

Buddha Ngram Viewer: a Ngram Visualization Tool of Chinese Buddhist Translations

Jen Jou
HUNG

Dharma Drum Buddhist College

Over the years scholars have leveraged traditional text-critical methods to research change in Chinese language to the extent that this is documented by the corpus of canonical and paracanonical Buddhist texts, as well as outside this corpus. Philologically-trained Buddhologists, on the other hand, have similarly exploited the methodology of textual studies and criticism (including doctrinal, literary and linguistic aspects) in order to ascertain, e.g. ,the texts’ authorship, dating, authenticity etc. Yet traditional philology has its scale limit and limitations in feasibility are becoming more evident. Digital philology as a branch of digital humanities, with its application to Buddhist materials, is now opening new research horizons. As of today, the first 55 volumes of the Taishō edition of the Chinese Buddhist canon are freely and publicly available digitally through the efforts of the Chinese Buddhist Electronic Texts Association (CBETA). However, the potential of information technology applied to this field of research has yet to be recognized, let alone developed and be introduced as common practice. With the help of computer information technology, we are now able to apply statistical methods or artificial intelligence algorithms to the analysis of this corpus of digitized texts. The performance of quantitative analysis on digital texts still requires high level skills in computer programming and advanced statistical knowledge, which creates a high barrier for scholars in the humanities who are now attempting to navigate these tools.

Therefore, the present project aims at assisting humanities scholars in overcoming the major hindrances to progress in their active use of computational linguistic resources and information technology in general. We propose to employ ‘Google Ngram Viewer’ as the key reference model, and to integrate it with the textual content of CBETA digital canon, in order to develop a new online tool called the “Buddha Ngram Viewer”. This tool will allow users visualize the over time occurrences of inputted phrases in Chinese Buddhist texts through a simple procedure of a few steps. Further, we will make the Ngram statistic data set generated on the basis of the CBETA data set freely available for download. We believe that the “Buddha Ngram Viewer” tool and the Ngram statistic data set will not only enable scholars to discover important language features of Buddhist texts, but it will also encourage them to apply digital resources and quantitative research method in their research.