Abstract

Cloud computing and Optical Character Recognition (OCR) technology has become an extremely attractive area of research over the last few years. Sometimes it is difficult to retrieve text from the image because of different size, style, orientation and complex background of image. There is need to convert paper books and documents into text. OCR is still imperfect as it occasionally mis-recognizes letters and falsely identifies scanned text, leading to misspellings and linguistics errors in the OCR output text. A cloud based Optical Character Recognition Technology was used. This was powered on Microsoft Windows Azure in form of a Web Application Programming Interface that load images to an Optical Character Recognition server, process with necessary recognition, export parameters, and obtain the results of the processing. The key idea is to bring together the advantages of the Optical Character Recognition technology and cloud computing in one place in other to enable quicker access and faster turn out time, processing period, and increasing efficiency across the board for application users. The methodology adopted is the object oriented methodology. This was achieved using JAVA programming language.
References

8. Malakar Samir (2012)."Text line extraction from handwritten document pages using spiral run length smearing algorithm". IEEE international conference on communications, Devices and intelligent system (CODIS)


Index Terms

Computer Science  Information Sciences

Keywords

Cloud computing, optical character recognition (OCR), Microsoft Window Azure, image, text, ABBYY.