

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
18 June 2009 (18.06.2009)

PCT

(10) International Publication Number
WO 2009/076010 A1

- (51) International Patent Classification:
G06F 17/24 (2006.01) G06F 17/21 (2006.01)
- (21) International Application Number:
PCT/US2008/083862
- (22) International Filing Date:
18 November 2008 (18.11.2008)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
11/951,973 6 December 2007 (06.12.2007) US
- (71) Applicant (for all designated States except US): MICROSOFT CORPORATION [US/US]; One Microsoft Way, Redmond, Washington 98052-6399 (US).
- (72) Inventors: BAILOR, Jonathan Beckett; c/o Microsoft Corporation, One Microsoft Way, Redmond, Washington 98052-6399 (US). BERNSTEIN, Ethan Joseph; c/o Microsoft Corporation, One Microsoft Way, Redmond, Washington 98052-6399 (US). KNIGHT, Mark Rolland; c/o Microsoft Corporation, One Microsoft Way, Redmond,

Washington 98052-6399 (US). ANTOS, Christopher James; c/o Microsoft Corporation, One Microsoft Way, Redmond, Washington 98052-6399 (US).

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

[Continued on next page]

(54) Title: DOCUMENT MERGE

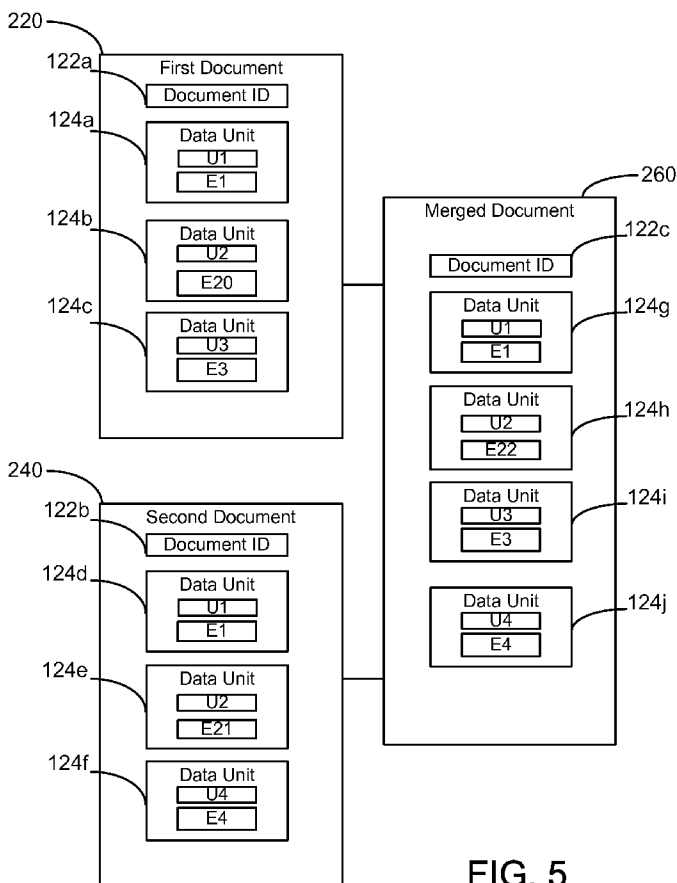


FIG. 5

(57) Abstract: Methods and systems of merging a first document with at least a second document to generate a third document. Each of the first and second documents includes at least one data unit. Each data unit is associated with a unit identifier and an edit identifier. The unit identifiers of each of the documents are compared to the unit identifiers of the other document to determine whether each unit identifier is a matching or non-matching unit identifier. Edit identifiers of the data units associated with the matching unit identifiers are compared to each other. Data units associated with the non-matching unit identifiers are inserted directly into the third document. Data units associated with the matching unit identifiers and the matching edit identifiers are inserted directly into the third document. Data units associated with the matching unit identifiers and the non-matching edit identifiers are merged into the third document.

WO 2009/076010 A1



NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG,
CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

— *as to the applicant's entitlement to claim the priority of the
earlier application (Rule 4.17(iii))*

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a
patent (Rule 4.17(ii))*

Published:

— *with international search report*

DOCUMENT MERGE

BACKGROUND

[0001] Merging is an act of reconciling multiple changes made to different copies of a same document. Most often, it is necessary when a document is modified by two
5 people at the same time. Later, these changes need to be merged into a single new document reflecting the intent of both authors.

[0002] A conventional merge can be cumbersome and time-consuming. A more efficient merge is desired.

SUMMARY

10 [0003] As discussed herein, an aspect of some embodiments relates to a method of merging a first document with at least a second document to generate a third document. Each of the first and second documents includes at least one data unit. Each data unit is associated with a unit identifier and an edit identifier. The method includes obtaining the unit identifiers of the data units of the first document. The method also includes obtaining
15 the unit identifiers of the data units of the second document. The unit identifiers of each of the first and second documents are compared to the unit identifiers of the other of the first and second documents to determine whether each of the unit identifiers is a matching unit identifier or a non-matching unit identifier. The matching unit identifiers indicate corresponding data units of the first and second documents. Each of the non-matching
20 unit identifiers indicates a data unit of one of the first and second documents that does not correspond with any data unit of the other of the first and second documents. Edit identifiers of the data units associated with the matching unit identifiers are compared to each other to ascertain matching edit identifiers. Each of the matching edit identifiers indicates the data unit of one of the first and second documents matches the corresponding
25 data unit of the other of the first and second documents. Data units of the first and second documents associated with non-matching unit identifiers are inserted directly into the third document. In addition, data units of the first and second documents associated with matching unit identifiers and matching edit identifiers are inserted directly into the third

document. Finally, data units of the first and second documents associated with matching unit identifiers and non-matching edit identifiers are merged into the third document.

[0004] Another aspect of some embodiments relates to a document processing system. The document processing system includes a document processing application and a first document. The document processing application is configured to merge together at least two documents to generate a third document. The first document is configured to be processed by the document processing application. The first document stores content and metadata. The content includes at least one data unit. The metadata includes a unit identifier and an edit identifier associated with each data unit. The unit identifier of each data unit is generated when the data unit is created. The edit identifier of each data unit is generated when a modification to the data unit is saved.

[0005] Further aspects relate to a computer-readable storage medium having computer-executable instructions for implementing a method of generating a merged document from a first document and a second document. The method includes determining whether the first document has a first document identifier and whether the second document has a second document identifier. The method also includes obtaining one of the first and second document identifiers if one of the first and second documents has one of the first and second document identifiers. The method further includes comparing the first document identifier of the first document to the second document identifier of the second document. The methods includes performing a conventional merge of the first and second documents to generate the merged document if any of the first and second documents does not have a document identifier. Moreover, the method includes performing a conventional merge of the first and second documents to generate the merged document if the first document identifier differs from the second document identifier. Finally, the method includes performing an accelerated merge of the first and second documents to generate the merged document if the first document identifier matches the second document identifier.

[0006] This summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This Summary is not intended to identify key or essential features of the claimed subject matter, nor is it intended to be used in any way as to limit the scope of the claimed subject matter.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] FIG. 1 is a block diagram of an exemplary document processing system;

[0008] FIG. 2 is a block diagram illustrating an exemplary document configured to be processed by an exemplary document processing application of FIG. 1;

5 [0009] FIG. 3 is a block diagram illustrating an exemplary format for the document shown in FIG. 2;

[0010] FIG. 4 is a block diagram illustrating another exemplary document similar to that shown in FIG. 3, except that a metadata file is stored separately from the document;

[0011] FIG. 5 is a block diagram illustrating merging a first document with a
10 second document to generate a third document;

[0012] FIG. 6 is an exemplary computing system for implementing aspects of the present disclosure;

[0013] FIG. 7 illustrates an example method of merging a first document with a second document to generate a third document;

15 [0014] FIG. 8 illustrates an example method of assigning and storing a unit identifier and an edit identifier to each data unit of a document; and

[0015] FIG. 9 illustrates another example method of merging a first document with a second document to generate a merged document.

DETAILED DESCRIPTION

20 [0016] This disclosure will now more fully describe exemplary embodiments with reference to the accompanying drawings, in which specific embodiments are shown. Other aspects may, however, be embodied many different forms and the inclusion of specific embodiments in the disclosure should not be construed as limiting such aspects to the embodiments set forth herein. Rather, the embodiments depicted in the drawings are
25 included to provide a disclosure that is thorough and complete and which fully conveys the intended scope to those skilled in the art. When referring to the figures, like structures and elements shown throughout are indicated with like reference numerals.

[0017] Embodiments of the present disclosure generally relate to a document merge. In one non-limiting example, merging a first document with at least a second
30 document to generate a third document is described. Each of the first and second

documents includes at least one data unit. Each data unit is associated with a unit identifier and an edit identifier. Further embodiments relate to a document processing system. The document processing system includes a document processing application which is configured to merge together at least two documents to generate a third
5 document.

[0018] FIG. 1 is a block diagram of an exemplary document processing system. The document processing system 100 includes a document processing application 110 and documents 120. The document processing application 110 is configured to merge together the documents 120 to generate a merged document. In alternative embodiments, the
10 document processing application 110 is configured to merge together three, four or more documents to generate a merged document.

[0019] In one possible embodiment, the document processing application 110 is a word processing application such as, but not limited to, a MICROSOFT® Office Word application. In other possible embodiments, the document processing system 100 is an
15 image, data, voice, or any other suitable document processing system. Similarly, the document processing application 110 can also be an image, data, voice, or any other suitable document processing application such as, but not limited to, a Microsoft® Office POWERPOINT®, EXCEL®, VISIO®, ACCESS™, and OUTLOOK® software application. The documents 120 can be text or rich text documents, presentations,
20 spreadsheets, drawings, data, emails, and any other documents in a suitable format.

[0020] FIG. 2 is a block diagram illustrating an exemplary document 120 which is configured to be processed by an exemplary document processing application 110 in FIG. 1. The document 120 is divided into one or more data units 124. Each data unit 124 is assigned a unique unit identifier 126 that persists for the life of the data unit 124. The unit
25 identifier 126 identifies and distinguishes the associated data unit 124 from any other data units in the document 120.

[0021] The data unit 124 is also assigned an edit identifier 128 when the data unit 124 is created, edited or modified. In particular, whenever the data unit 124 is created, edited or modified, a new edit identifier 128 is assigned to the data unit 124 so that the
30 new edit identifier 128 replaces the previous edit identifier (if any) of the data unit 124.

The new edit identifier 128 is different from the previous edit identifier (if any) of the data unit 124. In one possible embodiment, the edit identifier 128 is randomly generated.

[0022] In example embodiments, the document 120 is broken into data units 124 by using a variety of methods. For example, if the document 120 is a text document, each paragraph of the document can be broken into a separate data unit. In other words, in the example of a text document, the data unit 124 can also be called a paragraph. In such an example, the unit identifiers 126 are referred to as paragraph identifiers. In other types of documents, other methods can be used to break the document 120 into data units. For example, objects, graphics, slides, or other selectable items within a document can be used to break the document into separate data units.

[0023] The unit identifier 126 and the edit identifier 128 can be in any suitable format to identify the data unit 124. One possible exemplary format of the unit identifier 126 and the edit identifier 128 is a combination of numbers, letters, and/or symbols, such as a 32-bit integer. A person skilled in the art will appreciate many other suitable identifier formats for the unit identifier 126 and the edit identifier 128. The unit identifier 126 and the edit identifier 128 can be stored in the document 120, in a separate file or in any other suitable storable and accessible medium. Some exemplary formats of storing the unit identifier 126 and the edit identifier 128 are discussed below in more detail in FIGs. 3-4.

[0024] The document 120 also includes a document identifier 122 to identify the document 120. In one possible embodiment, the document identifier 122 is assigned to the document 120 when the document 120 is newly created. In another possible embodiment, the document identifier 122 is assigned to the document 120 when the document 120 has no document identifier. For example, a document identifier 122 is assigned to the document 120 when the document 120 is previously edited by an application that does not recognize unit or edit identifiers.

[0025] The document identifier 122 is unique and persistent when the document identifier 122 is assigned to the document 120. In other words, the document identifier 122 persists for the life of the document 120. Even when the document 120 is edited and modified, the document identifier 122 does not change. In addition, when a separate document is created from the document 120, the document identifier 122 stays with the

separate document. This can occur when a user copies a document to a new document, or when multiple user copies of the document are created to allow for multiple users to edit the same document.

[0026] FIG. 3 is a block diagram illustrating an exemplary format for the document 120 shown in FIG. 2. In particular, the document 120 includes a content 130 and a metadata 140. The document 120 stores the content 130 and the metadata 140. The content 130 includes at least one data unit 124. The metadata 140 includes a unit identifier 126 and an edit identifier 128 which are associated with each data unit 124. The unit identifier 126 is generated when the data unit 124 is created. The edit identifier 128 is generated whenever a modification to the data unit 124 is stored within the document. The metadata 140 is stored in the document 120.

[0027] FIG. 4 is a block diagram illustrating another exemplary document 120 similar to that shown in FIG. 3, except that a metadata file 150 is stored separately from the document 120. Preferably, the metadata file 150 includes a metadata 140 which pertains to content 130 of the document 120. The metadata 140 of the metadata file 150 includes a unit identifier 126 which indicates to which data unit 124 the metadata 140 pertains. In one possible embodiment, the metadata 140 can also include an edit identifier 128. The data unit 124 is indicated by the unit identifier 126 included in the metadata 140.

[0028] FIG. 5 is a block diagram illustrating merging a first document 220 with a second document 240 to generate a third document 260. Exemplary structures of the first document 220 and the second document 240 are described above in FIGs 1-4. The third document 260 is a merged document created by merging the first document 220 and the second document 240.

[0029] In an example shown in FIG. 5, the first document 220 includes a first document identifier 122a and data units 124a-c. The first document identifier 122a identifies the first document 220. Three data units 124a-c are shown in FIG. 5, however, the first document 220 may include more or less than three data units. As to the shown data units 124a-c, the data unit 124a has a unit identifier U1 and an edit identifier E1 which are associated with the data unit 124a. The data unit 124b has a unit identifier U2 and an edit identifier E20 associated with the data unit 124b. The data unit 124c has a unit identifier U3 and an edit identifier E3 associated with the data unit 124c.

[0030] Similarly, the second document 240 includes a second document identifier 122b and data units 124d-f. The second document identifier 122b identifies the second document 240. Three data units 124d-f are shown in FIG. 5, however, the second document 240 may include more or less than three data units. The data unit 124d has a unit identifier U1 and an edit identifier E1 associated with the data unit 124d. The data unit 124e has a unit identifier U2 and an edit identifier E21 associated with the data unit 124e. As it is noted, the edit identifier E21 of the data unit 124e in the second document 240 is different from the edit identifier E20 of the data unit 124b in the first document 220. The data unit 124f has a unit identifier U4 and an edit identifier E4 associated with the data unit 124f.

[0031] In the exemplary merge shown in FIG. 5, the first document identifier 122a of the first document 220 is same as the second document identifier 122b of the second document 240. The first document 220 and the second document 240 are thus performed in an accelerated merge as described in detail as follows. Since the first document 220 and the second document 240 have the same document identifier, the first and second documents 220, 240 share a common origin or parent. As such, the first document 220 can be merged with the second document 240 using one or more of the accelerated merging processes described below.

The data unit 124a of the first document 220 has a same unit identifier U1 and edit identifier E1 as the data unit 124d of the second document 240. The data unit 124a therefore is indicated as the same as the data unit 124d. In other words, the data unit 124d has not been changed or modified from the data unit 124a. Accordingly, no merge is needed between the data unit 124a and the data unit 124d. Rather, the data unit 124a (or the data unit 124d which is same as the data unit 124a) is directly inserted into the third document 260 as noted as a data unit 124g. Since there is no merge necessary between the data unit 124a and the data unit 124d, the data unit 124g in the merged document (i.e., the third document) 260 keeps the unit identifier U1 and the edit identifier E1 associated with it. This direct insertion without further comparing the contents in the data unit 124a with the contents in the data unit 124d can save time, improve efficiency, and enhance merge precision.

[0032] The data unit 124b and the data unit 124e have a same unit identifier U2, but have different edit identifiers E20, E21. Different edit identifiers E20, E21 indicate that the data unit 124b and the data unit 124e may be different. In other words, a merge between the data unit 124b and the data unit 124e is needed. Therefore, a data unit 124h
5 in the third document 260 is generated from a merge between the data unit 124b and the data unit 124e. The unit identifier U2 is kept in the data unit 124h. A new edit identifier E22 is created in the merge of the two data units 124b and 124e.

[0033] The data unit 124c of the first document 220 cannot be found to have any data unit in the second document 240 which has a matching unit identifier with the unit
10 identifier U3 of the data unit 124c. Thus, the data unit 124c is directly inserted into the third document 260 without a merge and is noted as a data unit 124i in the third document 260. The data unit 124i thus includes the unit identifier U3 and the edit identifier E3 associated with the data unit 124i. The data unit 124i in the third document 260 is exactly same as the data unit 124c in the first document 220. In one possible embodiment, the
15 data unit 124c is directly attached into the end of the third document 260. In another possible embodiment, the data unit 124c (without a merge of contents of the data unit 124c) is sorted to be inserted into a proper location of the third document 260 according to relationships among the data units 124a-j.

[0034] Similarly, the data unit 124f of the second document cannot be found to
20 have any data unit in the first document 220 which has a matching unit identifier with a unit identifier U4 of the data unit 124f. Therefore, the data unit 124f is directly inserted into the third document 260 without a merge and is noted as a data unit 124j in the third document 260. The data unit 124j thus includes the unit identifier U4 and the edit identifier E4 associated with the data unit 124j. The data unit 124j in the third document
25 260 is exactly same as the data unit 124f in the second document 240. In one possible embodiment, the data unit 124f is directly attached into the end of the third document 260. In another possible embodiment, the data unit 124f (without a merge of contents of the data unit 124f) is sorted to be inserted into a proper location of the third document 260 according to relationships among the data units 124a-j.

[0035] The merged document 260 thus includes the data units 124g-j. Each of the
30 data units 124g-j is also associated with a unit identifier and an edit identifier. In addition,

the document identifier 122a (or 122b which is same as 122a) stays in the merged document 260 and is noted as a document identifier 122c in FIG. 5.

[0036] In another possible embodiment, if the document identifier 122a of the first document 220 is different from the document identifier 122b of the second document 240, a conventional merge is performed between the first document and the second document. In yet another possible embodiment, if any of the first and second documents 220, 240 does not have a document identifier, a conventional merge will be performed to generate the third document.

[0037] The example in FIG. 5 shows merging two documents together to a merged third document. However, it is not limited to merging only two documents. Instead, the concept and principle illustrated herein can apply to merging more than two documents.

[0038] FIG. 6 is an exemplary computing system 600 for implementing aspects of the present disclosure. For example, computing system 600 can be a document processing system 100 as discussed above that is used to create one and/or merge two or more documents such as the documents 220, 240 described above. In its most basic configuration, computing system 600 typically includes at least one processing unit 602 and memory 604. Depending on the exact configuration and type of computing system, memory 604 may be volatile (such as RAM), non-volatile (such as ROM, flash memory, etc.) or some combination of the two. This most basic configuration is illustrated in FIG. 6 by dashed line 606. Additionally, computing system 600 may also have additional features/functionality. For example, computing system 600 may also include additional storage (removable and/or non-removable) including, but not limited to, magnetic or optical disks or tape. Such additional storage is illustrated in FIG. 6 by removable storage 608 and non-removable storage 610. Computer storage media includes volatile and nonvolatile, removable and non-removable media implemented in any method or technology for storage of information such as computer readable instructions, data structures, program modules or other data. Memory 604, removable storage 608 and non-removable storage 610 are all examples of computer storage media. Computer storage media includes, but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage

devices, or any other medium which can be used to store the desired information and which can be accessed by computing system 600. Any such computer storage media may be part of computing system 600.

5 [0039] Computing system 600 may also contain communications connection(s) 612 that allow the computing system to communicate with other devices. Communications connection(s) 612 is an example of communication media. Communication media typically embodies computer readable instructions, data structures, program modules or other data in a modulated data signal such as a carrier wave or other transport mechanism and includes any information delivery media. The term “modulated data signal” means a signal that has one or more of its characteristics set or changed in 10 such a manner as to encode information in the signal. By way of example, and not limitation, communication media includes wired media such as a wired network or direct-wired connection, and wireless media such as acoustic, RF, infrared and other wireless media. The term computer readable media as used herein includes both storage media and 15 communication media.

[0040] Computing system 600 may also have input device(s) 614 such as keyboard, mouse, pen, voice input device, touch input device, etc. Output device(s) 616 such as a display, speakers, printer, etc. may also be included. All these devices are well known in the art and need not be discussed at length here.

20 [0041] In some embodiments, memory 604 includes one or more of operating system 620, application programs 622, other program modules 624, and program data 626. For example, application programs 622 can include a text or word processing application that is used to create and/or merge two or more documents.

[0042] FIG. 7 illustrates an example method 700 of merging a first document with 25 a second document to generate a third document. In the method shown, each of the first and second documents includes at least one data unit. Each data unit is associated with a unit identifier and an edit identifier. A document processing application is configured to merge together the first and second documents to generate the third document. As discussed above, in one possible embodiment, the documents are text documents.

30 [0043] At operation 702, the document processing application obtains the unit identifiers of the data units of the first document. In one possible embodiment, the unit

identifiers are stored within the first document. In another possible embodiment, the unit identifiers are stored in a metadata file which is separated from the first document.

5 [0044] Operational flow proceeds to an operation 704. At the operation 704, the document processing application obtains the unit identifiers of the data units of the second document. In one possible embodiment, the unit identifiers are stored within the second document. In another possible embodiment, the unit identifiers are stored in a metadata file which is separated from the second document.

10 [0045] Operational flow proceeds to an operation 706. At the operation 706, the document processing application compares the unit identifiers of each of the first and second documents to the unit identifiers of the other of the first and second documents to determine whether each of the unit identifiers is a matching unit identifier or a non-matching unit identifier. A matching unit identifier means that a unit identifier in the first document is found to be same as a unit identifier in the second document. These matching unit identifiers indicate corresponding data units of the first and second documents. A
15 non-matching unit identifier means that a unit identifier in one of the first and second documents cannot be found to be same as any unit identifier in the other of the first and second documents. Each of the non-matching unit identifiers indicates a data unit of one of the first and second documents that does not correspond with any data unit of the other of the first and second documents.

20 [0046] Operational flow proceeds to a comparison operation 708. The comparison operation 708 determines whether or not any unit identifier in one of the first and second document matches any unit identifier in the other of the first and second document. If the comparison operation 708 determines that a unit identifier in one of the first and second document does not match any unit identifier in the other of the first and second document,
25 then operational flow branches “No” and operational flow proceeds to an operation 710. At the operation 710, the data unit associated with the non-matching unit identifier is directly inserted into the third document.

[0047] If the comparison operation 708 determines that a unit identifier in one of the first and second document matches any unit identifier in the other of the first and
30 second document, then operational flow branches “Yes” and operational flow proceeds to an operation 712.

[0048] At the operation 712, the document processing application compares edit identifiers of the data units associated with the matching unit identifiers to ascertain matching edit identifiers. Each of the matching edit identifiers indicates the data unit of one of the first and second documents matches the corresponding data unit of the other of the first and second documents. On the other hand, each of the non-matching edit identifiers indicates the data unit of one of the first and second documents does not match the corresponding data unit of the other of the first and second documents.

[0049] Operational flow proceeds to a comparison operation 714. The comparison operation 714 determines whether or not the edit identifiers of the data units associated with the matching unit identifiers match each other. If the comparison operation 708 determines that the edit identifiers match each other, then operational flow branches “Yes” and operational flow proceeds to the operation 710. At the operation 710, the data unit associated with the matching unit identifiers and the matching edit identifiers is directly inserted into the third document.

[0050] If the comparison operation 714 determines that the edit identifiers of the data units associated with the matching unit identifiers do not match each other, then operational flow branches “No” and operational flow proceeds to the operation 716. At the operation 716, the data units associated with the matching unit identifiers and the non-matching edit identifiers are merged into a data unit to the third document. This can be done, for example, using a convention merging process such as a three-way merge algorithm. Other conventional merging processes can also be used.

[0051] The example shown in FIG. 7 shows a method of merging two documents together to a merged third document. However, it is not limited to merging two documents. Instead, the concept and principle illustrated herein can apply to merging more than two documents.

[0052] FIG. 8 illustrates an example method 800 of assigning and storing a unit identifier and an edit identifier to each data unit of a document. In particular, at operation 802, a unit identifier is assigned to each data unit of the document. The unit identifier is unique and persistent. The unit identifier persists for the life of the data unit. The unit identifier identifies and distinguishes the associated data unit from any other data units in the document.

[0053] Operational flow proceeds to an operation 804. At the operation 804, the unit identifiers are stored within the document. In one possible embodiment, the unit identifiers are stored in a metadata within the document. In another possible embodiment, the unit identifiers are stored in a metadata file which is separated from the document.

5 [0054] Operational flow proceeds to an operation 806. At the operation 806, an edit identifier is assigned to each data unit of the document whenever the data unit is created or edited. Thus, the edit identifier indicates that the data unit is newly created or edited. The edit identifier is separated from the unit identifier. The newly assigned edit identifier is different from and replaces previous edit identifier (if any) of the data unit
10 when the edit identifier is assigned to the data unit. In one possible embodiment, the edit identifier is randomly generated. In other words, the edit identifier can be assigned with a random identifier.

[0055] Operational flow proceeds to an operation 808. At the operation 808, the edit identifiers are stored within the document. In one possible embodiment, the edit
15 identifiers are stored in a metadata within the document. In another possible embodiment, the edit identifiers are stored in a metadata file which is separated from the document. In yet another possible embodiment, when the user saves the document, the edit identifiers are stored therewith.

[0056] FIG. 9 illustrates another example method 900 of merging a first document
20 with a second document to generate a merged document. A document processing application is configured to merge together the first and second documents to generate the merged document. The method 900 particularly illustrates how to process the first and second documents which may not have a document identifier associated with the first and/or second documents. In other words, in one possible embodiment, the method 900
25 shows how to retroactively handle documents created and/or edited by previous versions of a document processing application. The method 900 illustrates how the document processing application handles document backward compatibility with previous application versions.

[0057] Operational flow starts at an operation 902. The operation 902 determines
30 whether the first document has a first document identifier associated with the first document. If the operation 902 determines that the first document does not have a first

document identifier associated with it, then operational flow branches “No” and operational flow proceeds to an operation 914 (to be discussed in detail below). If the operation 902 determines that the first document has a first document identifier associated with the first document, then operational flow branches “Yes” and operational flow
5 proceeds to an operation 904. At the operation 904, the first document identifier is obtained from the first document.

[0058] Operational flow proceeds to an operation 906. The operation 906 determines whether the second document has a second document identifier associated with the second document. If the operation 906 determines that the second document does not
10 have a second document identifier associated with the second document, then operational flow branches “No” and operational flow proceeds to an operation 914. If the operation 906 determines that the second document has a second document identifier associated with the second document, then operational flow branches “Yes” and operational flow proceeds to an operation 908. At the operation 908, the second document identifier is obtained from
15 the second document.

[0059] Operational flow proceeds to a comparison operation 910. The comparison operation 910 determines whether the first document identifier of the first document matches the second document identifier of the second document. If the comparison operation 910 determines that the first document identifier of the first document matches
20 the second document identifier of the second document, then operational flow branches “Yes” and operational flow proceeds to an operation 912. At the operation 912, an accelerated document merge is performed as discussed above in FIGs. 7-8. If the comparison operation 910 determines that the first document identifier of the first document does not match the second document identifier of the second document, then
25 operational flow branches “No” and operational flow proceeds to an operation 914. At the operation 914, a conventional document merge is performed.

[0060] After the conventional document merge is completed at the operation 914. Operational flow proceeds to an operation 916. At the operation 916, a document identifier is assigned to the merged document. The document identifier is unique and
30 persistent. The document identifier persists for the life of the merged document.

[0061] Operational flow proceeds to an operation 918. At the operation 918, each data unit in the merged document is assigned a unique unit identifier that persists for the life of the data unit. The unit identifier identifies and distinguishes the associated data unit from any other data units in the merged document.

5 [0062] Operational flow proceeds to an operation 920. At the operation 920, the unit identifiers are stored within the merged document. In one possible embodiment, the unit identifiers are stored in a metadata within the document. In another possible embodiment, the unit identifiers are stored in a metadata file which is separated from the merged document.

10 [0063] Operational flow proceeds to an operation 922. At the operation 922, an edit identifier is assigned to each of the data units. The edit identifier indicates that the data unit is newly created or edited. The edit identifier is separated from the unit identifier.

[0064] Operational flow proceeds to an operation 924. At the operation 924, the
15 edit identifiers are stored within the merged document. In one possible embodiment, the edit identifiers are stored in a metadata within the merged document. In another possible embodiment, the edit identifiers are stored in a metadata file which is separated from the merged document. In yet another possible embodiment, the edit identifiers are stored when a user saves the document. In addition, an edit identifier can be assigned with a
20 random identifier.

[0065] After the operations 916-924, the merged document is converted into a document which is ready for an accelerated merge when the merged document is edited and/or duplicated later. For example, if the merged document is modified to a fourth document by one user and to a fifth document by another user, the fourth document and
25 the fifth document can be merged in an accelerated merge method as described above.

[0066] The example shown in FIG. 9 illustrates a method of merging two documents together to a merged third document. However, it is not limited to merging two documents only. Instead, the concept and principle illustrated herein can apply to merging more than two documents.

30 [0067] Although the subject matter has been described in language specific to structural features and/or methodological acts, it is to be understood that the subject matter

defined in the appended claims is not necessarily limited to the specific features or acts described above. Rather, the specific features and acts described above are disclosed as example forms of implementing the claims.

WHAT IS CLAIMED IS:

1. A method of merging a first document (220) with at least a second document (240) to generate a third document (260), each of the first and second documents (220, 240) including one or more data units (124a, 124b, 124c, 124d, 124e, 124f), each data unit (124a, 124b, 124c, 124d, 124e, 124f) being associated with a unit identifier (U1, U2, U3, U4) and an edit identifier (E1, E3, E4 E20, E21), the method comprising:
 - obtaining the unit identifiers (U1, U2, U3) of the data units (124a, 124b, 124c) of the first document (220);
 - obtaining the unit identifiers (U1, U2, U4) of the data units (124d, 124e, 124f) of the second document (240);
 - comparing the unit identifiers (U1, U2, U3, U4) of each of the first and second documents (220, 240) to the unit identifiers (U1, U2, U3, U4) of the other of the first and second documents (220, 240) to determine whether each of the unit identifiers (U1, U2, U3, U4) is a matching unit identifier or a non-matching unit identifier, wherein each of the matching unit identifiers indicates corresponding data units (124a, 124b, 124c, 124d, 124e, 124f) of the first and second documents (220, 240), and wherein each of the non-matching unit identifiers indicates a data unit (124a, 124b, 124c, 124d, 124e, 124f) of one of the first and second documents (220, 240) that does not correspond with any data unit (124a, 124b, 124c, 124d, 124e, 124f) of the other of the first and second documents (220, 240);
 - comparing edit identifiers (E1, E3, E4 E20, E21) of the data units (124a, 124b, 124c, 124d, 124e, 124f) associated with the matching unit identifiers to ascertain matching edit identifiers, each of the matching edit identifiers indicating the data unit (124a, 124b, 124c, 124d, 124e, 124f) of one of the first and second documents (220, 240) matches the corresponding data unit (124a, 124b, 124c, 124d, 124e, 124f) of the other of the first and second documents (220, 240);
 - inserting directly into the third document (260) data units (124a, 124b, 124c, 124d, 124e, 124f) of the first and second documents (220, 240) that are associated with non-matching unit identifiers;

inserting directly into the third document (260) data units (124a, 124b, 124c, 124d, 124e, 124f) of the first and second documents (220, 240) that are associated with matching unit identifiers and matching edit identifiers; and

merging into the third document (260) data units (124a, 124b, 124c, 124d, 124e, 124f) of the first and second documents (220, 240) that are associated with matching unit identifiers and non-matching edit identifiers.

2. The method of claim 1, further comprising:

assigning a unit identifier to each data unit of the first document; and

10 storing the unit identifiers within the first document.

3. The method of claim 1, further comprising:

assigning an edit identifier to each data unit of the first document; and

storing the edit identifiers within the first document.

15

4. The method of claim 3, wherein storing the edit identifiers comprises

storing the edit identifiers when a user saves the first document.

5. The method of claim 3, wherein assigning the edit identifier comprises

20 generating a random identifier.

6. The method of claim 1, wherein obtaining the unit identifiers of the data

units of the first document comprises obtaining paragraph identifiers of paragraphs of the first document.

25

7. A document processing system (100) comprising:

a document processing application (110) configured to merge together at least two documents (220, 240) to generate a third document (260); and

a first document (120) configured to be processed by the document processing application (110), the first document (120) storing content (130) and metadata (140), the content (130) of the first document (120) including at least one data unit (124), the metadata (140) of the first document (120) including a unit identifier (126) and an edit

30

identifier (128) associated with each data unit, the unit identifier (126) of each data unit being generated when the data unit (124) is created, the edit identifier (128) of each data unit being generated when a modification to the data unit (124) is saved.

5 8. The document processing system of claim 7, wherein the document processing application is configured to generate the unit identifier of each data unit when the data unit is created.

10 9. The document processing system of claim 7, wherein the document processing application is configured to generate the edit identifier of each data unit when the modification is stored within the first document.

15 10. The document processing system of claim 7, further comprising:
a metadata file stored separately from the first document, the metadata file including a metadata pertaining to the content of the first document, the metadata of the metadata file including a unit identifier indicating to which data unit the metadata pertains.

20 11. The document processing system of claim 10, wherein the metadata stored in the metadata file is associated with the data unit indicated by the unit identifier included in the metadata.

25 12. The document processing system of claim 7, wherein the document processing application is configured to compare the unit identifiers of the first document to unit identifiers of a second document to determine whether each of the unit identifiers is a matching unit identifier or a non-matching unit identifier, wherein each of the matching unit identifiers indicates corresponding data units of the first and second documents, and wherein each of the non-matching unit identifiers indicates a data unit of one of the first and second documents that does not correspond with any data unit of the other of the first and second documents .

30

13. The document processing system of claim 12, wherein the document processing application is configured to compare edit identifiers of the data units associated

with the matching unit identifiers to ascertain matching edit identifiers, each of the matching edit identifiers indicating the data unit of one of the first and second documents matches the corresponding data unit of the other of the first and second documents.

5 14. The document processing system of claim 7, wherein the document is a text document.

 15. The document processing system of claim 7, wherein the data unit is a paragraph.

10

 16. A computer-readable storage medium (604) having computer-executable instructions for implementing a method of generating a merged document (260) from a first document (220) and a second document (240), the method comprising:

 determining whether the first document (220) has a first document identifier (U1, U2, U3), the first document identifier (U1, U2, U3) identifying the first document(220);

15

 obtaining the first document identifier (U1, U2, U3) if the first document (220) has a first document identifier (U1, U2, U3);

 determining whether the second document (240) has a second document identifier (U1, U2, U4), the second document identifier (U1, U2, U4) identifying the second

20

document (240);

 obtaining the second document identifier (U1, U2, U4) if the second document (220) has a second document identifier (U1, U2, U4);

 comparing the first document identifier (U1, U2, U3) of the first document (220) with the second document identifier (U1, U2, U4) of the second document (240);

25

 performing a conventional merge of the first and second documents (220, 240) to generate the merged document (260) if any of the first and second documents (220, 240) does not have a document identifier (U1, U2, U3, U4);

 performing a conventional merge of the first and second documents (220, 240) to generate the merged document (260) if the first document identifier (U1, U2, U3) of the first document (220) differs from the second document identifier (U1, U2, U4) of the second document (240); and

30

performing an accelerated merge of the first and second documents (220, 240) to generate the merged document (260) if the first document identifier (U1, U2, U3) of the first document (220) matches the second document identifier (U1, U2, U4) of the second document (240).

5

17. The computer-readable storage medium of claim 16, wherein performing the accelerated merge comprises:

identifying which data units of the first document match data units of the second document by comparing unique, persistent identifiers of each data unit of the first

10 document to unique, persistent identifiers of each data unit of the second document;

inserting the identified data units into the merged document directly to create data units of the merged document; and

merging each data unit of one of the first and second documents that does not match any data unit of the other of the first and second documents using a conventional
15 merge to further create the data units of the merged document.

18. The computer-readable storage medium of claim 16, wherein the method further comprises:

adding unique, persistent identifiers to each of the data units of the merged
20 document if any of the first and second documents does not have a document identifier; and

adding unique, persistent identifiers to each of the data units of the merged document if the document identifier of the first document differs from the document identifier of the second document.

25

19. The computer-readable storage medium of claim 18, wherein adding the unique, persistent identifiers to each of the data units of the merged document comprises storing the unique, persistent identifiers within the merged document.

30 20. The computer-readable storage medium of claim 17, wherein comparing the unique, persistent identifiers of each data unit of the first document to the unique, persistent identifiers of each data unit of the second document comprises:

comparing a unit identifier and an edit identifier of each data unit of the first document to a unit identifier and an edit identifier of each data unit of the second document.

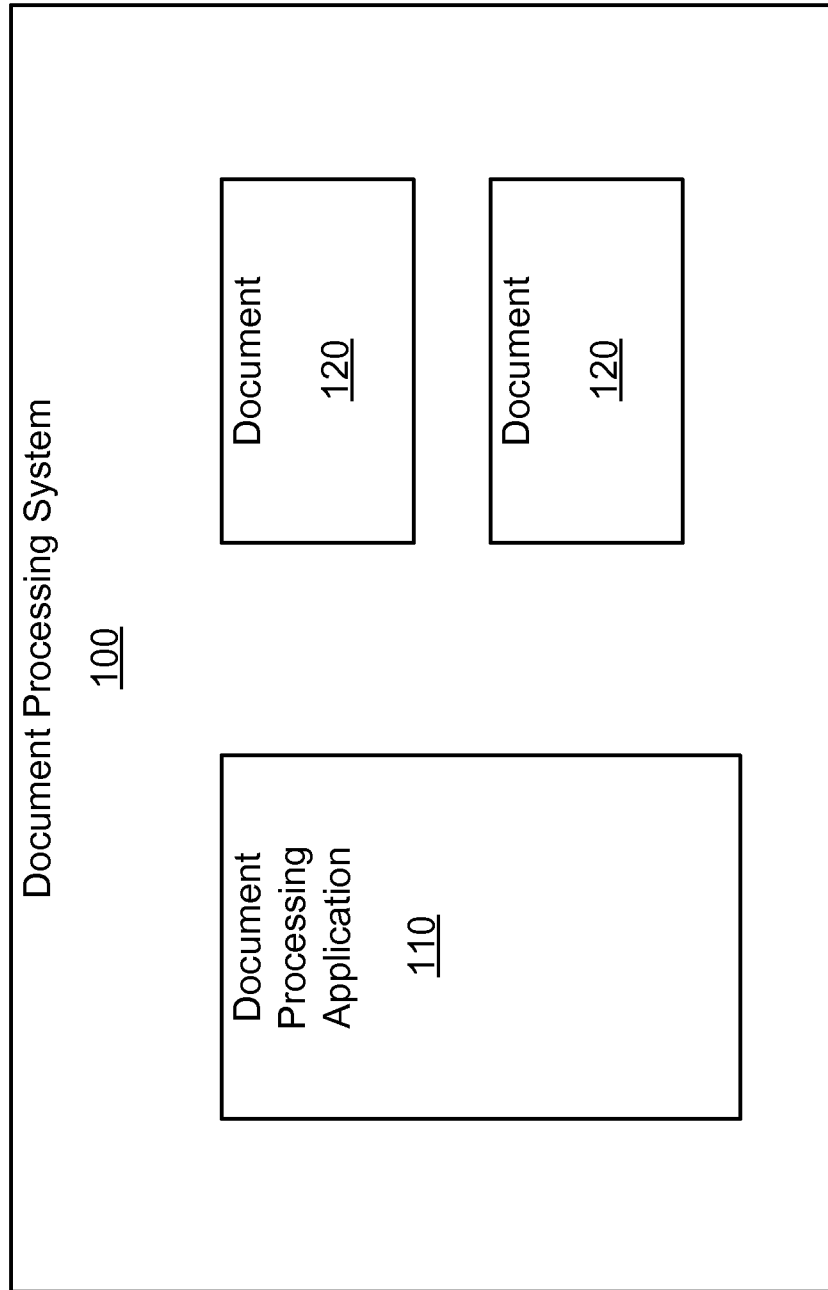


FIG. 1

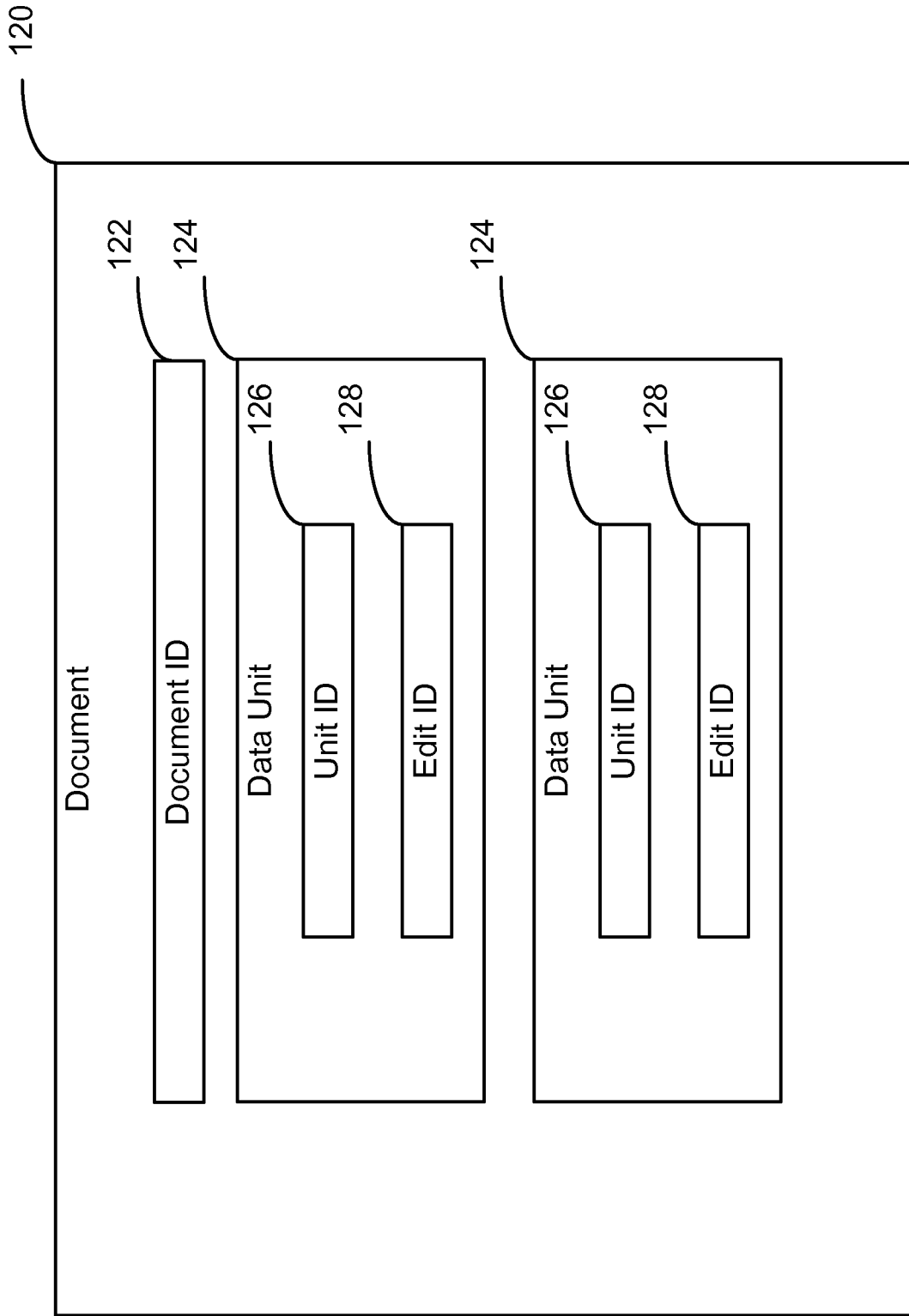


FIG. 2

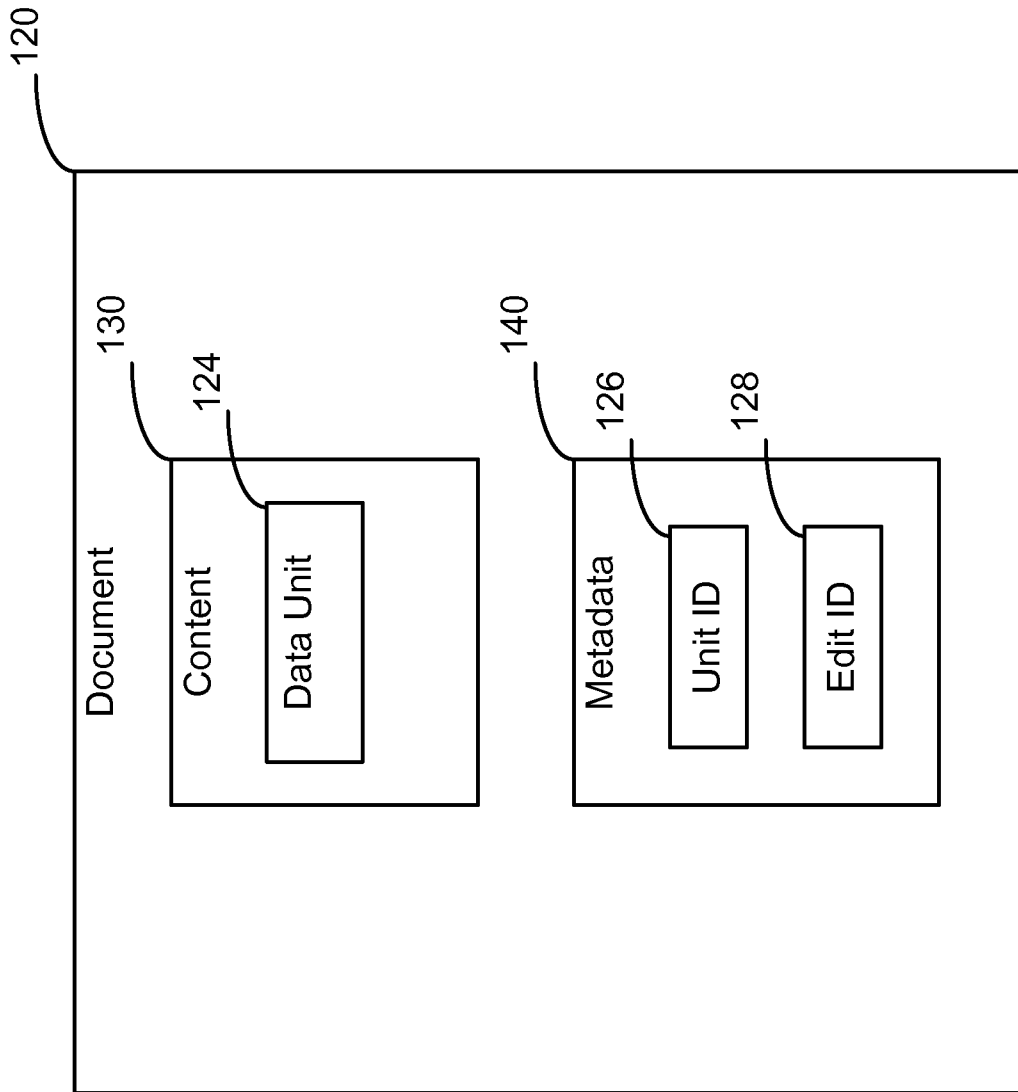


FIG. 3

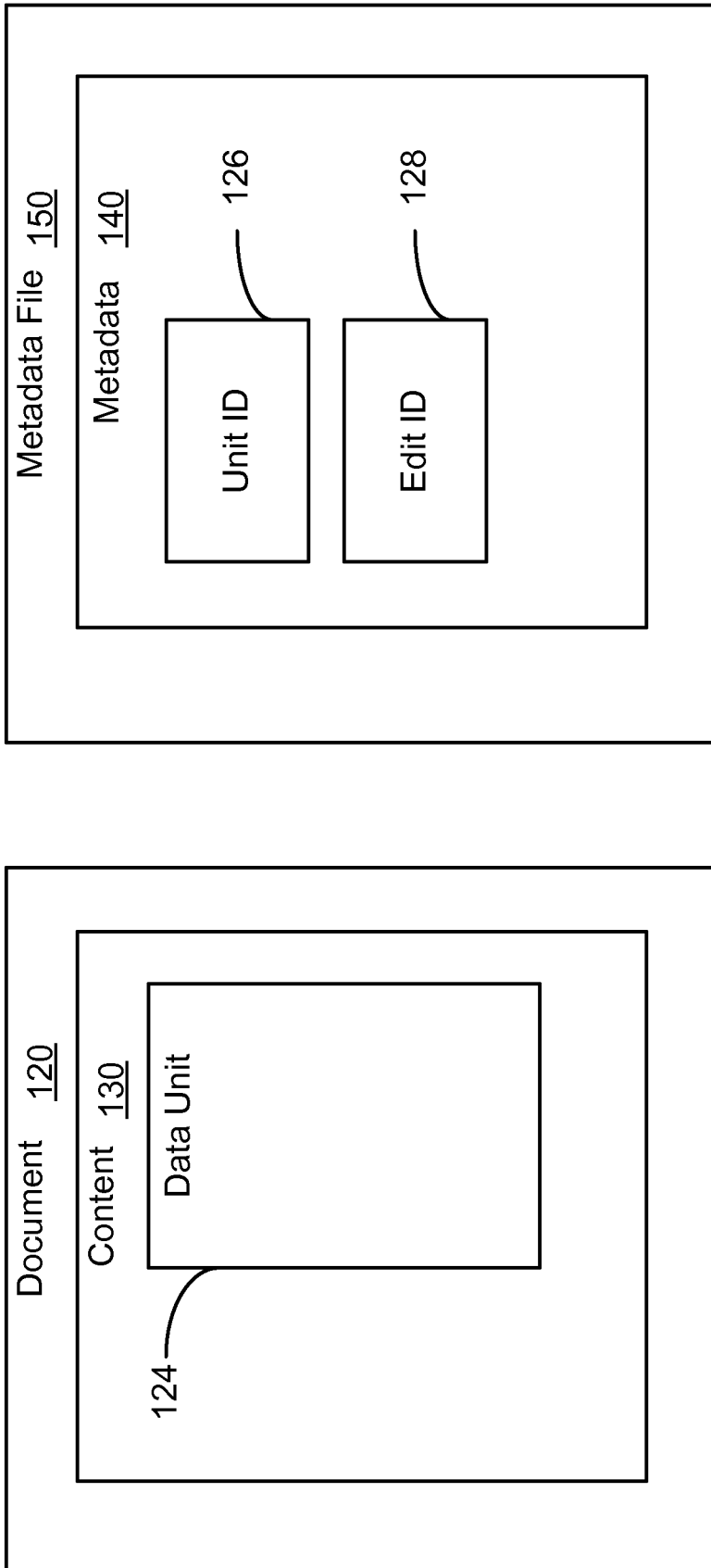


FIG. 4

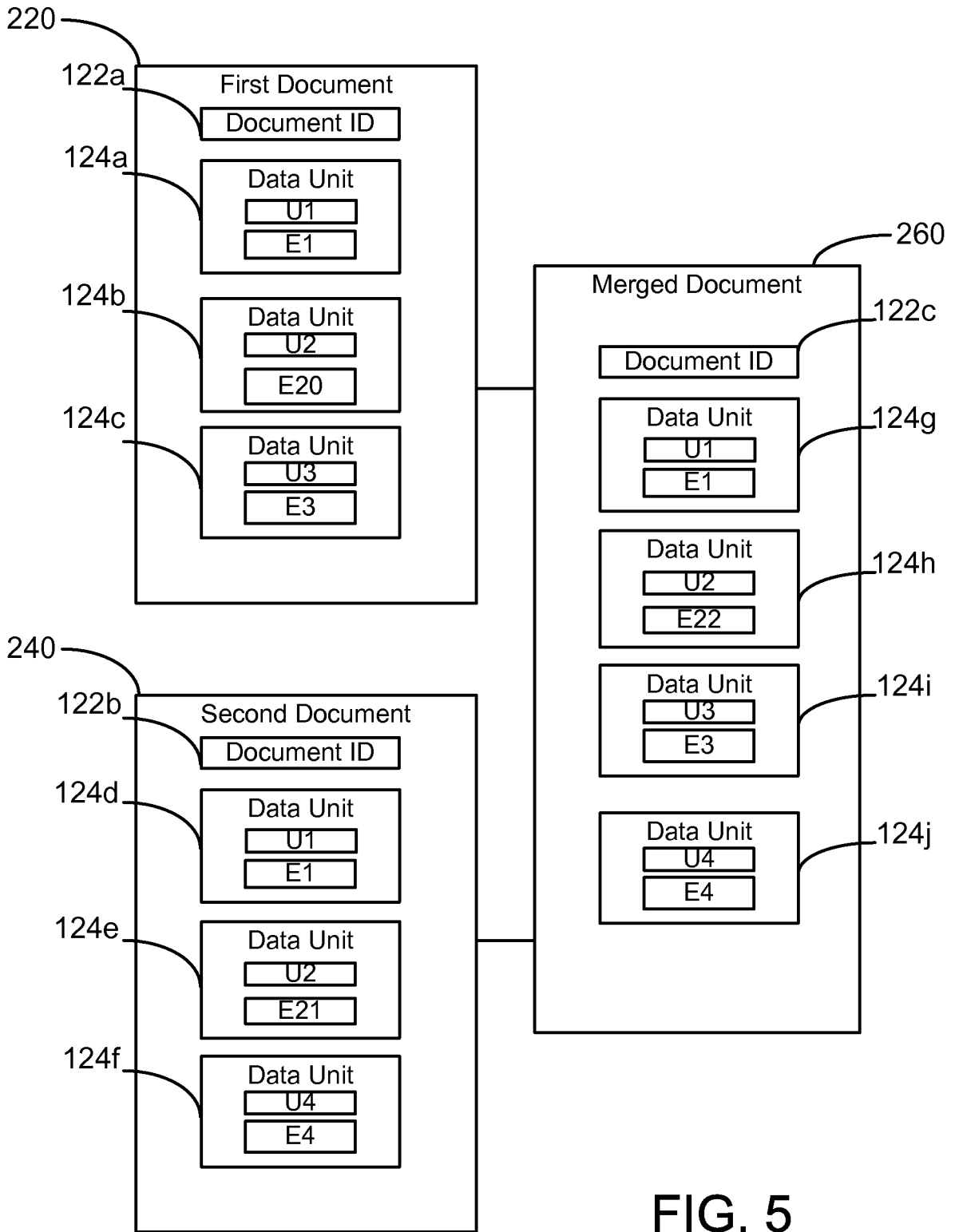


FIG. 5

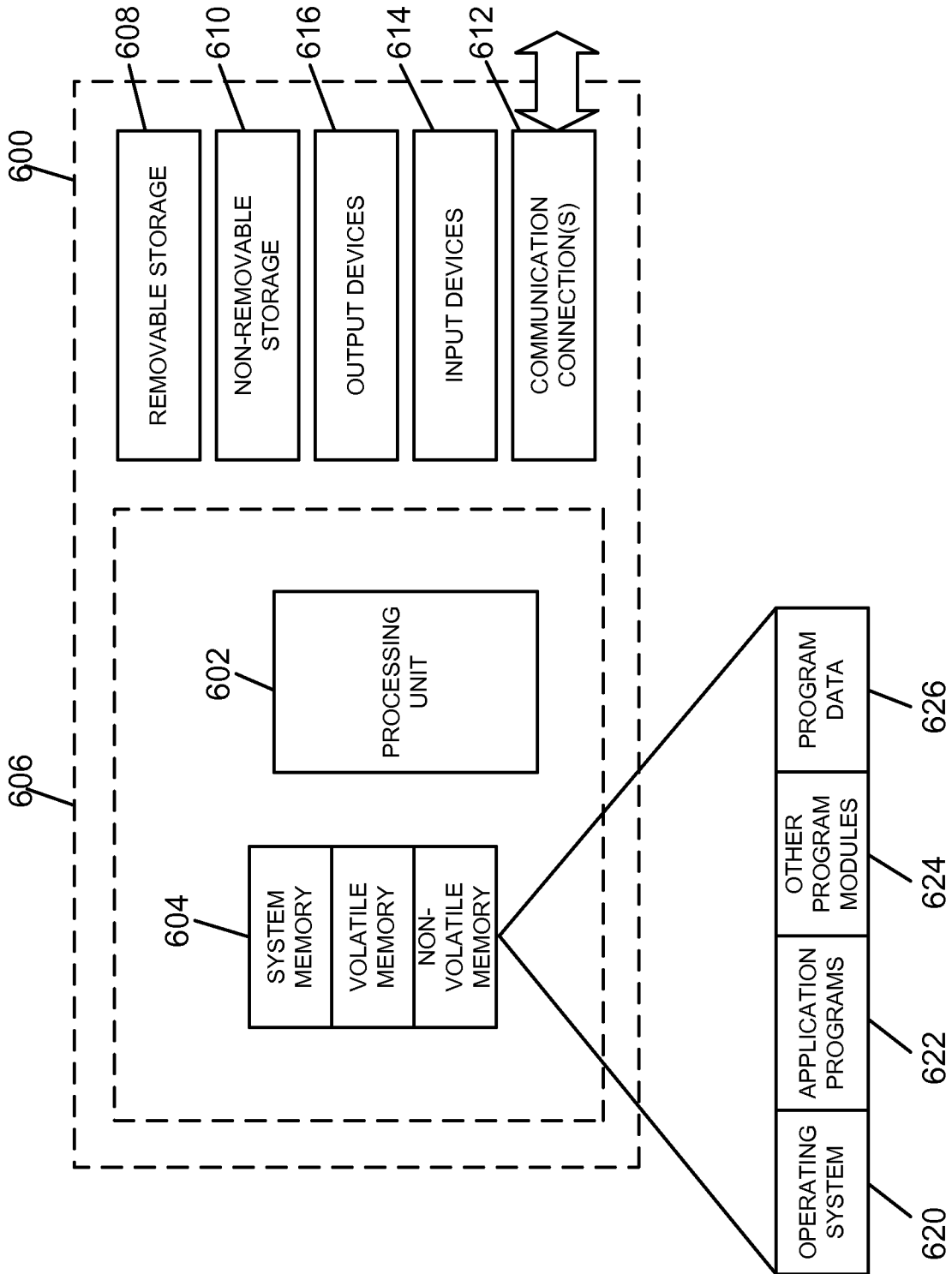


FIG. 6

700

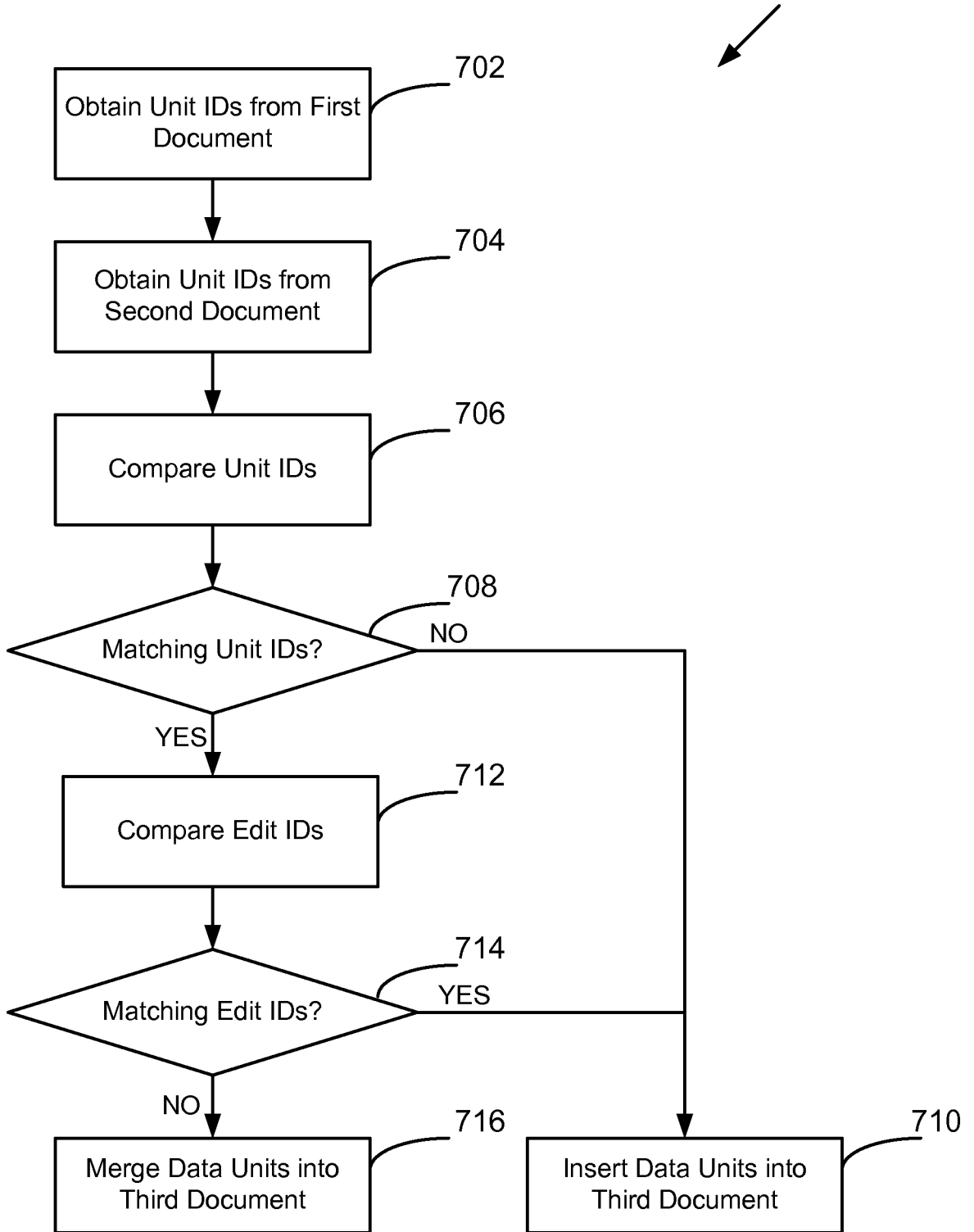


FIG. 7

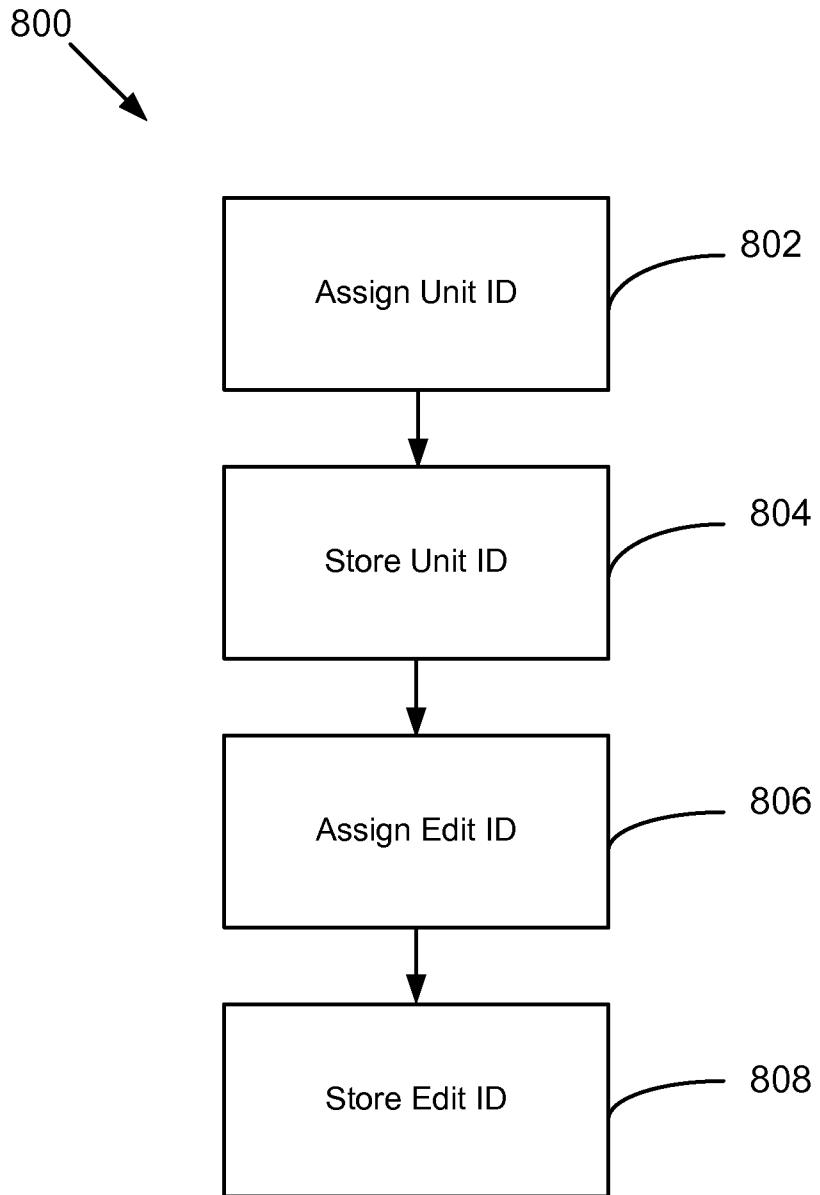


FIG. 8

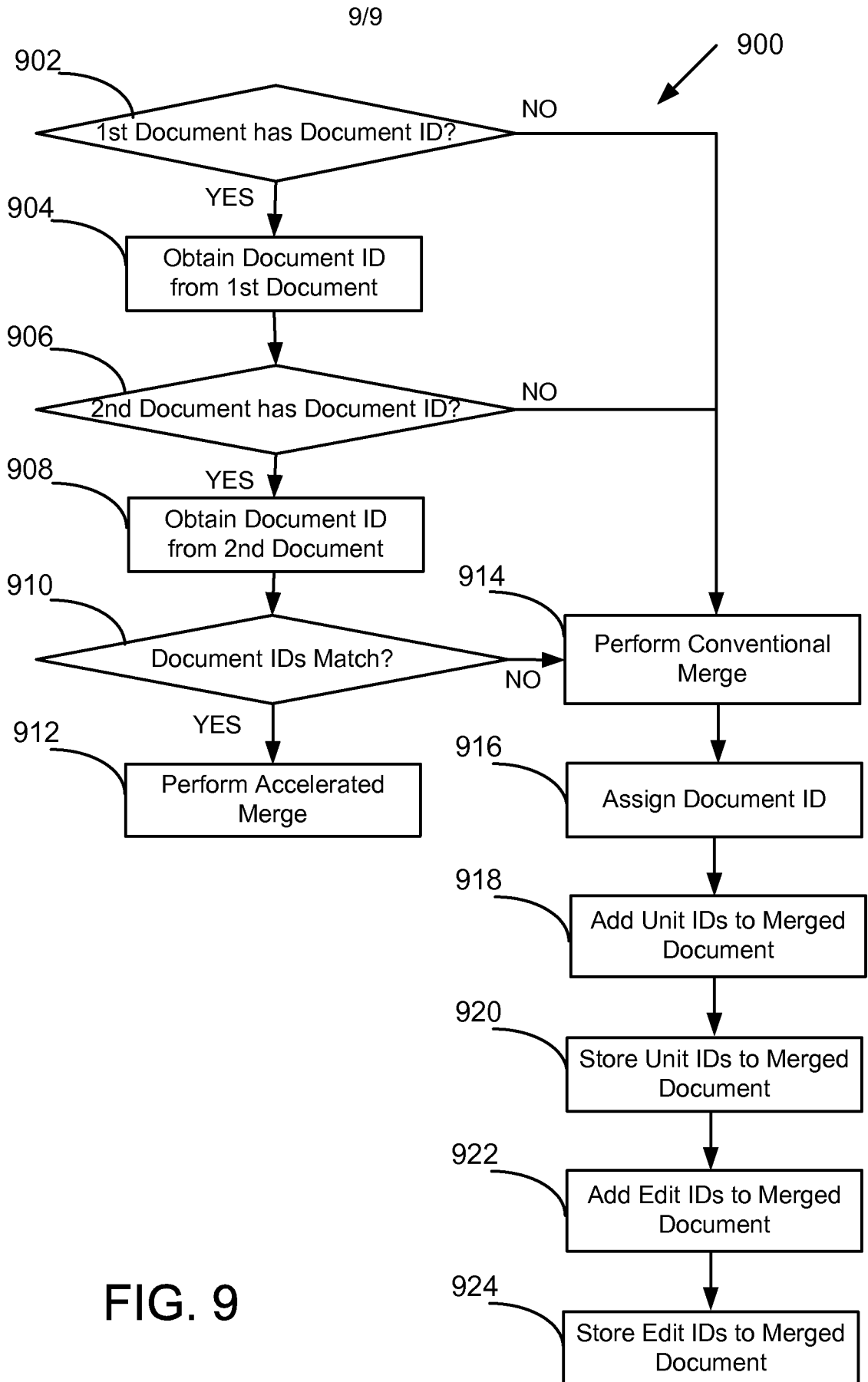


FIG. 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2008/083862**A. CLASSIFICATION OF SUBJECT MATTER****G06F 17/24(2006.01)i, G06F 17/21(2006.01)i**

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)

IPC : G06F

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

Korean Utility models and applications for Utility models since 1975
Japanese Utility models and applications for Utility models since 1975Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
eKIPASS(KIPO internal) "merge, identifier, data unit, edit, match, metadata, compare, insert"**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 06772165 B2 (GARRETT O'CARROLL) 3 AUGUST 2004 see abstract, column 8 line 16 - column 10 line 7, Figures 1, 9-11.	1-20
A	US 07200668 B2 (MINGCHI S. MAK et al.) 3 APRIL 2007 see abstract, column 17 line 42 - column 19 line 21, Figures 2-3B.	1-20
A	US 06757678 B2 (JUSSI PETRI MYLLYMAKI) 29 JUNE 2004 see abstract, column 5 line 38 - column 7 line 15, Figures 2-3.	1-20

 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 application or patent 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 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

"&" document member of the same patent family

Date of the actual completion of the international search

31 MARCH 2009 (31.03.2009)

Date of mailing of the international search report

31 MARCH 2009 (31.03.2009)

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex-Daejeon, 139 Seonsa-ro, Seo-
gu, Daejeon 302-701, Republic of Korea

Facsimile No. 82-42-472-7140

Authorized officer

AN, BYUNG IL

Telephone No. 82-42-481-8471



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2008/083862

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 6772165 B2	03.08.2004	AT 297571 T	15.06.2005
		AU 2001-56612 A1	26.11.2001
		DE 60111376 D1	14.07.2005
		DE 60111376 T2	16.03.2006
		EP 1290575 A1	12.03.2003
		EP 1290575 B1	08.06.2005
		IE 20010467 A1	28.11.2001
		US 2003-0093755 A1	15.05.2003
		WO 01-88750 A1	22.11.2001
		=====	=====
US 7200668 B2	03.04.2007	US 2003-172168 A1	11.09.2003
=====	=====	=====	=====
US 6757678 B2	29.06.2004	US 2002-188598 A1	12.12.2002