(Wu 2017)

Wu Fangzhi [吴方枝]. 2017. *‘Shuzi renwen beijing xia minguo wenxian de shuzihua yanjiu’* [数字人文背景下民国文献的数字化研究] [A Review of Literature Digitization of the Republic of China Period in Digital Humanities]. *Tushuguan xue yanjiu* [图书馆学研究], no. 15: 18-21.27. https://doi.org/10.15941/j.cnki.issn1001-0424.2017.15.004.

Since the implementation of the "Republic of China Document Protection Program" “民国时期文献保护计划” in January 2012, various information technologies have been used to process the Republic of China documents, and a series of Republic of China document databases have been developed, and the digitization of the Republic of China documents has achieved remarkable results

In the future we should fully learn from the successful practices in the field of digital humanities research and provide new methods and perspectives for the in-depth development and utilization of the literature of the Republic of China

1. Theory based on Unsworth definition of DH

2005 ADHO, then centers in Nam/EU like Harvard, Stanford, Kyoto, UCL (CASA at Bartlett), King’s College (Center for Computing in the Humanities)

2009 Taiwan Research Center for Digital Humanities, Taida, and HK-Taida\_Beida jointly organised “Forum on Spatially Integrated Humanities and Social Science”

2011 PRC Wuhan University Digital Humanities Research Center, 2016 Peking University Library and Harvard’s CBDB “Digital Humanities Forum”

1. DH Digitisation projects

Defines four levels of DH projects

* (1) data collection and database construction数据收集及数据库建设;
* (2) development and use of humanities data tools人文数据工具的开发和使用;
* (3) innovative humanistic research methods and adoption of research paradigms创新人文研究方法和研究范式的采用;
* (4) Creative destruction and construction in the field of humanities 人文领域的创造性破坏与建设

At present, most of the digital humanities practice in the digitization of literature in the Republic of China stays at the first level, that is, the level of digital humanities database or data set construction, and rarely involves higher levels. They mainly include the two aspects of construction: large-scale comprehensive full-text databases and special feature databases

2.1) Comprehensive full-text database projects

2. 1. 1 Public libraries 公共图书馆  
2. 1. 2 University Library  
2. 1. 3 Commercial database institutions商业数据库机构   
2.2) thematic characteristic database projects

2. 2. 1 Index database索引数据库

2. 2. 2 Thematic database for specific content针对特定内容的专题数据库   
2. 2. 3 Featured Resource Database特色资源数据库

3) development direction of the digitization

due to the lack of a unified construction plan缺乏统一的建设规划, there are **many problems in database construction**, including serious repeated construction重复建设现象严重, single digital resource platform数字资源平台单一, lack of deep development 缺乏深层开发, etc.

3.1 In-depth mining 深度挖掘

Text mining 文本挖掘

emphasis of digital humanities is no longer pure textualization and digitization, it is based on the deep mining and intelligent analysis of text

some systems developed in the ancient book digitization, poetry analysis, automatic sentence reading, Tang Poetry Analysis System

For the development of periodical newspapers in the Republic of China, we no longer only provide the full text content of the literature, but increase the depth and relevance of the literature resource disclosure 深度和关联性 on the basis of indexing

1. Literature 文学类 building a library of works of writers, relationships within the literature
2. Historical archives 历史档案类 extract specific information such as place names, names, historical events
3. Art 艺术类 clustered according to the type of work, era, genre, and creative form

3. 2 Display and extract knowledge 展示和知识提取

Data visualization数据可视化

"China Biographical Database (CBDB)" “中国历代人物传记资料库( CBDB) ”

Visualization technology可视化技术 is promising in the literature digitization of the Republic of China, but the current applications are mainly concentrated in two aspects:

1. Visual display of the literature text: same-screen comparison 同屏对照of search results, esp. commercial products
2. Knowledge extraction知识提取

In the future digitization of the literature of the Republic of China, we should make full use of visualization technology, develop more visualization projects, and introduce some visualization tools such as Gephi to present the literature of the Republic of China more intuitively.

3. 3 Construction of the historical geographic information system历史地理信息系统

Generally speaking, the application of GIS technology in the historical field can be broadly classified into three aspects:

* the digitization of displaying historical data展示历史资料的数字化,
* the management and visualization of spatial historical data空间历史数据的管理和可视化,
* and the exploration of historical processes based on spatial analysis methods基于空间分析方法探知历史过程

we can make full use of GIS technology and use CCTS results as examples to establish a historical and cultural geographic information system

provided researchers with a geographic data information platform, retrieval statistical tools and models 地理数据信息平台、检索统计工具和模型 to realize geospatial data processing, analysis and visualization, and it was convenient for the historical researchers of the Republic of China to use the platform to carry out relevant special studies.

simply using a certain technology generally cannot achieve satisfactory results

4. Conclusion

we should not rely too much on and strive for digital humanities 不能对数字人文太过依赖和追捧 It is the information technology 信息技术 and humanities research 人文研究 that should find the convergence point of the development

give play to the respective strengths of digital technology and humanities research, and jointly promote the development of the digitization of the Republic of China literature in a deeper and broader direction