

Once the preformatted text tags are deleted (step 312) or the preformatted text is revised according to a current user preference (step 314), the process passes to step 316, which illustrates a determination of whether all preformatted text within the HTML document has been examined. This determination may be made, for example, by determining whether additional preformatted text tag pairs are present which delimit text which has not yet been examined.

If additional preformatted text within the HTML document remains to be examined, the process returns to step 308 to scan the text between an additional pair of preformatted text tags. If no additional preformatted text requires examination, however, the process proceeds to step 318, which illustrates passing the (possibly revised) document to the application utilized to format the document for display, as described above. During formatting of the text for display, reflowable text is divided into lines having an allowed line length (that is, a length which will fit in an allowed display width).

Referring to FIG. 3B, various special formatting indicia for determining whether preformatted text may be appropriately reflowed in accordance with a preferred embodiment of the present invention are illustrated. Determinations of whether such special formatting indicia are present, steps which may be readily incorporated into the process depicted in FIG. 3A, are illustrated. Step 310a depicts determining whether the preformatted text includes multiple successive tabs. Steps 310b and 310c depict determining whether the text includes multiple successive blank spaces which grammatically unnecessary—that is, multiple successive blank spaces which are not situated at the end of a line of text, at the beginning of a line, or following certain specific characters, such as periods, exclamation points, question marks, or colons.

Step 310d depicts a determination of whether the preformatted text includes special characters indicative of a table or other two-dimensional text device. The special characters which are utilized for this determination are typically those which are not letters, numerals, or standard punctuation (periods, commas, exclamation points, question marks, colons, semi-colons, hyphens, dashes, parentheses, and single or double quotation marks). The special characters which are the focus of the determination may include vertical bars |, multiple successive hyphens (-) or dashes (--), multiple successive underscores (___), plus signs (+), or box characters (e.g., □, ▤, ▥, ▦, ▧, etc.). A variant of this step determines whether similar special characters appears on multiple successive lines, which is especially indicative of a table or other two-dimensional structure. For example, multiple successive lines containing vertical bars suggests that the preformatted text includes a box or table. Another variant seeks to determine whether blank spaces, periods, or the beginning or ending points of words align vertically in multiple successive lines, which is also indicative of a table.

The special formatting indicia described above are merely exemplary. Other special formatting indicia or criteria may be identified by detailed analysis of the HTML documents most likely to require automatic conversion of preformatted text for a specific application. Such other special formatting indicia are considered to fall within the scope and spirit of the present invention.

In the context of more general compound documents, other appropriate indicia of special formatting may also be identified, depending on the particular approach used to create the compound document. Tags identifying tables or graphical image data in post-script format, for example, may

be found in some compound documents (including documents consistent with later versions of HTML). Other tags may indicate a stylized formatting of the text which the author or publisher wishes to be preserved, as in poetry, advertisements, or promotional materials. Such indicia may be utilized within the present invention in the manner described.

The present invention allows preformatted text tags to be stripped from an HTML document where the nature of the text is such that the text may be appropriately reflowed. Text contained within preformatted text tag pairs is examined to determine if any special formatting indicia—elements indicating that the preformatted text is a table or other two-dimensional device which should not be reflowed—are present. Where reflowing the preformatted text is appropriate, preformatted text tags may be eliminated. This allows the text to be displayed in a more readable manner and eliminates the need for users to scroll right and left for each line of text.

Although described in the context of automatically converting preformatted text for display on set-top box data processing systems, the present invention may be employed in a variety of other circumstances. For example, an application providing display windows within the overall display where users may modify the size of each display window may also benefit from the present invention.

A variety of other applications may benefit from automatically reflowing text where appropriate. For example, other markup languages may provide tags for protecting the formatting of a particular block of text. The present invention may be employed within applications compatible to such markup languages, employing the protective tags as indicia of special formatting. Alternatively, a USENET newsreader may display postings in a font which does not permit lines of text within the posting to fill or fit evenly within the display area. In such a circumstance, reflow of certain portions of the posting, such as the posting header or a signature file appended to the end, may be inappropriate. The present invention may be employed within the newsreader to selectively reflow text as appropriate.

It is important to note that while the present invention has been described in the context of a fully functional data processing system, those skilled in the art will appreciate that the mechanism of the present invention is capable of being distributed in the form of a computer readable medium of instructions in a variety of forms, and that the present invention applies equally regardless of the particular type of signal bearing media used to actually carry out the distribution. Examples of computer readable media include: recordable type media such as floppy disks and CD-ROMs and transmission type media such as digital and analog communication links.

While the invention has been particularly shown and described with reference to a preferred embodiment, it will be understood by those skilled in the art that various changes in form and detail may be made therein without departing from the spirit and scope of the invention.

What is claimed is:

1. A method of automatically converting preformatted text to reflowable text, comprising:
 - determining whether text includes preformatted text tags defining line formatting for text delimited by the preformatted text tags;
 - responsive to determining that the text includes preformatted text tags defining line formatting for text delimited by the preformatted text tags, determining whether