

Joint Research Group
Manuscriptology and Digital Humanities

Dharmamitra: New Tools for Sanskrit Translation, Grammatical Analysis, Search and Digital Philology

**09:15-10:45 Hrs**
(CET Germany)= 13:45-15:15 pm (IST India)
= 17:15-18:45 pm (JST Japan)**Friday 22.11.2024****ZOOM LINK**

Manuscriptology and Digital Humanities

Speaker: Sebastian Nehrdich
(UC Berkeley)

Abstract: Sanskrit presents unique challenges for digital processing due to the language's rich morphological complexity and the absence of word boundaries in written texts. While recent advances in Natural Language Processing have revolutionized the study of modern languages and made applications such as machine translation and reliable search engines possible, Sanskrit so far is lagging behind in these developments. In this talk, I will present Dharmamitra's Sanskrit-specific capabilities, particularly our new language model that achieves state-of-the-art accuracy in fundamental Sanskrit processing tasks such as word segmentation, lemmatization, and morphological analysis.

I will demonstrate how these technical advances translate into practical tools for Sanskrit scholars – from assisting in basic text analysis to enabling sophisticated corpus-wide semantic search and machine translation. The talk will showcase examples of how our system can provide detailed grammatical explanations, annotated translations, and facilitate textual research via semantic search even across language boundaries. These tools are designed to serve both beginning Sanskrit students and advanced scholars conducting specialized research. I will also demonstrate how Dharmamitra's capabilities can be used as building blocks for Sanskrit digitization and annotation projects.