

The cover features a world map background with various academic fields labeled: 'Asian Languages', 'European Languages', 'English', 'Chinese', 'Language', 'Western Culture', 'Science', 'Technology', and 'Ecology'. A large white 'X' shape is overlaid on the map. The text 'FACULTY OF HUMANITIES NEWSLETTER 2015 ISSUE NO. 2' is prominently displayed in the upper right.

# FACULTY OF HUMANITIES NEWSLETTER 2015 ISSUE NO. 2



Prof. HUANG Chu-Ren, Department of Chinese and Bilingual Studies

**From classical poetry to modern ontology: Bridging the knowledge divide with a linked data approach. Workshop on Digital humanities of Chinese during International Conference on Corpus Linguistics and Technology Advancement (CoLTA 2015). Department of Linguistics and Modern Language Studies, The Hong Kong Institute of Education, Hong Kong, 16-18 December 2015.**

**Abstract:**

One of the biggest challenges in digital humanities is to extract, formulate, and represent knowledge from textual database in a way that is relevant and accessible for other scientific disciplines. This challenge is critical for historical texts as these are often one of the few tangible sources for us to extract our knowledge from our past. However, the historical texts also present the widest divide to bridge because of the difficulty in interpreting the text is compounded by the lack of full and explicit representation of knowledge and conceptual systems of the specific historical era. We report three studies to bridge this divide with the linked data approach. In particular, we take the ontology-lexicon interface approach (OntoLex, Huang et al. 2010a), where a lexicon extracted from text can be conceptually organized and map to modern ontology as knowledge representation. The ontological representation then provide full digital accessibility. In terms of system of knowledge to be extracted, we propose the Shakespearean-garden approach where initial focus was on flora and fauna. These approach allows us to map attested and described flora and fauna to modern ontology, as well as to reconstruct period specific domain ontologies based on the text. Three studies are reported: a pilot based on the small sample of classical collection of 300 Tang poems, an single author ontology based on work by Su Shi, and a full scale taxonomy-to-ontology mapping based on the encyclopedia of plant terms in Chinese classics GuangQunFangPu. In addition to the original textual databases, we also rely on the Academia Sinica Bilingual Ontological Wordnet (Sinica BOW, Huang et al 2010b) and the WordNet-SUMO mapping (Pease and Fellbaum 2010) to link out data. We show that the OntoLex approach to linking classical data to modern ontology is indeed an effective and rewarding way to bridge the knowledge divide of classical poetry and modern sciences.