From Digitization to Digital Humanities: The Development of Digital Humanities in Taiwan

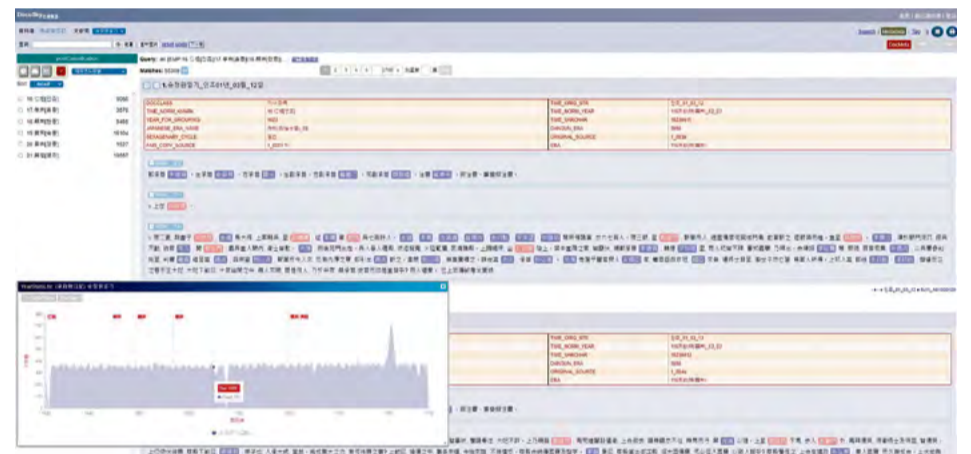Chijui Hu

The development of Digital Humanities in Taiwan is founded on decades of research in digital archives, including various works such as research tools, databases, and models. In the 1980s, several institutes and universities in Taiwan worked to achieve substantial progress in archives digitalization. From 2002 to 2012, the National Science Council of the Taiwan Government (NSC, which was renamed the Ministry of Science and Technology in 2014) conducted the National Digital Archives Project (NDAP).[1] This government support allowed many academic institutes, libraries, universities, and private institutions to digitize large amounts of archives, photos, scriptures, artifacts, maps, and video data. Based on these digital data, Digital Humanities researchers in Taiwan have been able to use these materials for many fields of research and innovation.

One of the achievements of NDAP was the construction of many databases in the early 21st century. However, these databases were built for institute-based researchers and experts, and so their main function is to retrieve and browse; their main principle was not only high precision but also high recall. As a result, users may find it difficult to identify the context of the data, merely ending up with many results from the database. Through an innovative process, a context discovery system was invented by the Research Center for Digital Humanities (RCDH) of National Taiwan University (NTU), which was led by Professor Jieh Hsiang. Using post-query classification methods, users can identify not only what was retrieved but also the inter-relationships among documents and the collective meanings of a sub-collection.[2] The basic assumption is that documents in a collection should have well-structured metadata, which is important for post-classification of a sub-collection. When the full text of the content is also available, more sophisticated analytical methods such as co-occurrence analysis can also be deployed.[3]

The context discovery system is, however, a closed database of sorts, making it difficult for users to add additional data or metadata into the database. In addition, the tools in the system were developed exclusively for the data of the system. This means the Digital Humanities tools are tied up with the system and that users cannot use them for their own data. This is why attempts are being made to move towards the development of a Digital Humanities platform. One of the goals of Digital Humanities platforms is to provide humanities researchers with the ability to integrate research material without the help of software engineers. Researchers can deal with the research data by themselves and upload it onto the platform; they can analyze, mark up, and reorganize metadata or produce statistics of that metadata; they can also visualize the uploaded data using the Digital Humanities tools embedded in the platform. The benefit of such a scheme is that researchers can save much time and effort, and the resources can be made much more accessible.[4]

There are several institutes in Taiwan committed to developing Digital Humanities platforms, examples being the DocuSky Collaboration Platform of RCDH,[5] the Digital Analysis System for Humanities (DASH) of the Academia Sinica Center for Digital Cultures,[6] and the CBETA Research Platform (CBETA RP) of Dharma Drum Institute of Liberal Arts.[7] CBETA RP is connected with the Buddhist digital canon of the Chinese Buddhist Electronic Texts Association (CBETA), which is a full-text database of high-quality Chinese Buddhist sutras. Not only is it possible to read and search sutras through the CBETA RP; it is also possible to analyze the terms of the result and present the results through different kinds of charts using the tools of the platform. DASH connects with several data repositories. Users can mark up the text and then calculate authority terms and N-gram statistics or conduct term co-occurrence

analysis; results are demonstrated through visualization tools in charts, word clouds, social analysis graphs, and maps. DocuSky was created by Dr. Hsieh-Chang Tu of NTU, and it is managed by Professor Jieh Hsiang. With the core format called DocuXml, many converters in DocuSky can be used to convert different format data into DocuXml. In other words, .txt, .xlsx, MARKUS[8] tagging files, or the text from many repositories (e.g., CBETA, CTEXT,[9] KANRIPO,[10] RISE,[11] or Wikisource[12]) can be converted into DocuXml with a couple of clicks. DocuXml can be used to build a personal database in DocuSky, and researchers can use the tools in DocuSky for tagging, metadata managing, adding relationship information for social network analysis, or even creating GIS layers using the webGIS tool called DocuGIS. The DocuXml upload in DocuSky lets users undertake analysis through the post-classification function of metadata and tags. Moreover, authors of DocuXml files can authorize RCDH to make their databases public through the DocuSky model to share their achievements with the world. From context discovery systems to a personal Digital Humanities platform, the main purpose is to make the connection between tools and databases more flexible and to allow users to operate and use the database designed by themselves more freely.

The formation of the Digital Humanities community in Taiwan was spearheaded by RCDH, which held the 1st International Conference of Digital Archives and Digital Humanities (DADH) in 2009. Since then, DADH has become an annual meeting for Digital Humanities scholars from Taiwan and overseas. With the increasing number of scholars studying Digital Humanities in Taiwan, the Taiwanese Association for Digital Humanities (TADH)[13] was formed in 2016. TADH, which formally became the organizer of the DADH annual meeting in 2016, has grown to become an important organization for the study of Digital Humanities in Taiwan. TADH also became a constituent organization

of the Alliance of Digital Humanities Organizations (ADHO) in 2018, officially joining the international Digital Humanities research community as a research partner. Moreover, the *Journal of Digital Archives and Digital Humanities*, published by TADH, has become a place for Digital Humanities scholars to publish their research. Through the association, annual meeting, and journal, the development of Digital Humanities research in Taiwan is expected to become more plentiful and diverse in the future.

Chijui Hu, Assistant Professor, National Changhua University of Education.

Notes

1 In 2008, NDAP changed its name to the Taiwan e-Learning and Digital Archives Program (TELDAP).

2 Szu-Pei Chen, Jieh Hsiang, Hsieh-Chang Tu, & Micha Wu, "On building a full-text digital library of Historical Documents, ICADL 2007 (Hanoi, Vietnam, 10-13 Dec. 2007)," *LNCS* 4822, pp. 49-60.

3 Hsiang Jieh, "Context discovery in historical documents – a case study with Taiwan History Digital Library (THDL)," July 16-22, Digital Humanities 2012, University of Hamburg.

4 Chijui Hu, "DOCUSKY AND DIGITAL HUMANITIES RESEARCH," Taiwan Insight. Available at https://taiwaninsight.org/2020/07/10/docusky-and-digital-humanities-research/ (Accessed March 10, 2022)

5 https://docusky.org.tw/

6 https://dh.ascdc.sinica.edu.tw/

7 https://cbeta-rp.dila.edu.tw/

8 Hou Ieong Brent Ho, and Hilde De Weerdt. MARKUS. Text Analysis and Reading Platform. 2014- http://dh.chinese-empires.eu/beta/ Funded by the European Research Council and the Digging into Data Challenge.

9 https://ctext.org/zh

10 https://www.kanripo.org/

11 https://rise.mpiwg-berlin.mpg.de/

12 https://wikisource.org/wiki/Main_Page

13 http://tadh.org.tw/en/