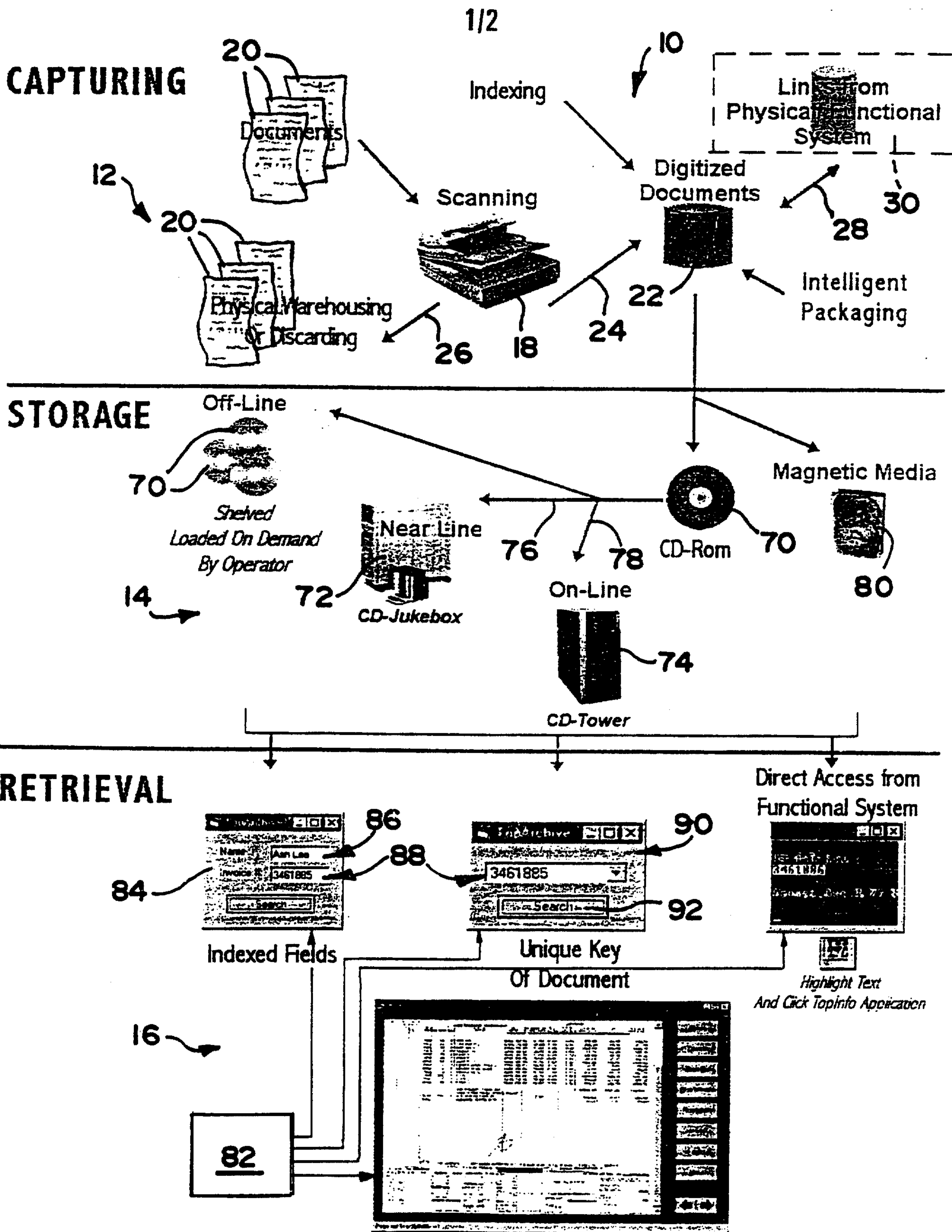


FIG 1

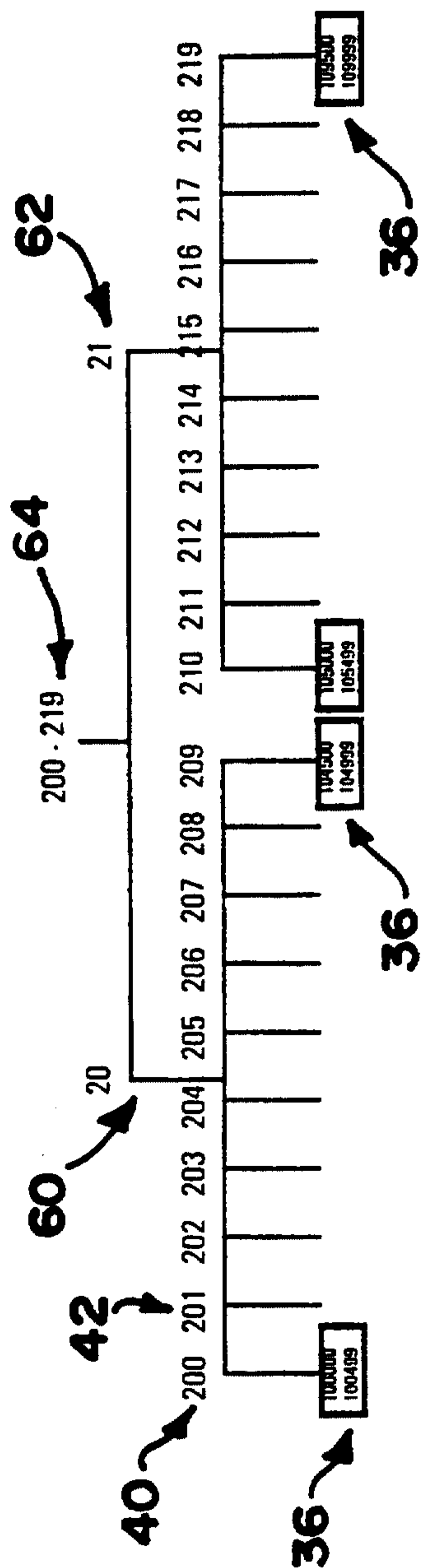


FIG 3

1.100

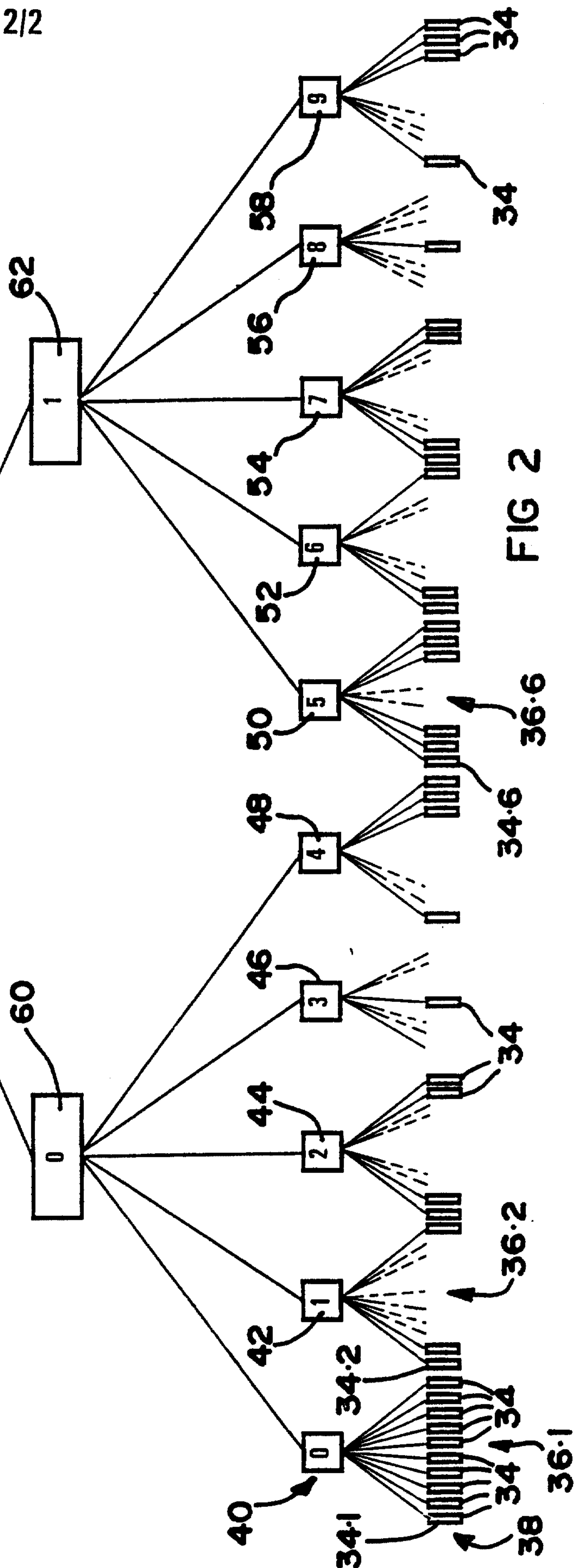


FIG 2