(Zhang 2019)

Zhang Wei [张玮]. 2019. *“Minguo baozhi shuzihua yanshou changjian wenti yanjiu——yi guojia tushuguan wei li”* [民国报纸数字化验收常见问题研究——以国家图书馆为例] [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]. *Tushu qingbao yanjiu* [图书情报研究], no. 3: 72–79.

1 Introduction

Acid paper -> urgent need for a new storage method to save the newspapers of the Republic of China, therefore creation of databases based on characteristics, specialities, and commercial use

Sichuan Uni Lib selects by numbers of microfilms used by users, established a platform 《新新新

闻》"New New News" shared platform, searching by Gregorian and Republican calendar, main/subtitles etc

Chongqing Lib started 2012 scanning of Republican newspapers into pdf format

Since 2014 National lib. digitization of newspaper microfilms, based on image scans created article title index

digital publishing companies created large-scale full-text databases, problems such as uneven development of various types of documents and lack of databases for subject types

The digitization of newspapers in the Republic of China has become the primary way to solve newspaper problems

How to improve the efficiency and quality of digitization > Xiao Hong (2017), but the article has no detailed analysis and introduction on quality inspection

2 common problems

digitized products usually include metadata 元数据 and object data对象数据.

Metadata: in MARC format

Object data: TIFF files, publishes jpg, dual-layer pdf, xml

NLC follows standards, uses professional verification software 专业数字化验收软件 developed by Hanwang 汉王

2014-2018 completed 114 newspapers, ca. 130.000 issues, >700.000 pages

Metadata: categories, labels, etc

Item data: file name, image quality etc. coordinates of boxes, content descriptions recorded in XML file

Use of Hanwang software for proofing

Problems: columns, titles, nesting articles, bounding boxes, authors, continued articles, images/image groups

Reasons for problems

1 Personnel lacks experience, problems in coordination of workflow

most of the processing personnel recruited by outsourcing companies外包公司 are junior high school students and high school students大部分是初中生、高中生.

not systematically trained, limited understanding of the tasks

communication issues within the companies, between different “assembly lines”

2 communication between digitisation staff and proofing staff not in time. Changes in the library policies to not get smoothly communicated via company supervisors to digitization staff

3 batch processing leads to stress quantity instead of quality. Repetitive processing enforced from same personnel, but leads to more errors

4 external companies lack experiences and equipment, imperfect results, ocr errors, problems with character recognition are show stoppers

Recommendations:

1 choose between companies, take the best based on previous experiences/tasks

2 calculate error rates and give outcome back if quality does not match

3 processing staff needs to be trained on specifications, in special cases have 1:1 training to get direct communication

4 train proofing personnel, all need to be familiar with the whole process, conduct visits at the external companies to see problems immediately

5 use proof readers according to their best capabilities