

# The Sanskrit Heritage Platform

**G rard Huet**

Inria Paris Center,  
France

## **Abstract**

The Sanskrit Heritage Platform (<http://sanskrit.inria.fr>) offers a number of Web services meant for the assisted processing of classical Sanskrit texts. Its main component is a Sanskrit Reader, that may be used for segmentation and tagging of text at the level of sentences. A graphical user interface allows the user to quickly select the intended segmented text (*padap tha*) over a possibly enormous number of solutions. The segments may be presented as morphological lemmas in a two-level system, accounting for derivative morphology as well as surface inflexion, for nominals as well as verbal forms. They index a choice of lexicons, namely the Heritage Sanskrit to French dictionary and the Monier-Williams Sanskrit to English dictionary. Various sub-services are provided, such as a declension/conjugation tool, and its inverse lemmatizer. A new experimental Corpus Manager allows a community of annotators to develop analysed corpus in a way similar to the Perseus site for Latin and Greek texts, using the Git distribution system for crowd sourcing.

These tools link with an experimental Sanskrit parser by Amba Kulkarni at Indian Institute of Advanced Studies, Simla, in order to offer dependency parsing (*k raka* analysis), up to generation of the dependency graph of the sentence. It is also possible to access the Heritage Reader from the Sanskrit Library digital corpus developed by Peter Scharf.

The talk will consist of a demonstration of the various services on typical situations, and of a discussion of its design rationale, its relationship with Paninian methods, and its current shortcomings.