

AUTOMATIC OPTICAL CHARACTER RECOGNITION (OCR) CORRECTION

ABSTRACT

An Optical Character Recognition (OCR) system, including: an acquisition device configured to obtain a digital image of a physical document; an image conversion device configured to convert the digital image of the physical document into corresponding machine-readable text; a correction device configured to: evaluate the machine-readable text using a trained Long short-term memory (LSTM) neural network language model to determine whether correction to the machine-readable text is required; if correction to the machine-readable text is required, determine a most similar text relative to the machine-readable text from a name and address corpus using a modified edit distance technique; and correct the machine-readable text with the determined most similar text; and an output device configured to output the corrected machine-readable text.

Figure 2