

**Contributor/Presenter:**

Ding Xiao Wen, Library of Jiangsu Administration Institute

**Title:**

***PDF Converter Production of Historical Newspaper Digitization: the picture experience of China's DaChengLaoJiu Database***

**Abstract:**

China's DaChengLaoJiu database includes more than 6000 kinds of China's publishing newspapers and magazines, a total of 120000 issues and more than 1.3 million articles from the late Qing dynasty to 1949. These literature are full of historical, scientific and literary research value. The digitalization of these historical newspapers and periodicals can use the PDF format. There are two forms of digitalization technology, one is double, the other is refactoring.

In general, to carry out the project of digitization of the historical newspapers, scanning, character recognition, correction, indexing, storage, constructing full-text databases and search websites will be just fine if only considering from the angle of protection and archiving. But if considering more from application of media terminals (such as iPhone, tablet computer), and the development of other products, we should take the technology of refactoring the PDF.

On the principle of availability and being artistic, refactoring PDF uses images and text data to make the whole graphic mixed rearrangement according to the original layout structure. It is a single layer, and the factors taken into consideration are as follows: (1) PDF file structure;(2) PDF format rearrangement;(3) the visual browsing;(4) printing;(5) format retrieval and positioning;(6) storage capacity;(7) error rate of words;(8) distribution channels;(9).producing.