



Hack Away II at Tibetan Buddhist Texts

Hackathon in the Galilee

February 5-9, 2017
Kibbutz Inbar, Israel

Organizers

- Orna Almogi (Universität Hamburg, Indian & Tibetan Studies)
- Nachum Dershowitz (Tel Aviv University, School of Computer Science)
- Dorji Wangchuk (Universität Hamburg, Asia-Africa Institute)
- Lior Wolf (Tel Aviv University, School of Computer Science)

The objective of this second hackathon is to further develop algorithmic tools that will facilitate the comprehensive study of canonical Buddhist texts, including scriptures, commentaries, and related treatises. The specific corpus of interest is the Tibetan Buddhist canon, consisting mostly of Indic Buddhist literature in Tibetan translation. This corpus was formed over a period of more than a thousand years and contains various layers of materials; its translation into Tibetan spanned several centuries.

We will probe methods of finding inexact quotations and borrowed texts within the corpus, analyzing influences among texts, and study the evolution and emergence of individual texts. We will work on

classification and clustering of texts, with the aim of allowing better understanding of the processes of translation, revision, and editorial activities.

We will also work on images of manuscripts and xylographs, including their paleographic analysis and character recognition.

The importance of the planned tools lies first and foremost in that it will allow scholars to uncover details of the emergence and history of transmission of individual texts and textual corpora and to better understand the intellectual history of Buddhism.

The tools developed for Tibetan should have wide applicability to other languages and corpora.