Articles on proofing and verification of indexing newspaper contents

(Zhang 2019; Xiao and Huai 2017; Xiao 2017)

Xiao, Hong [肖红]. 2017. “The Main Problems and Solutions in the Practice of Digitization of Newspapers in the Republic of China” [民国报纸数字化实践中的主要问题及处理策略]. *Researches on Library Science* [图书馆学研究] 2017 (4): 22–37. https://doi.org/10.15941/j.cnki.issn1001-0424.2017.04.004.

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.

Zhang Wei [张玮]. 2019. “Research on the Common Problems in Digital Acceptance of Newspapers in the Period of Republic of China: A Case Study of National Library of China” [民国报纸数字化验收常见问题研究——以国家图书馆为例]. *Library and Information Studies* [图书情报研究], no. 3: 72–79.

Digitization based on image scans

Provided by library or also digitized by external agencies, or did they digitize from MF rolls?

Outsourcing to structure scans, edit newspaper and issue metadata, and analyse content by page: draw boxes around columns, titles, authors, texts, identify parts of an article, record text

Text “OCR” - > manually or with software?

Staff for work mostly “junior high school and and high school students”

Problem of libraries: data output with many errors, outsourcing dominant way to digitize newspapers, many different libraries do it, therefore the articles as guideline

Different levels of errors

Suggestions for improvement (but automatical approaches e.g. to fulltext not mentioned)