(12) STANDARD PATENT (11) Application No. **AU 2015211150 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Document page identifiers from selected page region content

(51) International Patent Classification(s)
G06F 40/143 (2020.01) ***G06F 16/35*** (2019.01)

(21) Application No: **2015211150** (22) Date of Filing: **2015.01.27**

(87) WIPO No: **WO15/116602**

(30) Priority Data

(31) Number	(32) Date	(33) Country
14/171,300	**2014.02.03**	**US**

(43) Publication Date: **2015.08.06**
(44) Accepted Journal Date: **2020.03.12**

(71) Applicant(s)
Bluebeam Software, Inc.

(72) Inventor(s)
Hartmann, Brian;Noyes, Peter

(74) Agent / Attorney
Davies Collison Cave Pty Ltd, Level 15 1 Nicholson Street, MELBOURNE, VIC, 3000, AU

(56) Related Art
US 8566711 B1
US 2002/0118379 A1
US 2012/0047090 A1

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 August 2015 (06.08.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/116602 A1

(51) International Patent Classification:
G06F 17/21 (2006.01) *G06F 17/30* (2006.01)

(21) International Application Number:
PCT/US2015/013140

(22) International Filing Date:
27 January 2015 (27.01.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
14/171,300 3 February 2014 (03.02.2014) US

(71) Applicant: **BLUEBEAM SOFTWARE, INC.** [US/US]; 55 South Lake Avenue, Suite 900, Pasadena, California 91101 (US).

(72) Inventors: **HARTMANN, Brian**; 2720 Night Jasmine Drive, Simi Valley, California 93065 (US). **NOYES, Peter**; 8826 Riderwood Drive, Sunland, California 91040 (US).

(74) Agent: **TANEZAKI, Eric, L.**; STETINA BRUNDA GARRED & BRUCKER, 75 Enterprise, Suite 250, Aliso Viejo, California 92656 (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, BN, BR, BW, BY, BZ, CA, CH, CL, 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, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, 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, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,

[Continued on next page]

(54) Title: DOCUMENT PAGE IDENTIFIERS FROM SELECTED PAGE REGION CONTENT

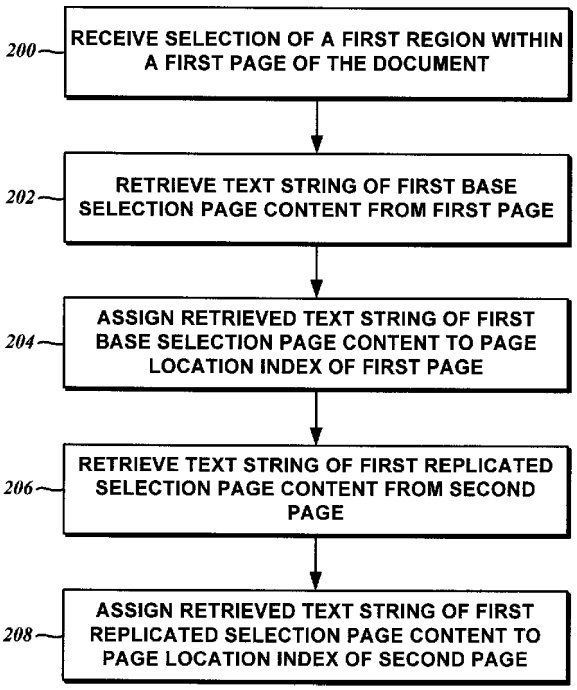

FIG. 6

(57) Abstract: Generating unique document identifiers from content within a selected page region is disclosed. A selection of a first region within a first page of the documents is received from a user, and is defined by a set of first boundaries relative to the first page. A text string of a first base selection page content within the first region is retrieved from the first page. Then the retrieved text string is assigned to a page location index associated with the first page. A text string of a first replicated selection page content is retrieved from a second page. The first replicated selection page content is included in the same first region defined by the set of first boundaries relative to the second page. The retrieved text string of the first replicated selection page content is assigned to a page location index of the second page.

WO 2015/116602 A1

SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

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 (Art. 21(3))*

DOCUMENT PAGE IDENTIFIERS FROM SELECTED PAGE REGION CONTENT

TECHNICAL FIELD

The present disclosure relates generally to electronic document management, and more particularly, to generating, automatically without user intervention, unique document page identifiers from content within a selected page region.

BACKGROUND

The creation, distribution, and management of information are core functions of business. Information or content can be presented in a variety of different ways, including word processing documents, spreadsheets, graphics, photographs, engineering drawings, architectural plans, and so forth. In electronic form, these are generally referred to as documents, and may be generated and manipulated by computer software applications that are specific thereto. The workflows of creating, reviewing, and/or editing electronic documents have evolved to accommodate the specific requirements of various fields, though the need for a device-independent, resolution-independent file format led to the widespread adoption of the Portable Document Format (PDF), amongst other competing formats. Accordingly, different platforms having a wide variety of operating systems, application programs, and processing and graphic display capabilities can be accommodated regardless of the particulars of the workflow.

The PDF standard is a combination of a number of technologies, including a simplified PostScript interpreter subsystem, a font embedding subsystem, and a storage subsystem. As those in the art will recognize, PostScript is a page description language for generating the layout and the graphics of a document. Further, per the requirements of the PDF storage subsystem, all elements of the document, including text, vector graphics, and raster (bitmap) graphics, collectively referred to herein as graphic elements, are encapsulated into a single file. The graphic elements are not encoded to a specific operating system, software application, or hardware, but are designed to be rendered in the same manner regardless of the specificities relating to the system writing or reading such data. The cross-platform capability of PDF aided

in its widespread adoption, and is now a de facto document exchange standard. Although originally proprietary, PDF has been released as an open standard published by the International Organization for Standardization (ISO) as ISO/IEC 3200-1:2008. Currently, PDF is utilized to encode a wide variety of document types, including those composed largely of text, and those composed largely of vector and raster graphics. Because of its versatility and universality, files in the PDF format are often preferred over more particularized file formats of specific applications.

In technical fields such as engineering and architecture, one project typically involves multiple aspects with numerous professionals spanning a wide range of disciplines. The planning documents, e.g., drawings, are specific to each discipline, though a change in one aspect may require a corresponding change in another aspect, and so on. For example, in a building construction project, there may be one set of plans for the structural aspect, while there may be another set of plans for the heating/ventilation/air conditioning (HVAC) aspect, and another set of plans for plumbing, another set for electrical, etc. A high level of detail is necessary in the planning documents to accurately convey the specifications of the project so that it can be correctly implemented. Although the ability to zoom in and zoom out of an electronic document alleviates this issue to a certain degree, the size and the amount of information contained in any one page must nevertheless remain manageable while retaining all the necessary detail so that viewing, editing, and annotating do not require complicated inputs/interface manipulations. Thus, the contents are separated into multiple pages.

In a typical set of drawings, whether stored in a PDF or otherwise, a standard convention is utilized to present, in an organized fashion, header information such as the title, drawing number, project name/identifier, facility identifier and/or address, measurement units, and so forth. This convention is typically the title block, which is usually positioned at the same location on each of the pages in the document. While this header information is useful when viewing the particular page on which it is located, it is a part of the document content itself, and cannot be used by the viewing/editing application to catalog and organize the document. The extent of any metadata that is stored in connection with a page is oftentimes

limited to the page number relative to the other pages in the document, without any further descriptors.

Adding such descriptive information to label each page is a painstaking, error-prone, and time-consuming process that requires human intervention. After visually searching for and ascertaining the desired header information from the contents of each page, conventional processes require the manual keying of the same into a form field via the user interface of the editing application. The added metadata could thereafter be used for subsequent searching and organization purposes. The aforementioned procedure was required regardless of whether a bookmark was being created for a particular location on a page or a label was being applied to the page. Although labeling/bookmarking a document having only one or two pages may be trivial, typical project planning documents span many tens to hundreds of pages. Furthermore, the information from multiple different parts of the page could be needed for generating precise descriptors. For such larger, more complex documents, the time necessary to complete this task can increase to several hours.

It is desired to provide a method that alleviates one or more difficulties of the prior art, or that at least provides a useful alternative.

SUMMARY

There is a need in the art for methods to generate, automatically without user intervention, unique document page identifiers from content within a selected page region. Accordingly, the automatic generation of bookmarks or page labels on a multi-page document with text data extracted from the content of one or more regions of the document content is disclosed. In accordance with one embodiment of the present disclosure, there is a method for auto-indexing an electronic document stored on a computer and having multiple pages each having page content. In accordance with the present invention, there is provided a method for auto-indexing an electronic document stored on a computer and having multiple pages, each page having page content, the method comprising:

receiving, from a user, a selection of a first region within a first one of the multiple pages of the document, the first region being defined by a set of first boundaries relative to the first one of the multiple pages and including a first base selection page content;

retrieving, without user intervention, a text string of the first base selection page content from the first one of the multiple pages;

assigning, without user intervention, the retrieved text string of the first base selection page content to a page location index associated with the first one of the multiple pages;

generating, without user intervention, a first replicated region within a second one of the multiple pages of the document by applying the set of first boundaries to the second one of the multiple pages, the first replicated region including a first replicated selection page content;

retrieving, without user intervention, a text string of the first replicated selection page content from the second one of the multiple pages; and

assigning, without user intervention, the retrieved text string of the first replicated selection page content to a page location index of the second one of the multiple pages.

The present disclosure will be best understood accompanying by reference to the following detailed description when read in conjunction with the drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

Some embodiments of the present invention are hereinafter described, by way of example only, with reference to the accompanying drawings, wherein:

FIG. 1 is an exemplary computer system that may be capable to perform functions for use with embodiments of the present invention including a display monitor;

FIG. 2 is a block diagram of system components of the exemplary computer system of FIG. 1;

FIG. 3 is a diagram showing an exemplary structure of the document including multiple pages each with content and a location index;

FIG. 4 is a dialog window of a user interface for a document management software application that implements various aspects of the contemplated methods for generating unique page identifiers from content within a selected page region, the dialog window

2015211150 18 Feb 2020

including interface elements for defining the selected page region as well as interface elements for initiating the procedure to generate the unique page identifiers;

FIG. 5A is an excerpted section of the user interface for the document management software application showing a first one of the multiple pages of a document rendered for viewing;

FIG. 5B is another excerpted section of the user interface for the document management software application in which a second one of the multiple pages of a document is rendered for viewing;

FIG. 6 is a flowchart depicting one embodiment of the method for generating unique page identifiers in accordance with the present disclosure;

FIG. 7A illustrates one implementation for receiving the selection of a first page region;

FIG. 7B illustrates receiving the selection of a second page region; and

FIG. 8 is another dialog window of a user interface for the document management software application, the dialog window including interface elements for defining the selected page region as well as interface elements for initiating the procedure to generate the unique page identifiers.

Common reference numerals are used throughout the drawings and the detailed description to indicate the same elements.

DETAILED DESCRIPTION

The detailed description set forth below in connection with the appended drawings is intended as a description of the presently preferred embodiment methods for generating unique document page identifiers from content within a selected page region. It is not intended to represent the only form in which the present method may be developed or utilized. The description sets forth the functions and the sequence of steps for developing and operating the invention in connection with the illustrated embodiment. It is to be understood, however, that the same or equivalent functions and sequences may be accomplished by different embodiments that are also intended to be encompassed within the spirit and scope of the invention. It is further understood that the use of relational terms such as first, second, and the like are used solely to distinguish one from another entity without necessarily requiring or implying any actual such relationship or order between such entities.

FIG. 1 depicts an exemplary hardware environment in which various aspects of the contemplated method of document management and arrangement may be implemented. Generally, there is a computer system 10 with a system unit 12 and a display device 14. The display device 14 produced a graphical output from the data processing operations performed by the system unit 12. Input devices including a keyboard 16 and a mouse 18, for example, may be manipulated by a user to generate corresponding inputs to the data processing operations, and are connected to the system unit 12 via ports 20. Various other input and output devices may be connected to the system unit 12, and different interconnection modalities are known in the art.

As shown in the block diagram of FIG. 2, the system unit 12 includes a processor (CPU) 22, which may be any conventional type. A system memory (RAM) 24 temporarily stores results of the data processing operations performed by the CPU 22, and is interconnected thereto typically via a dedicated memory channel 26. The system unit 12 may also include permanent storage devices such as a hard drive 28, which is also in communication with the CPU 22 over an input/output (I/O) bus 30. A dedicated graphics module 32 may also connected to the CPU 22 via a video bus 34, and transmits signals representative of display data to the display device 14. As

indicated above, the keyboard 16 and the mouse 18 are connected to the system unit 12 over the port 20. In the embodiments where the port 20 is a USB type, there may be a USB controller 35 that translates data and instructions to and from the CPU 22 for the external peripherals connected via the port 20. Additional devices such as printers, microphones, speakers, and the like may be connected to the system unit 12 thereby.

The system unit 12 may utilize any operating system having a graphical user interface (GUI), such as WINDOWS from Microsoft Corporation of Redmond, Washington, MAC OS from Apple, Inc. of Cupertino, CA, various versions of UNIX with the X-Windows windowing system, and so forth. The system unit 12 executes one or more computer programs, with the results thereof being displayed on the display device 14. Generally, the operating system and the computer programs are tangibly embodied in a computer-readable medium, e.g., the hard drive 28. Both the operating system and the computer programs may be loaded from the aforementioned data storage devices into the RAM 24 for execution by the CPU 22. The computer programs may comprise instructions, which, when read and executed by the CPU 22, cause the same to perform or execute the steps or features of the various embodiments set forth in the present disclosure.

According to one embodiment, a user can interact with the computer system 10, and specifically with the graphics being generated on the display device 14, via the mouse 18. The movement of a cursor generated on the display device 14 is tied to the movement of the mouse 18, with further interactivity being provided with input from the mouse buttons. Input from the keyboard 16 also provides interactivity with the computer system 10.

Although certain features of the present disclosure are described in relation to a computer system with such input and output capabilities, specifics thereof are presented by way of example only and not of limitation. Any alternative graphical user interfaces such as touch interfaces and pen/digitizer interfaces may be substituted. The analogs of those features will be readily appreciated, along with suitable modifications to accommodate these alternative interfaces while still achieving the same functionalities.

Along these lines, the foregoing computer system 10 represents only one exemplary apparatus of many otherwise suitable for implementing aspects of the

present disclosure, and only the most basic of the components thereof have been described. It is to be understood that the computer system 10 may include additional components not described herein, and may have different configurations and architectures. Any such alternative is deemed to be within the scope of the present disclosure.

Again, the present disclosure is directed to a process for automatically generating bookmarks and/or page labels on a multi-page document by incorporating text from one or more selected regions on each page. This process may be implemented as a series of executable instructions in a document management software application. As utilized herein, the term document refers to standalone representation of data that can be opened, rendered and/or edited by such document management software application. Accordingly, the document may be formatted and structured in a particular way so that its elements can be rendered consistently. In the exemplary embodiment discussed herein, the document management software application handles PDF documents, and the following description will be in reference thereto. Documents in any other format that are editable or viewable with any other software application may be substituted.

With reference to the diagram of FIG. 3, according to various embodiments of the present disclosure, a document 38 is comprised of one or more pages 40. In the particular example, there is a first page 40a, a second page 40b, and a third page 40c. There could be additional pages 40, though these are not depicted in the diagram. Each of the pages has an associated location index 42 and content 44, and is ordered according to a numerical page number value. When the document is rendered within a user interface of the document management software application, the content 44 within the respective pages 40 may be presented in the order defined by the numerical page number value. Typically, a single document is stored in a single file on a hierarchical file system of the hard drive 28 or any other accessible storage device.

The contemplated process may begin with a precursor step of activating a dialog window 46 as shown in FIG. 4. There are two options presented: a first option 48 to generate page labels based on specific bookmarks that may be included in or appended to the pages 40, and a second option 50 to generate page labels based on a page region. The dialog window 46 may be invoked in various ways from other parts of the user interface of the document management software application. As will be

described in more detail below, it is possible to select a region within the pages 40 from which the page labels are generated, and the specific selecting step is initiated by activating a select button 52. In addition, the range of pages over which the process is to be applied is selected via a dropdown menu 54. The default value may be all of the pages 40 in the document 38, though different ranges may be selected. The particulars of the dialog window 46 have been presented by way of example only and not of limitation, and any suitable interface may be presented to the user to initiate the process.

FIG. 5A shows an excerpted section of the user interface for the document management software application, specifically referred to as a primary window 55. Within the primary window 55, the document 38, and more particularly, the first page 40a, is displayed with a rendering of the content 44 thereof. By way of example only and not of limitation, the content 44 may include a title block 56 with various information therein specific to the first page 40a. FIG. 5B again shows the primary window 55, but with the second page 40b of the document 38 being displayed with a rendering of the particular content 44 thereof. The content 44 likewise includes the title block 56, though with different information therein that pertains specifically to the second page 40b. It is understood that the position of the title block 56 relative to the first page 40a is the same as the position of the title block 56 relative to the second page 40b. In the title block 56 of the first page 40a, an example text “S3.1” and “Foundation Details” is included.

Referring now to the flowchart of FIG. 6, a method for auto-indexing the document 38 stored on the computer system 10 proceeds with a step 200 of receiving a selection of a first region within a first page 40a of the document 38. This step may be initiated in response to detecting the activation of the select button 52 in the dialog window 46. As best shown in the enlarged view of the title block 56 in FIG. 7A, the first region 58 is defined by an upper left hand corner 60a and a diagonally opposed lower right hand corner 60b, and is therefore rectangular in shape. These boundaries are understood to be defined relative to the first page 40a. The first region 58, which encompasses the example text “S3.1”, may be selected by positioning a mouse cursor 62 on the upper left hand corner 60a, clicking the mouse button, and dragging the mouse cursor 62 to the lower right hand corner 60b. In this regard, the selected content, e.g., the example text “S3.1” may also be referenced as a first base selection

page content 64. Other ways for selecting a region on a graphical user interface are known, and such alternatives are deemed to be within the scope of the present disclosure. For instance, a multi-segment polygon could be a basis for selecting the region by designating each polygon segment graphically. In the context of a method particular to applying page labels as well as a method for generating bookmarks, the step 200 is understood to correspond to receiving one or more page region selections on a first one of the plurality of pages 40 that each include a label or bookmark content, respectively, within.

The method continues with a step 202 of retrieving, without user intervention, a text string of the first base selection page content 64 from the first page 40a. The first base selection page content 64 may be stored as basic text/character data within the document 38. In such case, this retrieving step may involve extracting such underlying text data that is determined to be positioned within the first region 58. Alternatively, the first page selection page content 64 may be stored as a vector or raster graphic. If this is the case, the retrieving step may further include parsing such content with an optical character recognition (OCR) module to generate the corresponding text string. The OCR module is externally called from the document management software application, and is seamlessly integrated therewith. For a method particular to applying page labels or a method for generating bookmarks, the step 202 corresponds to extracting, without user intervention, one or more label or bookmark contents, respectively, from a plurality of selected pages as text data. The specific text data that is extracted is based on page location selections that were made by the user in a previous step.

The retrieved text string of the first base selection page content 64 is then assigned, without user intervention, to the page location index 42a associated with the first page 40a in accordance with a step 204 of the method. The page location index 42a is understood to be, for example, a page label metadata that serves as a descriptor of the first page 40a for organization and searching purposes. Alternatively, the page location index 42 may be a bookmark that is appended to the first page 40a. The step 204 is understood to be a corollary to assigning the text data of the label or bookmark content to the particular page from which the particular text data was extracted in the earlier step directed in the methods for automatically applying labels or generating bookmarks.

The aforementioned procedure is repeated for each subsequent page 40 that is part of the selected page range of the document 38. In particular, the method includes a step 206 of retrieving, without user intervention, a text string of a first replicated selection page content 66. As indicated above, the user designates a selection of the first region 58, though in the spatial context of the first page 40a. The underlying text or image data found in the same first region 58 as defined by the same boundaries but in the spatial context of the second page 40b is retrieved. The data is parsed by the OCR engine as necessary to retrieve the text string. The example shown in FIG. 5B is that of the second page 40b, and the same region that was originally selected from the first page 40a includes the text "S4.1." According to a step 208, the method includes assigning, without user intervention, the retrieved text string of the first replicated selection page content 66 to a page location index 42b associated with the second page 40b. Thus, different content from the same part of each different page 40 may be assigned as the page location index 42, e.g., the corresponding page label or a bookmark linked to the page.

Before committing the assignments of the respective text strings retrieved from the first base selection page content 64 and the first replicated selection page content 66 to the respective page location indices 42 a preview of the retrieved text strings from the selection of the first region 58 may be presented. Referring to FIG. 8, an exemplary dialog window 68 may be displayed in response to activating select button 52 from the dialog window 46. The dialog window 68 includes a preview 70 of the text string as would be assigned to the page location index 42.

Along these lines, it is envisioned that the user may define the format of the page location index 42. Thus, it is possible to specify a formatting structure 72 in a text input field 74. Keywords corresponding to the selected first region 58, also referred to as field codes, may be designated with surrounding brackets []. The preview 70, as well as any assignment of the text string to the page location index 42 may follow the formatting as designated by the formatting structure 72. This involves substituting the field code with the actual retrieved text string, and retaining all other additional custom text including delimiters and the arrangement thereof as specified in the text input field 74.

As shown in the text input field 74, it is possible to select the first region 58 of the first page 40a for assigning the underlying content as the page location index 42.

With reference to FIG. 7B, it is also possible to select an additional second region 76 on the first page 40a for repeating the same procedure. Once the first region 58 has been selected, the dialog window 68 is presented again. By activating an add button 77, the display of the document 38 as shown in FIG. 5A returns, and the user can select the different second region 76 that includes a second base selection page content 78. The second region 76 is defined by an upper left hand corner 80a and a diagonally opposed lower right hand corner 80b, and is therefore rectangular in shape. These boundaries are also defined relative to the first page 40a. The second region 76, which encompasses the example text "Foundation Details", may be selected by positioning the mouse cursor 62 on the upper left hand corner 80a, clicking the mouse button, and dragging the mouse cursor 62 to the lower right hand corner 80b. The selected content, e.g., the example text "Foundation Details" is understood to be the aforementioned second base selection page content.

When such second region 76 is selected, as shown in the dialog window 68 of FIG. 8, the formatting structure 72 also includes a reference thereto, similarly designated by surrounding brackets []. The formatting structure 72 can further incorporate field delimiters such as a space character, a dash character, or any other suitable character that visually separates the reference to the first region and the reference to the second region.

The text string of the underlying second base selection page content 78 is retrieved in accordance with the procedures outlined above, and shown in the preview 70. After all of the desired regions on the first page 40a are selected, the aforementioned procedure of retrieving the replicated selection page content and assigning the text string to the respective page location index 42 is repeated for each selected region, and concatenated as defined by the formatting structure 72. As shown in FIG. 5B, when processing the second page 40b, the particular content that is positioned at the same relative location as the second region 76, that is, a second replicated selection page content 82, is retrieved. That is, the data from the same regions on subsequent pages are used to generate the page location index 42, e.g., the bookmark or page label, for the corresponding page 40. The processing of the pages 40 is initiated once the "OK" button 84 from the dialog window 68 is activated.

The particulars shown herein are by way of example and for purposes of illustrative discussion of the embodiments of the present disclosure only and are

2015211150 18 Feb 2020

presented in the cause of providing what is believed to be the most useful and readily understood description of the principles and conceptual aspects of the disclosure. In this regard, no attempt is made to show more details than is necessary for a fundamental understanding of the disclosure, the description taken with the drawings making apparent to those skilled in the art how the several forms of the presently disclosed methods may be embodied in practice.

Throughout this specification and claims which follow, unless the context requires otherwise, the word "comprise", and variations such as "comprises" and "comprising", will be understood to imply the inclusion of a stated integer or step or group of integers or steps but not the exclusion of any other integer or step or group of integers or steps.

The reference in this specification to any prior publication (or information derived from it), or to any matter which is known, is not, and should not be taken as an acknowledgment or admission or any form of suggestion that that prior publication (or information derived from it) or known matter forms part of the common general knowledge in the field of endeavour to which this specification relates.

2015211150 18 Feb 2020

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. A method for auto-indexing an electronic document stored on a computer and having multiple pages, each page having page content, the method comprising:

receiving, from a user, a selection of a first region within a first one of the multiple pages of the document, the first region being defined by a set of first boundaries relative to the first one of the multiple pages and including a first base selection page content;

retrieving, without user intervention, a text string of the first base selection page content from the first one of the multiple pages;

assigning, without user intervention, the retrieved text string of the first base selection page content to a page location index associated with the first one of the multiple pages;

generating, without user intervention, a first replicated region within a second one of the multiple pages of the document by applying the set of first boundaries to the second one of the multiple pages, the first replicated region including a first replicated selection page content;

retrieving, without user intervention, a text string of the first replicated selection page content from the second one of the multiple pages; and

assigning, without user intervention, the retrieved text string of the first replicated selection page content to a page location index of the second one of the multiple pages.

2. The method of Claim 1, wherein the page location index is a page label.

3. The method of Claim 1, wherein the one of the page location index is an appended page bookmark.

4. The method of Claim 1, wherein receiving the selection of the first region includes a graphically selected area on the document designated with a cursor input of a first point and a second point corresponding to the set of first boundaries.

5. The method of Claim 1, wherein retrieving the text string of the first base selection page content further includes:

extracting underlying text data positioned within the first base selection page content.

6. The method of Claim 1, wherein retrieving the text string of the first base selection page content further includes:

parsing the first base selection page content with an optical character recognition (OCR) module to generate the text string of the first base selection page content.

7. The method of Claim 6, wherein retrieving the text string of the first replicated selection page content further includes:

parsing the first replicated selection page content with the OCR module, to generate the text string of the first replicated selection page content.

8. The method of Claim 1, further comprising:

receiving a formatting structure including a first field code associated with the first region;

wherein retrieving the text string of the first base selection page content includes arranging the text string of the first base selection page content in accordance with the formatting structure, with the first base selection page content being substituted for the first field code.

9. The method of Claim 7, wherein retrieving the text string of the first replicated selection page content includes arranging the second text string in accordance with the formatting structure, with the first replicated selection page content being substituted for the first field code.

10. The method of Claim 7, further comprising:

displaying a preview of the text string of the first base selection page content as arranged in accordance with the formatting structure.

2015211150 18 Feb 2020

11. The method of Claim 1, wherein the first region is rectangular.

12. The method of Claim 1, further comprising:

receiving a selection of a second region within a first one of the multiple pages of the document, the second region being defined by a set of second boundaries relative to the first one of the multiple pages and including a second base selection page content;

retrieving a text string of the second base selection page content from the first one of the multiple pages;

assigning the retrieved text string of the second base selection page content to a page location index of the first one of the multiple pages;

generating, without user intervention, a second replicated region within the second one of the multiple pages of the document by applying the set of second boundaries to the second one of the multiple pages, the second replicated region including a second replicated selection page content;

retrieving, without user intervention, a text string of the second replicated selection page content from the second one of the multiple pages; and

assigning the retrieved text string of the second replicated selection page content to a page location index of the second one of the multiple pages.

13. The method of Claim 12, further comprising:

receiving a formatting structure including a first field code associated with the first region, a second field code associated with the second region, and at least one field delimiter;

wherein retrieving the text string of the first base selection page content includes concatenating the text strings of the first base selection page content and the second base selection page content and the at least one field delimiter in accordance with the formatting structure, with the first base selection page content being substituted for the first field code and the second base selection page content being substituted for the second field code.

14. The method of Claim 13, further comprising:

displaying a preview of the text strings of the first base selection page content and the second base selection page content as arranged in accordance with the formatting structure.

15. The method of Claim 12, wherein the field delimiter is a space character.

16. The method of Claim 12, wherein:

retrieving the text string of the first base selection page content further includes:

parsing the first base selection page content with an optical character recognition (OCR) module to generate the text string of the first base selection page content;

retrieving the text string of the first replicated selection page content further includes:

parsing the first replicated selection page content with the OCR module, to generate the text string of the first replicated selection page content;

retrieving the text string of the second base selection page content further includes:

parsing the second base selection page content with an optical character recognition (OCR) module to generate the text string of the second base selection page content; and

retrieving the text string of the second replicated selection page content further includes:

parsing the second replicated selection page content with the OCR module, to generate the text string of the second replicated selection page content.

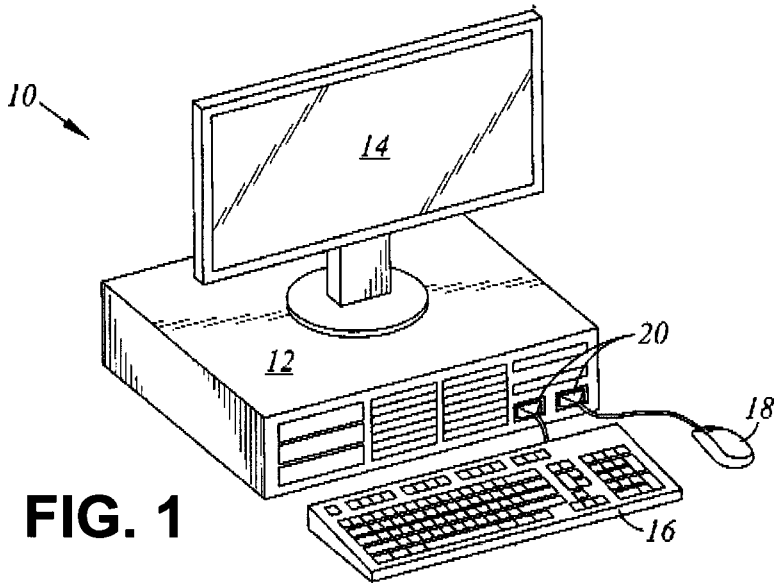
10
14
12
20
18
16

FIG. 1

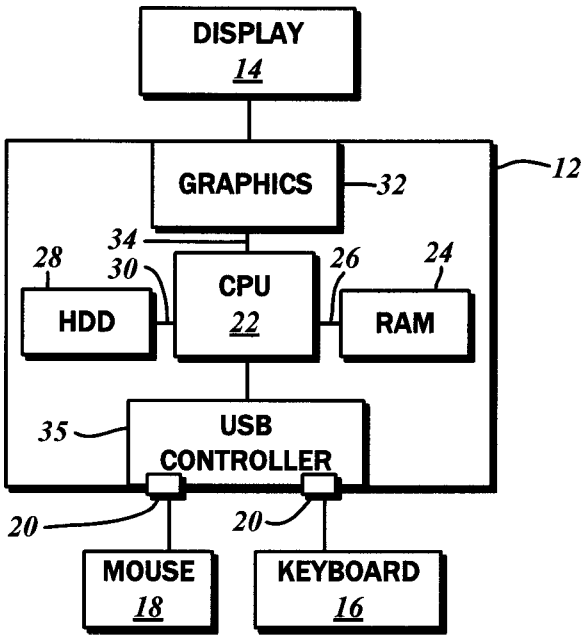
DISPLAY
14
GRAPHICS
32
12
34
28
30
26
24
CPU
22
HDD
RAM
35
USB
CONTROLLER
20
20
MOUSE
18
KEYBOARD
16

FIG. 2

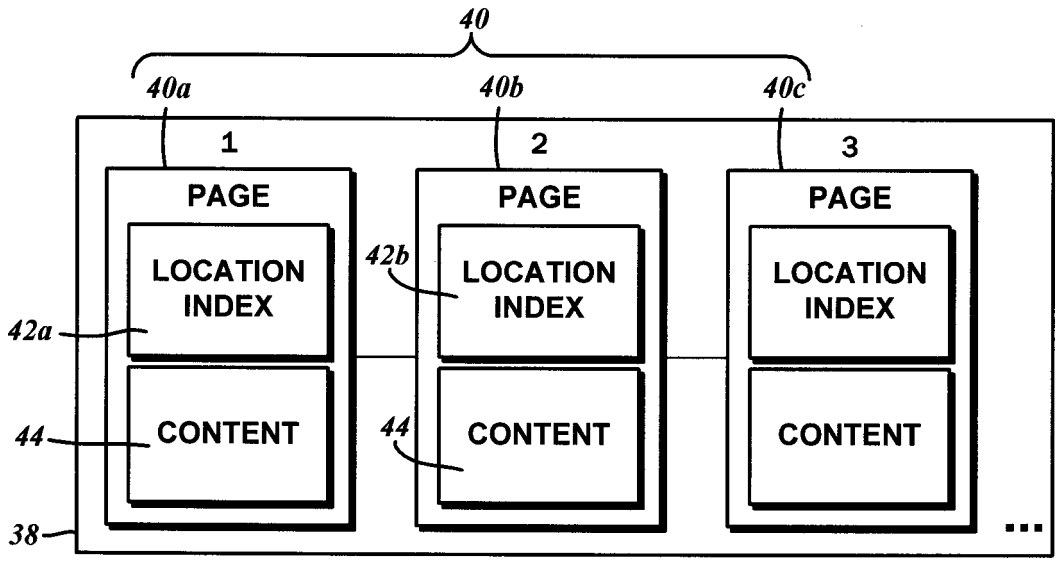
40
40a
40b
40c
1
2
3
PAGE
PAGE
PAGE
LOCATION INDEX
LOCATION INDEX
LOCATION INDEX
42a
42b
CONTENT
CONTENT
CONTENT
44
44
38
...

FIG. 3

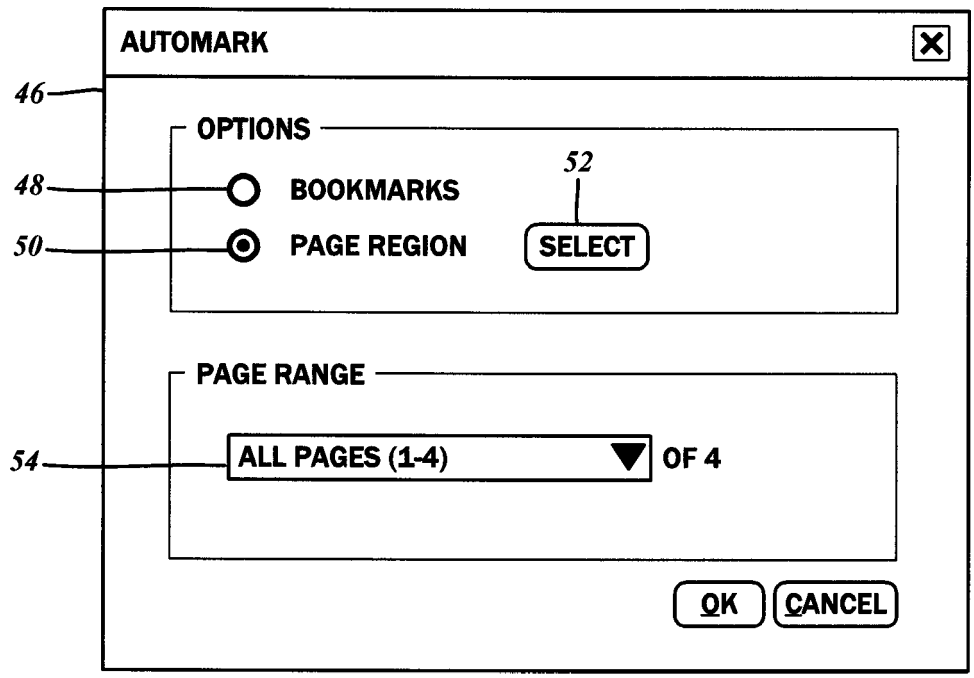
AUTOMARK
46
OPTIONS
48
BOOKMARKS
52
50
PAGE REGION
SELECT
PAGE RANGE
54
ALL PAGES (1-4)
OF 4
OK
CANCEL

FIG. 4

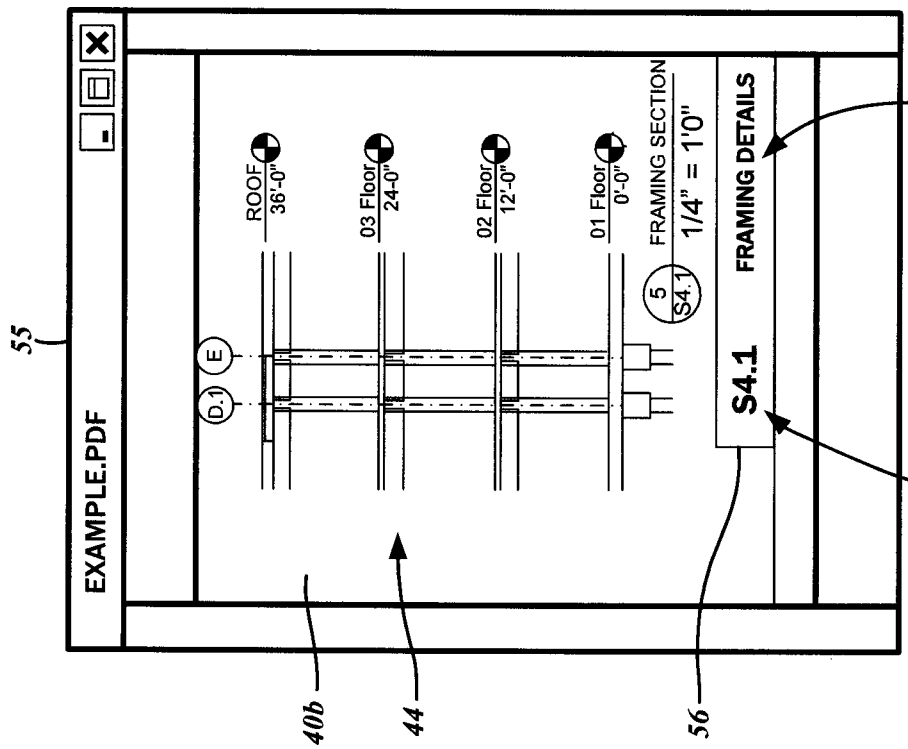
EXAMPLE.PDF
ROOF 36'-0"
03 Floor 24-0"
02 Floor 12'-0"
01 Floor 0'-0"
FRAMING SECTION
5 S4.1 1/4" = 1'0"
S4.1 FRAMING DETAILS
D.1
E
55
82
66
40b
44
56

FIG. 5B

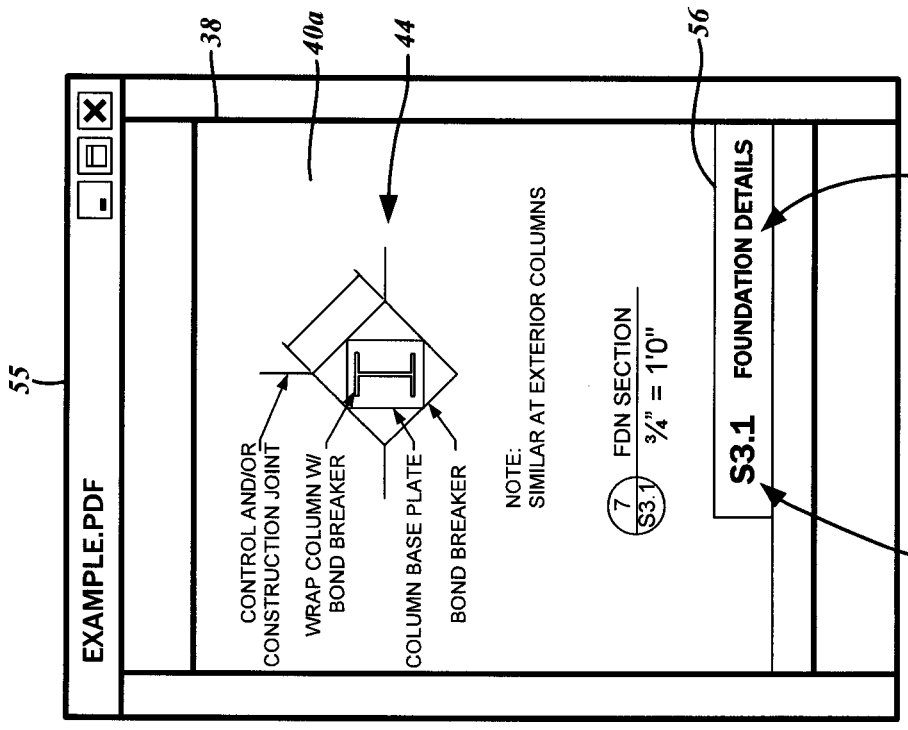
EXAMPLE.PDF
CONTROL AND/OR CONSTRUCTION JOINT
WRAP COLUMN W/ BOND BREAKER
COLUMN BASE PLATE
BOND BREAKER
NOTE: SIMILAR AT EXTERIOR COLUMNS
7 S3.1 FDN SECTION 3/4" = 1'0"
S3.1 FOUNDATION DETAILS
55
38
40a
44
56
78
64

FIG. 5A

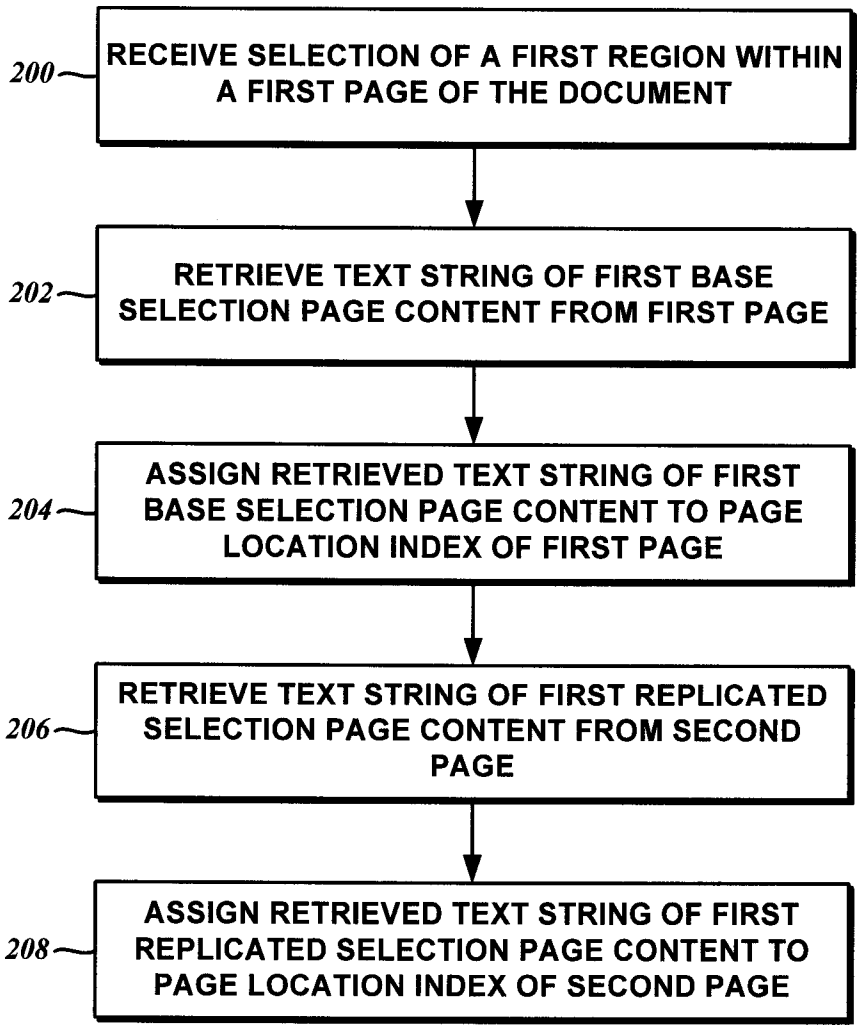
200
RECEIVE SELECTION OF A FIRST REGION WITHIN A FIRST PAGE OF THE DOCUMENT
202
RETRIEVE TEXT STRING OF FIRST BASE SELECTION PAGE CONTENT FROM FIRST PAGE
204
ASSIGN RETRIEVED TEXT STRING OF FIRST BASE SELECTION PAGE CONTENT TO PAGE LOCATION INDEX OF FIRST PAGE
206
RETRIEVE TEXT STRING OF FIRST REPLICATED SELECTION PAGE CONTENT FROM SECOND PAGE
208
ASSIGN RETRIEVED TEXT STRING OF FIRST REPLICATED SELECTION PAGE CONTENT TO PAGE LOCATION INDEX OF SECOND PAGE

FIG. 6

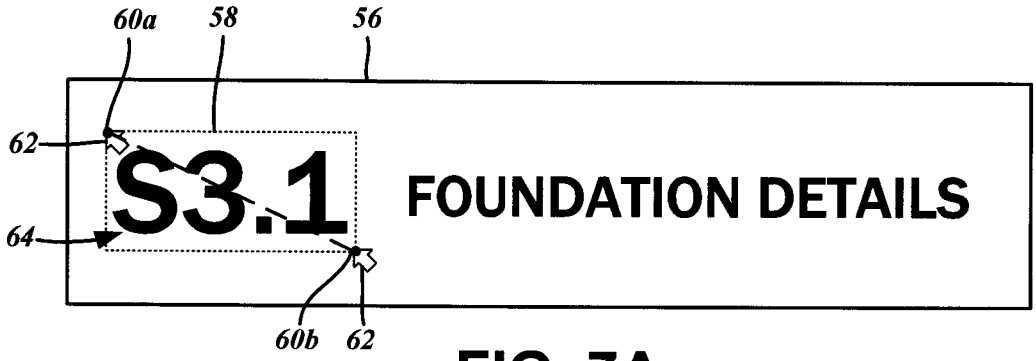

FIG. 7A

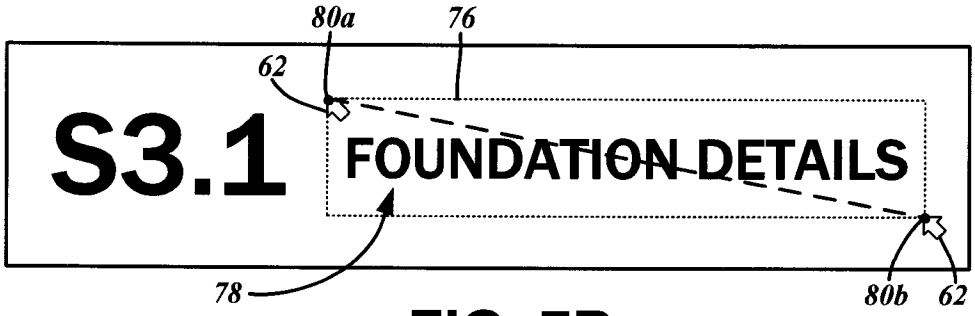

FIG. 7B

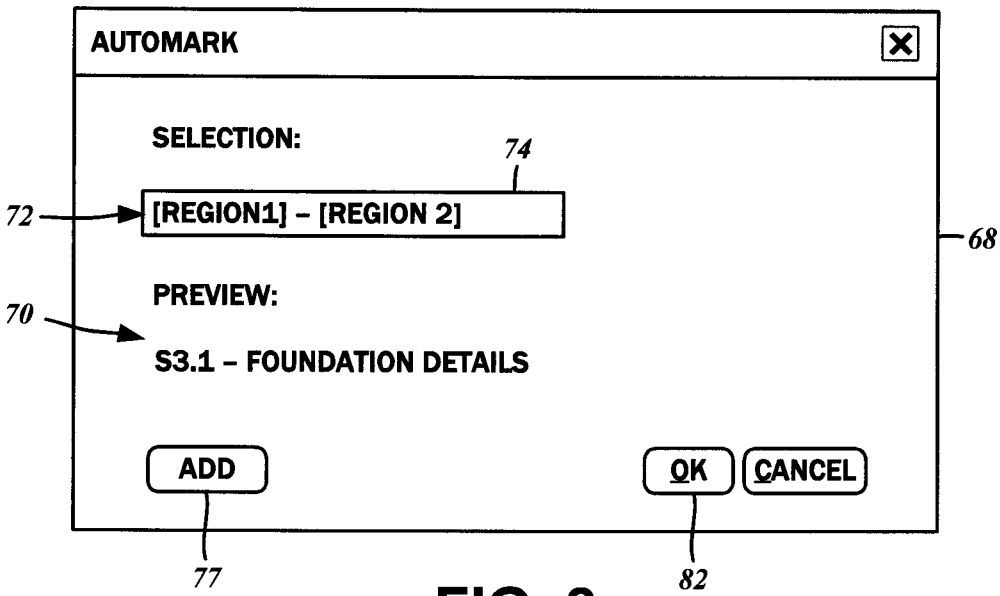

FIG. 8