

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
3 August 2006 (03.08.2006)

PCT

(10) International Publication Number
WO 2006/079575 A1

(51) International Patent Classification:
H04L 12/58 (2006.01) **G06Q 10/00** (2006.01)

(21) International Application Number:
PCT/EP2006/050043

(22) International Filing Date: 4 January 2006 (04.01.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0501779.3 28 January 2005 (28.01.2005) GB

(71) Applicant (for all designated States except US): **INTERNATIONAL BUSINESS MACHINES CORPORATION** [US/US]; New Orchard Road, Armonk, New York 10504 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **BRAITHWAITE, Kevin, Anthony** [GB/GB]; 11 Crusader Road Hedge End, Southampton Hampshire SO30 0PD (GB). **DUNN, Tim** [GB/GB]; Beechridge Lippen Lane, Warnford Hampshire SO32 3LE (GB).

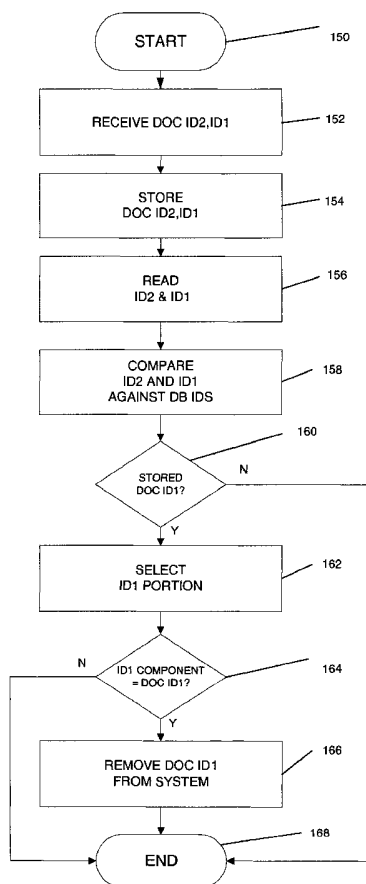
(74) Agent: **LITHERLAND, David, Peter**; IBM United Kingdom Limited, Intellectual Property Law, Hursley Park, Winchester Hampshire SO21 2JN (GB).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT,

[Continued on next page]

(54) Title: ELECTRONIC MAIL COMMUNICATION



(57) Abstract: A method is disclosed for managing emails in a communication network comprising a plurality of data processing systems. The method comprises generating an email at a first data processing system and assigning thereto a unique document ID; transferring the first email to a second of the data processing systems; generating at the second data processing system a second email incorporating the received email document and assigning to the second email another unique document ID and retaining the unique document ID of the received email; and transferring the second email including both unique document IDs to one of the data processing systems. These unique IDs may be employed to reduce the amount of storage required to store emails and/or to manage an email inbox by storing transmitted and received emails; identifying each stored email all whose ID(s) match some or all of the ID(s) of a received email; and deleting the stored email having the matched ID(s).

WO 2006/079575 A1



RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

ELECTRONIC MAIL COMMUNICATION**Field of the Invention**

5 The present invention relates generally to the field of communication via electronic mail (email) and more particularly to the management of email documents in email systems.

Background of the Invention

10 Email has become one of the most widely used business productivity applications. It is now commonplace using email systems to generate and send email documents and to reply to, or forward, received emails to other individuals. The forwarding and reforwarding of documents between a group
15 of individuals often results in a sequence of documents, commonly called a thread. Each individual may have a different subset of the complete thread depending on whether he or she has been copied on each document in the thread.

20 The growth in email communication and the need to manage large volumes of emails has led to frustration on the part of users. The amount of local and network storage required to store these emails has also increased - a problem which is exacerbated by the duplication of emails in original and forwarded copies of a thread. In corporate networks, it is common for a
25 limit to be placed on the amount of network storage allowed per user for email documents. This means that the user needs to carry out regular housekeeping operations on their email systems to avoid exceeding the storage limit.

30 There are a number of known techniques designed to address the problem of email storage. PCT published application WO 02/33981 describes one technique whereby redundant email messages are identified and removed from a user's message file. Identification of redundancy involves copying
35 messages stored in the message file into two different arrays and cleansing one of the arrays to remove formatting and header information. The cleansed messages are then compared to determine whether a particular message is wholly repeated in any other. Messages that are found to be repeated are removed from the other array. The messages remaining in the other array may then be substituted for the messages in the user's message
40 file, resulting in a minimized message file.

The management of a large email inbox can also be very time consuming, especially when the inbox has not been processed for a long period of time e.g. due to vacation. In addition to the usual collection of documents that do not require the user's attention e.g. circulars and spam email, the inbox will likely include a number of documents making up distinct threads of legitimate email. When processing a large inbox, it can become difficult to unravel a thread and in particular to determine which email document of the sequence is the latest and which therefore includes all the information required by the user to formulate a response.

It would be desirable to devise an email system which addresses one or more of these problems.

Summary of the Invention

According to a first aspect of the invention, there is provided a method for managing documents in an email communication network comprising a plurality of data processing systems, the method comprising: generating an email document at a first data processing system and assigning thereto a unique document ID; transferring the first email document to a second of the data processing systems; generating at the second data processing system a second email document incorporating the received email document and assigning to the second email document another unique document ID and retaining the unique document ID of the received email document; and transferring the second email document including both unique document IDs to one of the data processing systems.

Thus, an email document that incorporates the content of a number of other email documents has associated therewith its own unique identifier as well as the unique identifiers of the incorporated documents. This differs from the known prior art in which a single unique ID is assigned to each message in an email communication system and the IDs of incorporated documents are not retained.

This new ID generation technique is advantageously employed to identify and delete redundant copies of documents in an email system and in a preferred method comprises: storing transmitted and received email documents and associated unique IDs at a data processing system; receiving a document at a data processing system and determining the one or more unique IDs associated with the document; comparing the determined one or more unique IDs with the stored IDs; identifying each stored document, all

of whose ID(s) match some or all of the ID(s) of the received document;
and deleting the stored document having the matched ID(s).

5 Additionally or alternatively, the claimed ID generation technique may be
employed in managing an email inbox in a data processing system and in a
preferred method comprises: determining the unique IDs associated with
each document in an inbox of the data processing system; and on user
selection of one of the documents in the inbox, determining whether the
inbox contains a document whose associated IDs include all the IDs of the
10 selected document; and on such a determination notifying the user.

According to a second aspect of the invention there is provided an email
communication system comprising: means for receiving an email document
having assigned thereto a unique document ID; means for generating a
15 response email document incorporating the received email document and
assigning to the second email document another unique document ID and
retaining the unique document ID of the received email document; and means
for transferring the response email document to another email
communication system.

20 According to a third aspect there is provided a method for managing
documents in an email communication network comprising a plurality of data
processing systems, the method comprising: receiving at an email system an
email document having assigned thereto a unique document ID; generating a
25 second email document incorporating the received email document for
transmission to another of the data processing systems; assigning to the
second email document another unique document ID in addition to the unique
document ID of the received email document.

30 Other aspects and preferred features are defined in the appended claims.

Brief Description of the Drawings

35 Preferred embodiments of the invention will now be described, by way of
example only, with reference to the accompanying drawings in which:

Figure 1 is a block diagram of an email communication network including
email systems connected to email servers;

40 Figure 2 is a block diagram of one of the email systems of Figure 1
according to an embodiment of the invention;

Figures 3A and 3B are flow charts showing the steps involved in the generation of email documents according to an embodiment of the present invention;

Figure 4 is a flow chart showing an email document management process according to an embodiment of the present invention;

Figure 5 is a schematic illustration of an example sequence of email communications between email systems according to an embodiment of the present invention;

Figure 6 is a schematic illustration of an email inbox and document ID storage according to an embodiment of the present invention; and

Figures 7A and 7B are flow charts showing an inbox management technique according to an embodiment of the present invention.

Detailed Description of the Preferred Embodiment

Figure 1 illustrates a network environment 10 made up of a first network 12 of email systems 14A, 14B, 14C, 14D and 14E interconnected via email server 16, and a second network 18 made up of email systems 20A, 20B, 20C and 20D interconnected via email server 22. The email servers 16 and 22 communicate via the internet 24. If, for example, the user of email system 14A wishes to send an email to the email system 20B, the user generates the email including the email address of the recipient. The email is sent to email system 20B via email server 16, internet 24 and the email server 22. Copies of emails transmitted and received by an email system are commonly held in local storage associated with the email system and may also be held in storage associated with the email server. The communication protocols and processes by which this communication is achieved are well known.

Figure 2 is a block diagram of one of the email systems of Figure 1 illustrating the components relevant to an understanding of the present invention. The email system of Figure 2 includes a hardware platform (processor, memory, storage etc.), an operating system such as Linux or Windows or similar and an email application running on the operating system. The operating system defines a network interface 38 over which the email system communicates with an email server. The email system may take the form of a personal computer system including separate display, keyboard, mouse etc., in which case the operating system may be Windows XP

and the email application may be provided by Lotus Notes which is commercially available from IBM Corporation. Any other system that provides email communication capability (cable or wireless) may also be used e.g. cellphone, personal digital assistant or the like.

In Figure 2, the email application includes an outbound document processing component 40, a database 46 and an inbound document processing component 49. The database 46 comprises a number of folders including an inbox folder 47 and a document ID database 48 in which are stored the IDs of all documents stored in the system in a manner which will be described below.

It will be appreciated that the division into inbound and outbound document processing components is done for convenience only, as an aid to understanding the invention. There may be no such division in an actual email application.

Outbound document processing component 40 includes a document creator 42 by means of which a system user creates email documents, and a unique ID generator 44 by means of which the system creates a unique ID for each document. Once created, an outbound email document is stored in the database 46. The inbound document processing component 49 is made up of an ID comparator 50, a document comparator 52 and a document deleter 54, the functions of which will be described below. An incoming document is placed into an inbox folder defined by the email application and stored in the database.

Next will be described, with reference to Figures 3A and 3B, the steps involved in generating email documents and assigning a unique ID to each document.

The process starts at 100, where, for example, a user on email system 14A opens the email inbox of the email application. The user creates 110 an email document e.g. by selecting a 'create new doc' option of the document creator component 42, typing in the text of the email document and adding the email address of the recipient e.g. at email system 20B. When the document is complete and ready to send, the ID generator 44 generates 112 a unique ID for the completed document. The document ID may take any suitable form that makes it unique but advantageously may be based on the identity of the originating email system and a next-in-sequence document ID for that system. For convenience, the unique ID is represented in the following description and Figures 3A and 3B as ID1. The document ID1

(hereafter referred to as DOC ID1 and comprising the content of document and the unique ID, ID1) is then sent at step 114 to the connected email server 16 and a copy of DOC ID1 is stored at step 116 in the email database of the email system. The document ID database is updated 118 by adding an entry for ID1, linked to the copy of DOC ID1. The document creation process ends at 120.

Figure 3B shows the steps involved in a user at recipient email system 20B creating and sending an email document that incorporates DOC ID1. The process starts at step 130 where, for example, the user opens the email system inbox. At step 132, DOC ID1 is received into the inbox and at step 134 is opened by the user who decides to send a document incorporating DOC ID1 to a number of addressees including the user of email system 14A. At step 136, the user creates a response including DOC ID1 by, for example, selecting a 'forward document' option in the email application, adding text and specifying the email addresses of the recipients. When the response document is ready to send, the ID generator generates 138 a unique ID - ID2 - for the response document which again may be based on a combination of email system ID and next-in-sequence document ID for that email system. In total therefore, the response comprises the content of DOC ID1, the material added by the user to create the response PLUS the document IDs ID1 and ID2. For convenience, the response document is referred to in the following description as DOC ID2,ID1. At step 140, on selection by the user of a 'send document' option, DOC ID2,ID1 is transferred to email server 22 for transmission to the addressees including email system 14A. At step 142, DOC ID2,ID1 is stored on the email system 20B and at step 144, the ID database is updated at step 144 by adding an entry for ID2,ID1 linked to the copy of DOC ID2,ID1. The process ends at step 146.

Figure 4 shows the steps involved in receiving and processing DOC ID2,ID1 in accordance with an embodiment of the present invention. The process starts at 150 where for example the user of email system 14A starts the email application and/or opens the inbox folder. At step 152, DOC ID2,ID1, sent from email system 20B, is received into the inbox folder and stored 154 on the email system. At step 156, the ID comparator component 50 reads the IDs associated with the received document i.e. ID2 and ID1, and compares them 158 with the IDs stored on ID database 48. On a determination at step 160 that the database contains an entry for ID1 - thereby indicating the existence of DOC ID1 on email system 14A - the process moves to step 162. If, at step 160, the database is determined not to contain an entry for ID1 which would be the case where DOC ID1 had

previously been deleted from email system 14A, the process ends at step 168.

5 When a match is made, a document integrity check is next optionally carried out in order to determine whether the content of the ID1 portion of the received document matches the content of DOC ID1. This is designed to address the possible situation where the sender of DOC ID2, ID1 has removed some of the content of DOC ID1; in which case it would not be desirable to delete the copy of DOC ID1 from email system 14A. At step 162
10 therefore, document comparator 52 (see Figure 2) retrieves the copy of DOC ID1 from storage and selects the portion corresponding to DOC ID1 from DOC ID2, ID1. A comparison between the two is made at step 164. There are various ways in which such a comparison may be made. One suitable technique is described in US Patent Application Publication US2003/0006919
15 where a search is made in the later email of a thread for characters or keywords such as FW: or RE: which indicate a forwarded or responded-to portion and then selecting the text in that portion for the comparison. Alternative integrity checks are also possible.

20 If a determination is made at step 164 that there is a match, the process moves on to step 166 where DOC ID1 and the ID1 database entry are deleted from storage at email system 14A. The process then ends at 168. If however a determination is made that there is no match then DOC ID1 is not deleted, the ID entry is kept, and the process ends.

25 In the above, the document identification, comparison and deletion processes have been described as taking place on-the-fly when the document is received. This has the advantage that the document storage is constantly optimized. In an alternative arrangement, the document is
30 received into the inbox and read, responded to or moved to another folder in the conventional manner. Steps 156 to 166 of Figure 4 are then integrated into a later process in which all the emails stored in the email system are processed to find and delete duplicate documents.

35 Thus have been described the steps involved in two email systems exchanging an email document and a single response document. The same principles are employed in more complex exchanges between multiple email systems and involving multiple threads of email documents. This may be understood with reference to Figure 5 which shows in schematic form the
40 exchange of email documents amongst three email systems A, B and C. In Figure 5, email system A sends DOC ID1 (identified as '1') to system B. System B in turn sends DOC ID2, ID1 ('2,1') to systems A and C. System C

then sends DOC ID3, ID2, ID1 ('3,2,1') to system A which then sends DOC ID4, ID3, ID2, ID1 ('4,3,2,1') to system B. It can be seen in Figure 1 that at the end of this exchange, there are four documents stored on system A, three documents on system B and two documents on system C. When the process of Figure 4 is applied to email system A (either on-the-fly or at a later stage), only DOC ID4, ID3, ID2, ID1 is retained as it includes all of the other documents. In email system B, the same document is retained and in system C, only DOC ID3, ID2, ID1 is retained. The storage required to retain the entire history of a thread received at a particular system is thereby reduced.

The use of unique IDs in the manner described above may also be employed to improve the identification and management of document threads in an email inbox folder. With reference to Figure 6, there is shown a representation of an email inbox 200 including a number of email documents 201 to 207. For the sake of simplicity only seven documents are shown in Figure 6. In real systems, the number may be in the tens or even hundreds especially when the inbox has not been accessed for a long period of time e.g. on returning from vacation. Associated with the inbox folder is an inbox document ID list 250 which is maintained by the email system. This inbox list may be separate to or part of the overall document ID database 48.

Figure 7A shows the steps involved in receiving a document into the inbox folder 200. The process starts at step 300 when, for example, the user opens the email application and connects to the email server to receive email documents. At step 302, a document having a unique ID, generated according to the process of Figure 3A for example, is received into the inbox. At step 304, the unique Doc ID is read and an entry linked to the document is added to the Inbox List 250. In Figure 6, a number of email documents have been processed in this manner to create entries in the Inbox List. The first processed document is DOC ID1 and the last processed is DOC ID7.

The inbox owner will typically process the inbox documents starting with the earliest. In conventional systems, the user on opening the first document may decide to immediately respond to the document, either without looking for other inbox documents that form part of the same thread or else methodically going through the whole inbox, opening and closing documents to find the most recent document in the thread, to which a reply is sent. This can take a good deal of time.

Figure 7B shows the steps involved in identifying documents in a thread which allows the user to more easily work with the latest document. The process starts at step 320 where, for example, the user starts the email application and downloads unread documents from the email server. The documents are received into the inbox and the document IDs are read and stored in the manner of Figure 7A. At step 322, the user opens the inbox and after scrolling through the sequence of documents, selects the earliest document, in this case DOC ID1. At steps 324 and 326, the entries in the Inbox listing are examined to determine whether any entries relate to documents incorporating DOC ID1. In Figure 6 it can be seen that there are three such entries, namely ID3,ID1; ID5,ID1 and ID6,ID3,ID1. At step 328 the user is notified of the existence of these documents in the inbox and is presented with the option to select either any one of the documents or the latest, which in this example is DOC ID6,ID3,ID1. The user may then open 330 the selected document. The process ends at step 330. On a determination at step 326 that there are no entries in the inbox listing relate to documents incorporating DOC ID1 then the process ends at step 332. The process of Figure 7B is repeated for each document in the inbox.

CLAIMS

1. A method for managing documents in an email communication network comprising a plurality of data processing systems, the method comprising:

generating an email document at a first data processing system and assigning thereto a unique document ID;

transferring the first email document to a second of the data processing systems;

generating at the second data processing system a second email document incorporating the received email document and assigning to the second email document another unique document ID and retaining the unique document ID of the received email document; and

transferring the second email document including both unique document IDs to one of the data processing systems.

2. The method of claim 1, further comprising:

storing transmitted and received email documents and associated unique IDs at a data processing system;

receiving a document at a data processing system and determining the one or more unique IDs associated with the document;

comparing the determined one or more unique IDs with the stored IDs;

identifying each stored document, all of whose ID(s) match some or all of the ID(s) of the received document; and

deleting the stored document having the matched ID(s).

3. The method of claim 2, further comprising:

prior to deleting the document having the matched ID(s), verifying that the content of the received document matches the content of the portion of the stored document corresponding to the matched IDs.

4. The method of claim 1, further comprising:

determining the unique IDs associated with each document in an inbox of the data processing system;

on user selection of one of the documents in the inbox, determining whether the inbox contains another document whose associated IDs include all the IDs of the selected document; and on such a determination notifying the user.

5. A method for managing documents in an email communication network comprising a plurality of data processing systems, the method comprising:

receiving at an email system an email document having assigned thereto a unique document ID;

generating a second email document incorporating the received email document for transmission to another of the data processing systems;

assigning to the second email document another unique document ID in addition to the unique document ID of the received email document.

6. An email communication system comprising:

means for receiving an email document having assigned thereto a unique document ID;

means for generating a response email document incorporating the received email document and assigning to the second email document another unique document ID and retaining the unique document ID of the received email document; and

means for transferring the response email document to another email communication system.

7. The email communication system of claim 6, further comprising:

means for storing transmitted and received email documents and associated unique IDs;

means for receiving an email document and determining the one or more unique IDs associated with the document;

means for comparing the determined one or more unique IDs with the stored IDs;

means for identifying each stored document, all of whose ID(s) match some or all of the ID(s) of the received document; and

means for deleting the stored document having the matched ID(s).

8. The email communication system of claim 6, further comprising:

means for determining the unique IDs associated with each document in an inbox of the data processing system;

means responsive to user selection of one of the documents in the inbox, for determining whether the inbox contains a document whose associated IDs include all the IDs of the selected document; and

means for notifying the user on such a determination.

9. A computer program product loadable into the internal memory of a digital computer, comprising software code portions for performing, when said product is run on a computer, the steps of any of claims 1 to 5.

1/7

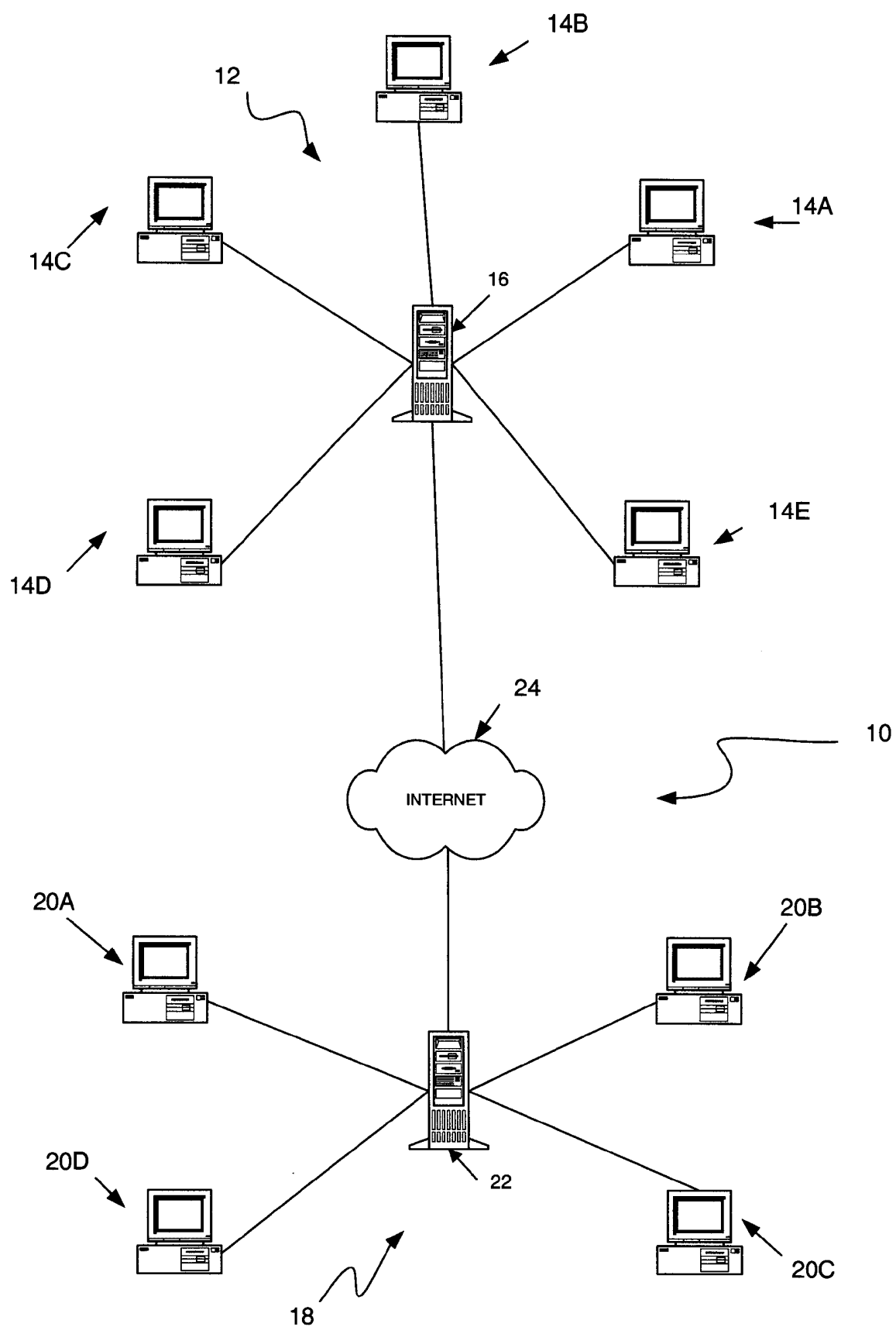


FIGURE 1

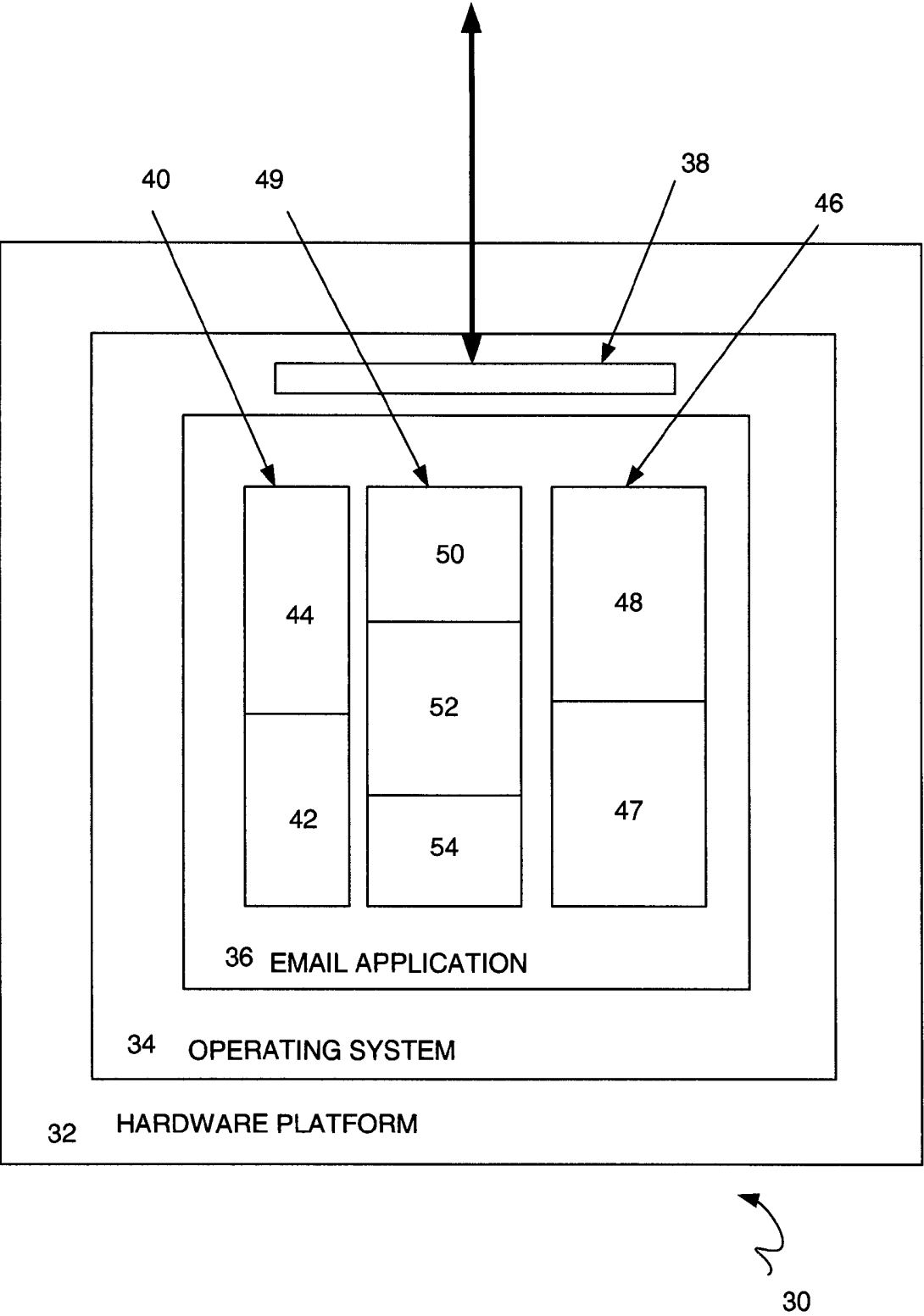


FIGURE 2

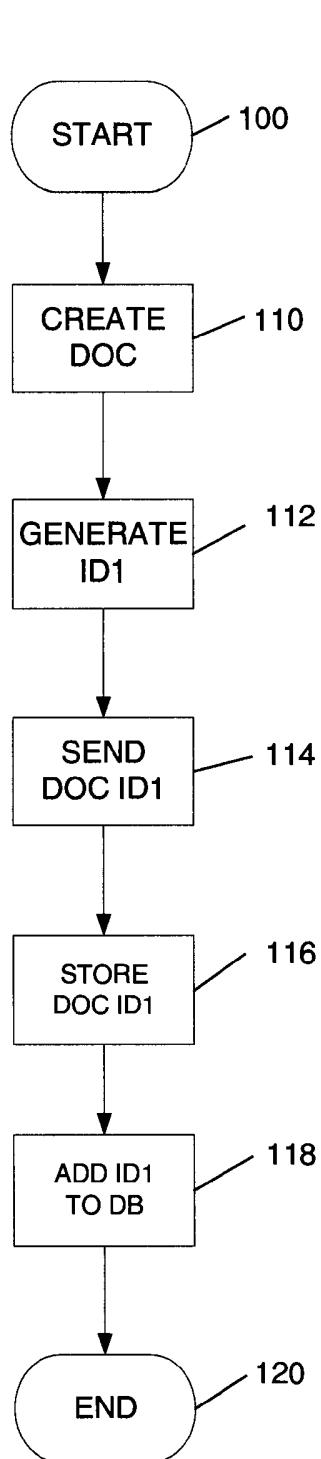


FIG 3A

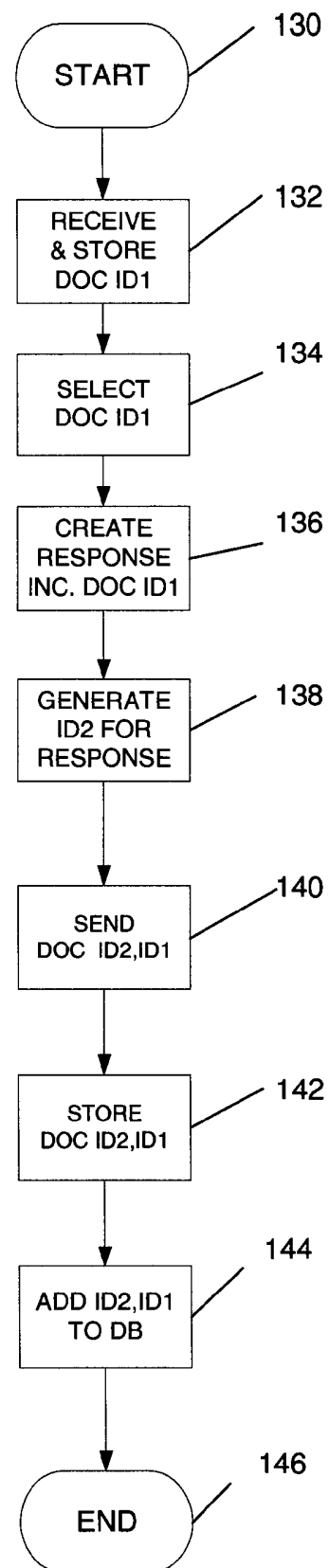


FIG 3B

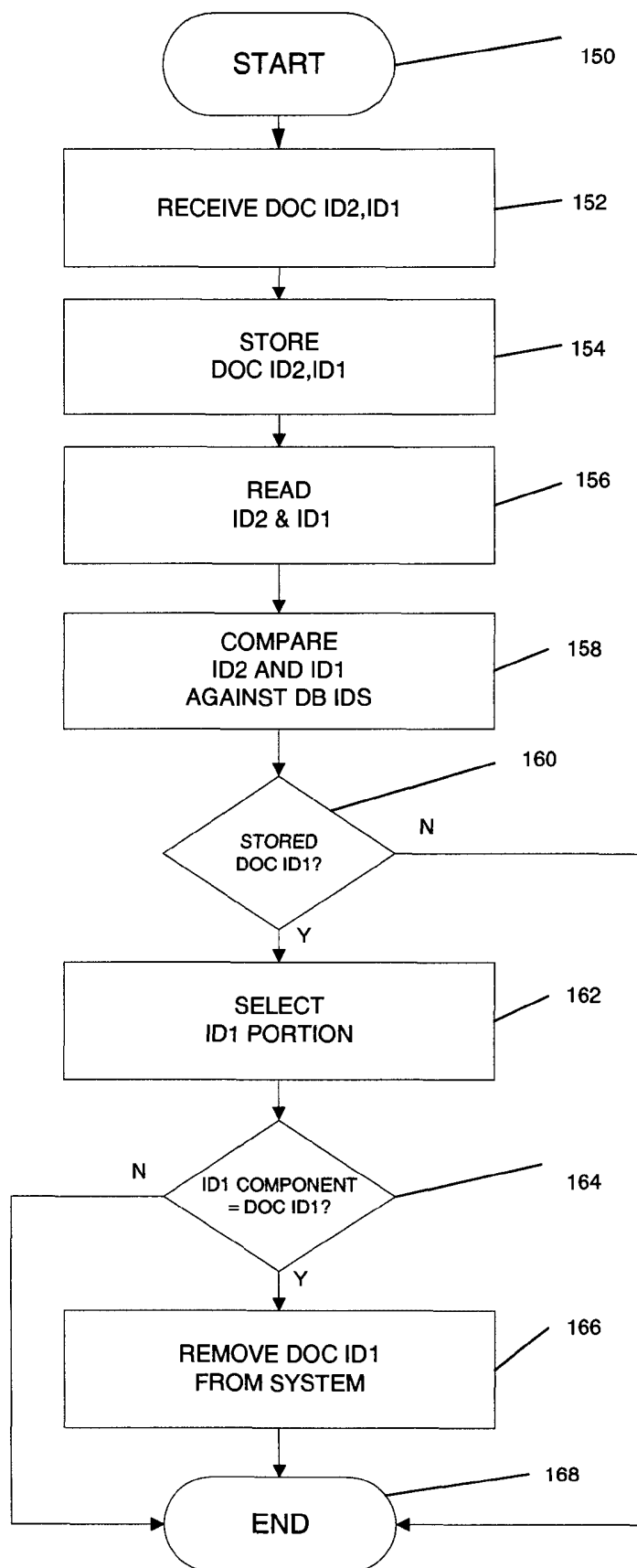


FIGURE 4

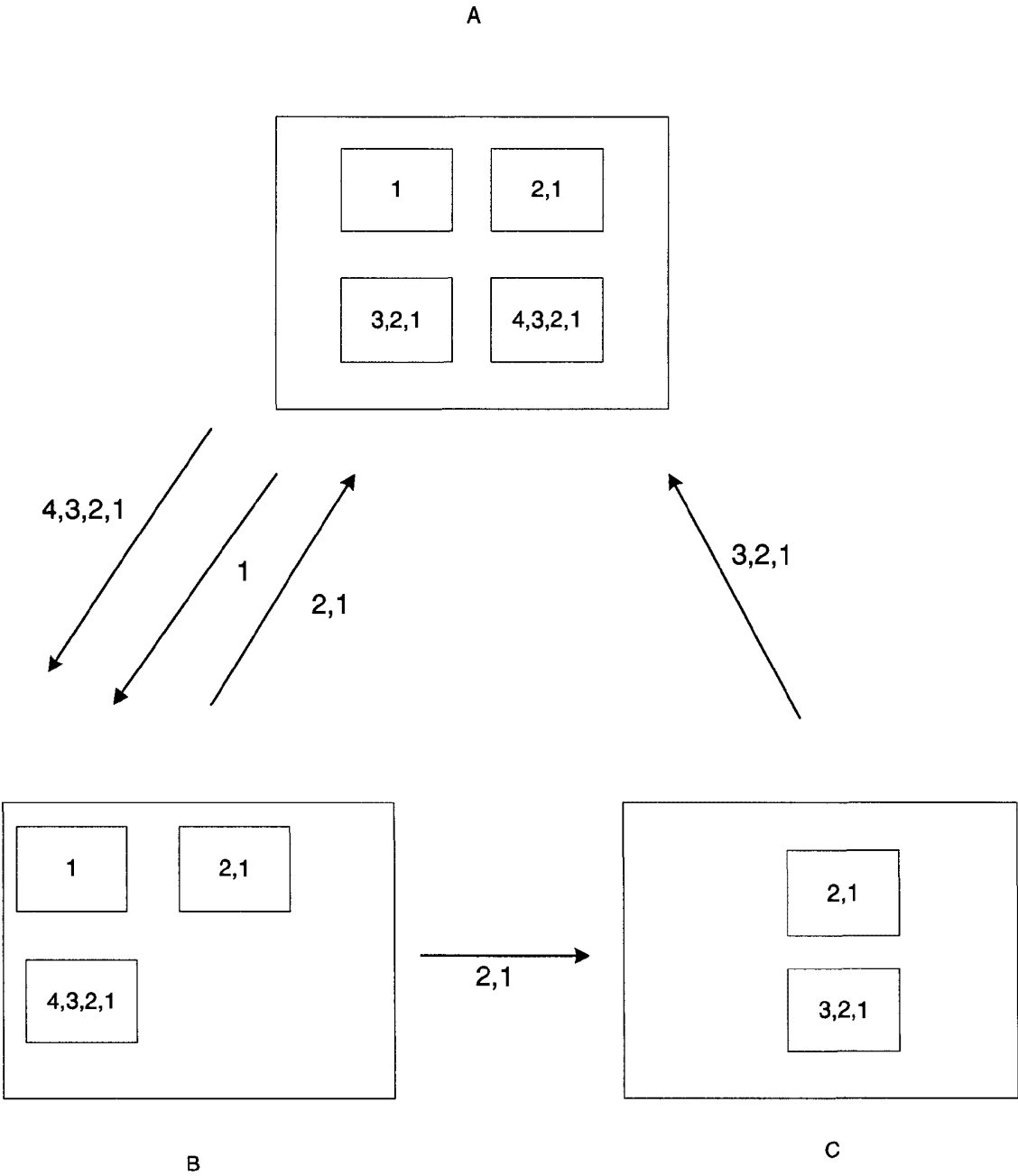


FIGURE 5

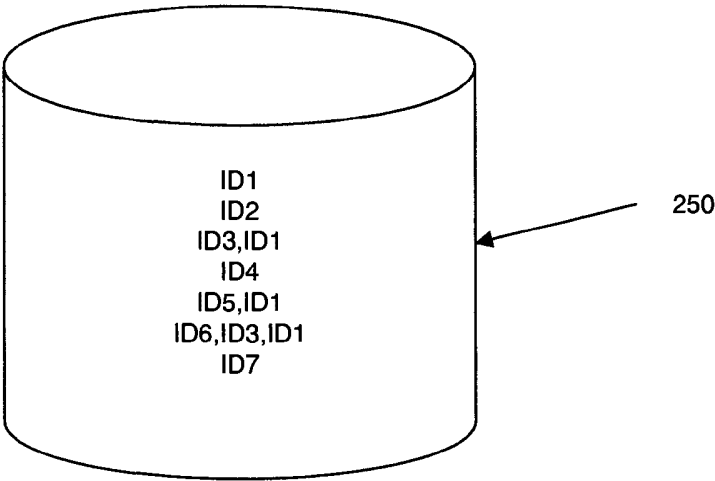
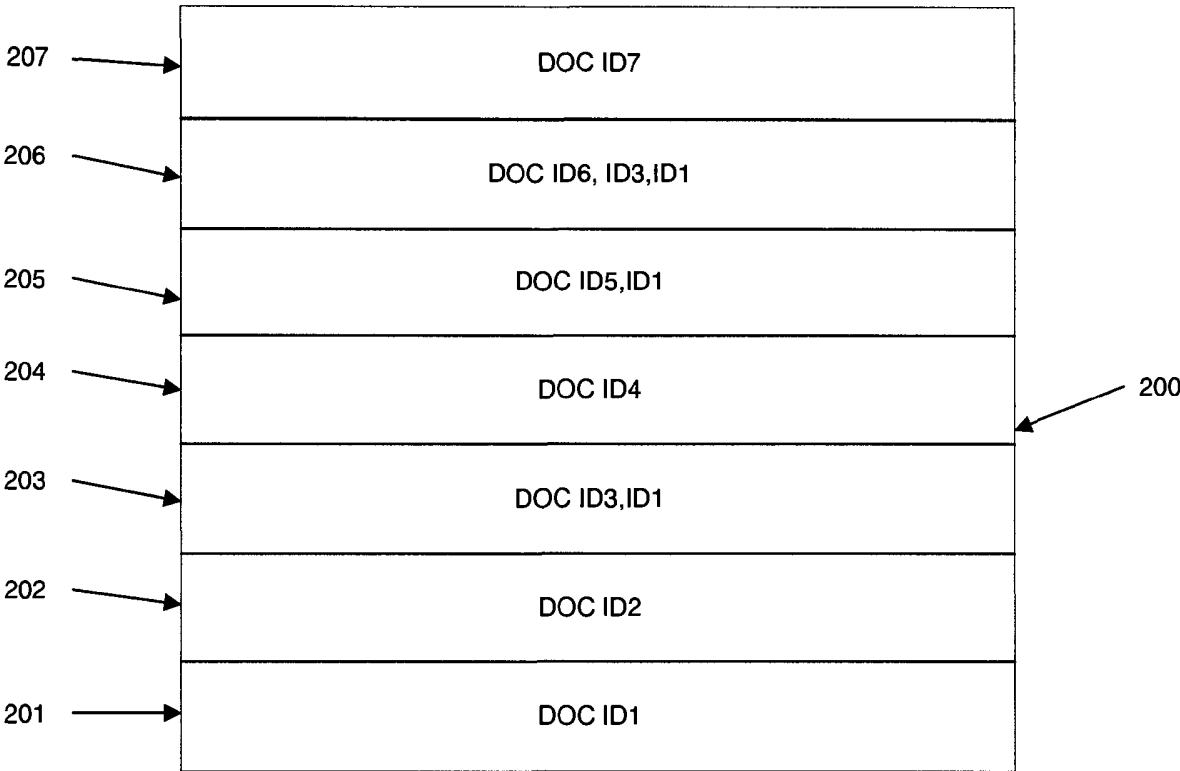


FIGURE 6

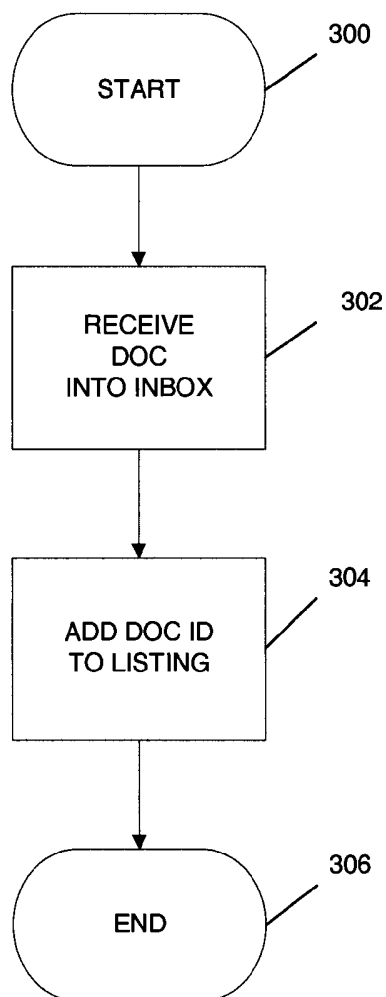


FIGURE 7A

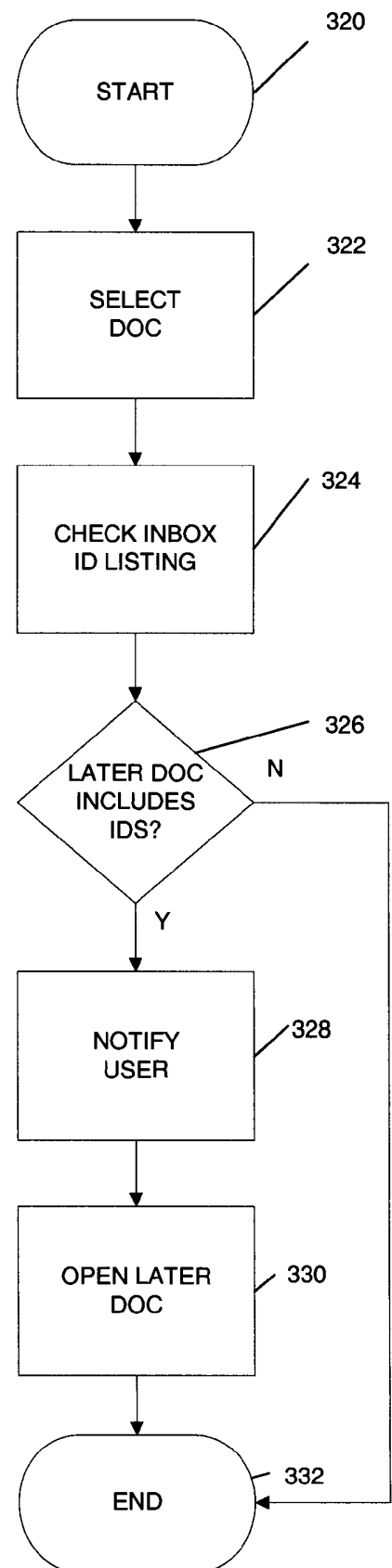


FIGURE 7B

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2006/050043

A. CLASSIFICATION OF SUBJECT MATTER

H04L12/58 G06Q10/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EP0-Internal, WPI Data, PAJ, INSPEC, COMPENDEX, IBM-TDB

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 6 631 398 B1 (KLEIN DEAN A) 7 October 2003 (2003-10-07) column 1 - column 11, line 7; figures 2-8 -----	1-9
X	US 2004/199590 A1 (KEOHANE SUSANN MARIE ET AL) 7 October 2004 (2004-10-07) paragraph [0008] paragraph [0034] - paragraph [0052]; figures 4-10 -----	1,4-6,8, 9



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *Z* document member of the same patent family

Date of the actual completion of the international search

22 March 2006

Date of mailing of the international search report

29/03/2006

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

García Bolos, R

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2006/050043

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 6631398	B1	07-10-2003	NONE
US 2004199590	A1	07-10-2004	NONE