

consolidated markup document. In step 402, the routine continues until all paragraphs of the original document are checked and the routine finishes in step 404.

The process of comparing one paragraph against another paragraph to see if they are the same or similar enough that we consider them the same paragraph from the two documents is illustrated in FIGS. 5A–5C. In the following discussion, it will be assumed that Paragraph 1 is a paragraph from the original document, and Paragraph 2 is a paragraph from an edited copy.

The routine begins in step 500 and proceeds to steps 502 and 504 in which the next words in paragraphs 1 and 2 are retrieved. Text followed by a white space is considered a “word” for these retrievals. These retrieved words are then compared in step 506 and a match is checked in step 510. To perform the comparison the routine only examines textual characters; font attributes, which simply make the words look different, are ignored. If a match is detected a match count is incremented in step 512.

In step 514, the routine counts how many words have matched. If four consecutive words match, then the paragraphs are similar enough to count them as a match for the previously described document comparison. In this case, a match is declared in step 518. The routine continues on to determine whether any differences exist between the words in paragraph 1 and paragraph 2. If differences exist they are marked in paragraph 2 as indicated in step 519. The routine finishes in step 522.

If, in step 510, it is determined that words that don’t match before four consecutive words are detected, then the match count is reset in step 508 and the words being considered in Paragraph 1 and Paragraph 2 are marked in step 516.

In order to continue, the routine uses the same comparison process as previously used for paragraphs, but instead compares words. In particular, the routine proceeds, via off-page connectors 520 and 524, to step 526 where the marked word in Paragraph 1 is compared to each word in Paragraph 2 in order to get “re-synched”. If a match is detected in step 528, the routine also must detect at least four consecutive words that match. This process is illustrated in steps 530 and 532. In step 530, upon the detection of a match, a match count is incremented. If four consecutive words match as determined in step 532, then the routine proceeds, via off-page connectors 533 and 539 to step 544 where a paragraph match is declared. The routine continues searching for, and marking, differences in the two paragraphs until the end of the paragraph is reached in step 548. The routine then ends in step 546.

Alternatively, if the comparisons performed in step 526 do not yield four consecutive words that match, then the routine proceeds to step 529 in which a check is made to determine if the end of the paragraph has been reached before four consecutive matches are found. If the last word in the paragraph has not been examined, the match count is reset in step 531 and the routine returns to step 526 to continue checking for four consecutive matches.

If the end of the paragraph is reached before four consecutive matches are detected as determined in step 529, the routine proceeds, via off-page connectors 535 and 537 to step 534 in which the marked word in Paragraph 2 is compared to each word in Paragraph 1. The presence of four consecutive words that match is detected in steps 538 and 540. If four consecutive words match, then the routine proceeds to step 544 where a paragraph match is declared, further differences are detected and marked in step 548 and the routine ends in step 546.

Alternatively, if four consecutive words that match are not detected, the routine checks for the end of the paragraph in step 541. If the end of the paragraph has not been reached, then the match count is reset in step 543 and the routine returns to step 534 to continue checking for four consecutive matches.

Alternatively, if the end of the paragraph has been reached as determined in step 541 before four consecutive word matches have been found, then non-matching paragraphs are declared in step 542 and the routine finishes in step 546.

If the routine in FIGS. 5A–5C determines that Paragraphs 1 and 2 are similar enough for a match, then a separate routine (not shown) examines Paragraph 1 and Paragraph 2 looking for embedded objects. For example, these objects can consist of a table embedded in a paragraph or a text frame that contains some text. The routine which examines the paragraphs for embedded objects is only performed if it is determined that the original document and the edited copy have the same “heritage”. or the edited copy is a continuation of the original document. Heritage is determined by storing, at the time of creation, a creator ID for the creator of the document and a timestamp indicating when the document was created. If both the creator ID and the timestamp match for the original document and the edited copy, then the documents are checked for embedded objects.

If embedded objects are discovered, the object ID for each embedded object is checked. If two embedded objects have the same timestamp, then they are the same embedded objects in the original and edited copy documents. The contents of the two embedded objects are then compared using the same process as described with paragraphs. In particular, the stream of paragraphs within the frame or table are compared in the same manner as a stream of paragraphs in two documents are compared as illustrated in FIGS. 4A–4C.

Although only one embodiment of the invention has been disclosed, it will be apparent to those skilled in the art that various changes and modifications can be made which will achieve some of the advantages of the invention without departing from the spirit and scope of the invention. For example, alternative comparison routines could be used to compare paragraphs. These and other obvious modifications are intended to be covered by the appended claims.

What is claimed is:

1. An apparatus for producing a formal document from a first document and a second document in response to user commands, the apparatus comprising:

means responsive to a user command for comparing the first document with the second document in order to identify a first section of the first document which contains unedited portions identical to portions of a second section of the second document;

means responsive to the comparison of the first and second documents for creating a consolidated document which displays both edited and unedited portions of each of the first and the second sections;

means responsive to the creation of the consolidated document for determining changes between the first section and the second section;

editing means controlled by user commands for selecting desired changes to be made in the first section from the determined changes; and

means responsive to a user command for forming a final document which includes the unedited portions of the first section and the selected desired changes.

2. The apparatus according to claim 1 wherein the editing means comprises: