



5

ABSTRACT OF THE DISCLOSURE

Method and apparatus for storing document images, for creating a retrieval index by which the document images may be retrieved, and for displaying the retrieved document images. When a document image is obtained, the document image is subjected to rule-based block selection techniques whereby individual regions within the document region are identified, and the types of regions are also identified, such as title-type regions, text-type regions, line art-type regions, halftone-type regions and color image-type regions. The identification is used to create structural information and both the document image and the structural information is stored. A word-based retrieval index is created based on title-type regions and/or text-type regions, the retrieval index being used in conjunction with a search query so as to be able to retrieve documents which match the search query. The retrieved documents are displayed in either a full image mode or a rapid browsing mode. In the rapid browsing mode, the full image of the document is not displayed, but rather only an abstract structural view of the document image based on the stored structural information. The level of abstraction may be specified by the operator in connection with the identified structural regions of the document, whereby, for example, only a structural view is displayed, or only title-type regions are displayed mixed with remaining structure.

E-2898