(Xiao and Huai 2017)

Xiao, Hong [肖红], and Yan Huai [槐燕]. 2017. “An Analysis of Quality Checking Problems in the Practice of Digitization of Newspapers of Republic of China” [民国报纸数字化实践中的质检问题探析]. Researches on Library Science [图书馆学研究], no. 07: 61-78+87.

DOI:10.15941/j.cnki.issn1001-0424.2017.07.010

Questions:

Which OCR engine used

National Library started digitizing newspapers from the Republic of China in 2014

Image scan of each page, OCR of article headings and titles, headings and titles marked with boxes (for title search)

Quality control: number and naming of documents, the resolution 分辨率, size, and bit depth 位深 of scanned images and other technical parameters, and the recording of OCR character recognition the content of the title in XML text file.