

The Origins and Current State of Digitization of Humanities in Japan

Muller, A. Charles

acmuller@jj.em-net.ne.jp
University of Tokyo

Hachimura, Kōzaburō

Ritsumeikan University

Hara, Shoichiro

Kyoto University

Ogiso, Toshinobu

National Institute for Japanese Language and Linguistics

Aida, Mitsuru

National Institute for Japanese Literature

Yasuoka, Koichi

Kyoto University

Akama, Ryo

Ritsumeikan University

Shimoda, Masahiro

University of Tokyo

Tabata, Tomoji

University of Osaka

Nagasaki, Kiyonori

International Institute for Digital Humanities

Digital Humanities in Japan has been in progress since an early period. Recently, due to the spread and the development of advanced digital environments, individual humanities researchers are coming to use digital materials in various forms, and according to the continually growing needs of users, cooperation and organization between projects has steadily increased. However, there is some extent to which the large framework known Humanities Computing in Japan has lagged behind in its efforts to develop cooperation with similar projects overseas. Therefore, this panel aims to take a step in the right direction by introducing the origins and current state of Humanities Computing in Japan on digitization by featuring the reflections of the representatives of the

projects and organizations that have worked in this area from a relatively early time.

Humanities digitization projects in Japan, being developed by various research centers and research organizations, have already garnered over 3.5 billion yen in the form of known large-scale grants which were funded by Japanese Ministry of Education, Culture, Sports, Science and Technology. This does not include the numerous small grants that have been received for various digitization projects by individual researchers and small groups and budgets which were assigned by each organization itself which was promoting such projects. In order to provide a venue for the presentation and publication of the results of these funded projects, a number of groups and associations engaged in this work were established. One of the more prominent is Special Interest Group for Computers and Humanities (SIG-CH), which was established in 1989 under the auspices of the Information Processing Society of Japan which is the largest society of informatics in Japan. SIG-CH has served as the major organ for communication among researchers interested in these projects.

Kōzaburō Hachimura, Ritsumeikan University, representing SIG-CH

The meetings of SIG-CH have been held on a regular basis, about four times a year since 1989. At each meeting approximately 8 research papers are presented, with the proceedings being published as the “IPSJ SIG Notes” series. Up to the present, we have held 84 meetings, which have included a total of 720 paper presentations. The group consists of over 200 researchers belonging to the academic organizations of informatics or humanities.

Here, processing, analysis or mining of texts, images, digital archiving of texts, bibliographies or other digitized materials, and especially 3-D motion capture, etc. are the major themes. The technique, tools or study results in which they are applied etc. are presented. In recent years, study results that use GIS have been increasing in number. The various Humanities fields in Japan are represented by literature, linguistics, history, archeology,

museum studies, anthropology, dance studies, and Buddhist Studies, etc. In earlier periods, system-oriented thought was dominant, but recently, presentations tend to be characterized by an increase in content orientation, as well as on local and international cooperation between projects.

Toshinobu Ogiso, representing the National Institute for Japanese Language and Linguistics NINJAL)

NINJAL, established in 1948, has created various Japanese corpora. One example is the "*Taiyō* Corpus," which is the first major Humanities database in Japan created using XML. It is a tag-structure rendition of the 19-20th century magazine *Taiyō*, in which tags created for the purpose of linguistic research applied to about 14.5 million text characters. *Taiyō* was a typical magazine in Japan during the period of its publication from 1895 to 1925, and thus is an invaluable resource for understanding the foundations of the modern Japanese language which were formed during that period.

Also underway at this institute is the KOTONOHA plan, which seeks to integrate various corpora (including the *Taiyō* corpus). One part of this effort is the presently-underway project of the 'Balanced Contemporary Corpus of Written Japanese' (BCCWJ), containing 100 million words. In addition to this, the construction of a corpus is planned aimed at compiling premodern data. This project must especially address the peculiarities of written Japanese, which does not include spaces between words, and includes Chinese characters, *hiragana*, and *katakana*, making it very difficult to indicate word information with pauses between phrases, parts of speech, etc.

Mitsuru Aida, representing the National Institute for Japanese Literature (NIJL)

NIJL was established in 1972, making one of the earliest efforts to digitize Japanese literature. Researchers there have worked at converting the research information into database format, and in its inclusion of words, text descriptions, and literary indexes, has become Japan's prototype textual research database. In the early 1990's, NIJL defined an original standard for tagging Japanese literatures based on SGML, and upon this built a large full-text database. It has played a major role as the mechanism for a general database of the human culture research.

Koichi Yasuoka, Kyoto University, Institute for Humanities Research (Jinbunken)

In 1980 the Jinbunken began the digitization of the *Ming Dynasty Civil Examination Index*. The following year, the institute initiated the digitization of the *Index of Shanwen Liyi* and *Catalog for the Study of East Asian Documents*. Moreover, the Institute has held an Annual Workshop for Oriental Studies Computing (ORICOM) every year in 1990. The research conducted here for the past 20 years extends to many areas, including multilingual text processing, character-code issues, digital catalogs, and GIS. Most notably, the Jinbunken has been actively engaged in critique and development of Japanese *kanji* character sets, dealing with issues concerning the relationship between ancient characters, JIS X 0213, Unicode and so forth.

Ryo Akama, representing the Ritsumeikan University Art Research Center (ARC)

The extensive works of ARC have their origins in the digitization of the *kabuki* material that Prof. Akama initiated in the Waseda University Theater Museum in 1988. This approach led

to the establishment of the ARC in 1998. The ARC has worked on the digitization of various material and intangible Japanese cultural treasures through the aid of the Ministry of Education. The research is carried out making good use of various information technology skills, which include not only text and images, but also 3D images and motion capture, etc. At present, they have received a Global COE grant from the Ministry of Education, whereby they are serving as a base for Digital Humanities intended for Japanese culture as a whole.

Masahiro Shimoda, representing The University of Tokyo Center for Evolving Humanities (CEH)

One of the major aims of the CEH is, along with the development of its own Buddhist texts information system, to bring about cooperation with a wide range of digital projects related to the study of Indian Philosophy and Buddhism, and in so doing, to demonstrate a solid example of the possibilities of digital humanities studies to the Japanese academic world. The project includes the development of an extensive bibliographical database for the field of Buddhist Studies in Japan, which now includes about 70,000 entries. In addition, the Daizōkyō Text Database Research Committee, established in 1994, has completed and released a set of highly accurate text data of 600MB, covering the major portion of the East Asian Buddhist canon. This is known as the SAT Database, which is now fully interactive with the above mentioned article database, as well as the online reference work, the *Digital Dictionary of Buddhism*. The project is working toward further expansion of cooperation with other data bases in the field.

Conclusion

The projects represented here constitute only a very small portion of what is going on in Japan. For example, on the educational front, the Faculty of Culture and Information Science of Doshisha University, established in 2005, is aiming to teach methods of research for

the analysis of cultural information at both the undergraduate and graduate levels, and in the future is expected to be a major source for the development of talented researchers in the digitization of humanities. A few decades have now passed for Humanities Computing in Japan, and a wealth of data, tools, and techniques have been produced as a result. We are convinced that this panel will be an important step for locating digitization of humanities in Japan in the flow of worldwide Digital Humanities, and in charting our own future.