A universal subhāṣita database, and deduplicating sanskrit texts

Vishvas Vasuki

Dyugaṅgā, Beṅgaḷūru - 92

`vishvas.vasuki+ACAD@gmail`

Abstract

Subhāṣita-s are popular and beautiful Sanskrit quotations. In this paper, we propose a universal open-source subhāṣita database. As a part of it, we present an algorithm for deduplicating sanskrit texts, as well as one for identifying non-duplicate vairants of a given quote.

1 Motivation

Subhāṣitas are popular and beautiful Sanskrit quotations - they're usually verses. One of the greatest (and useful) pleasures I've had in tough times is retreat for a while into the world of beautiful Subhāṣitas- so as to burst forth with renewed wisdom and energy.

Since ages, they have been lovingly compiled (Parab and Pansikar., 1952) and memorized. I especially like online collections curated by some friends and myself since a book is not always available, and I want to collect + easily access choice ones for future enjoyment. But it is tedious (atleast for me) to sit in front of a computer to do the following:

- read them,

- or scour the internet for new ones

- or collect favorites in a spreadsheet

- or just annotate them with comments.

So, it is desirable to make the above as simple and easy as possible, and to share our collective labor so that we can benefit more easily from each others' work.

2 A universal database

We've set out to build a database of Subhāṣitas - which is:

- **Universal**

  - Its goal is to contain within it every worthy subhAShita*- ever composed.
  - In fact, the ambition encompasses all *languages*, *verse- and* prose- forms.

- **Freely and easily available**. Anyone should be able to

  - Access it
  - Export it to other formats
  - Present it in any way users will find convenient.

- **Growing constantly in number**, thanks to contemporary compositions.

- **Growing constantly in annotations/ ratings**, where annotations include rating, description, translations, metre, flaws, sources ….

3 Expected extensions

We hope that this will motivate other such long-sought-after universal databases for sanskrit connoiseurs, like: one for metres.

The front-end clients built for this database could serve as a model for other kAvya readers. Similarly, one can build a collaboratively annotated and rated collection of verses/ sentences within the context of long sequential works (rather than free floating subhAShita-s).

References

Kashinatha Pandurang Parab and Wasudeva Laxman Shastri Pansikar. 1952. *Subhasita-Ratna-Bhandagara or Gems of Sanskrit Poetry : Being a Collection of Witty, Epigrammatic, Instructive and Descriptive Verses.* Nirnaya Sagar Press.