

Name :

Class/No:

Change the verb in brackets into the correct form!

Optical character reader (OCR) (1) _____ (be) the electronic conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document, a photo of a document, or a scene-photo.

Widely used as a form of data entry from printed paper data records, it (2) _____ (be) a common method of digitizing printed texts so that they can be electronically (3) _____ (edit), (4) _____ (search), (5) _____ (store) more compactly, (6) _____ (display) on-line, and (7) _____ (use) in machine processes such as cognitive computing, machine translation, (8) _____ (extract) text-to-speech, key data and text mining. OCR (9) _____ (be) a field of research in pattern recognition, artificial intelligence and computer vision.

Early versions (10) _____ (need) to be trained with images of each character, and (11) _____ (work) on one font at a time. Advanced systems (12) _____ (capable) of producing a high degree of recognition accuracy for most fonts (13) _____ (be) now common, and with support for a variety of digital image file format inputs. Some systems (14) _____ (be) capable of reproducing formatted output that closely (15) _____ (approximate) the original page including images, columns, and other non-textual components.

OCR engines have been developed into many kinds of domain-specific OCR applications, such as receipt OCR, invoice OCR, check OCR and legal billing document OCR.