## Jñānasaṅgrahaḥ: A Collection of Computational Applications related to Sanskrit

Hrishikesh Terdalkar hrishirt@cse.iitk.ac.in

Arnab Bhattacharya arnabb@cse.iitk.ac.in

Dept. of Computer Science and Engineering, Indian Institute of Technology Kanpur, India.

Jñānasaṅgrahaḥ is a web-based collection of several computational applications related to the Sanskrit language. The aim is to highlight the features of Sanskrit language in a way that is approachable for an enthusiastic user, even if she has a limited Sanskrit background. Jñānasaṅgrahaḥ is available at https://sanskrit.iitk.ac.in/jnanasangraha/. The applications part of Jñānasaṅgrahaḥ are described in the following sections.

#### Sankhyāpaddhatih

In the ancient India, it was a common practice to represent numeric values using letters, syllables or words from a natural language. The primary reason to use such systems is, ease of remembrance of numbers. We present a user-friendly web-based interface, Saṅkhyāpaddhatiḥ, which implements three such ancient Indian numeral systems, Kaṭapayādi Saṅkhyā (Terdalkar and Bhattacharya, 2019), Āryabhaṭīya Saṅkhyā and Bhūtasaṅkhyā.

The former two are alpha-syllabic numeral systems, while the latter is a number notation that uses ordinary words having implication of numeral values.

The central idea of an alpha-syllabic systems is that numeric values of the syllables are defined based on the constituent consonants and vowels. Usually, more than one syllable is assigned the same numerical value, however, every syllable has a unique numerical value, i.e. a many-to-one mapping of syllables to numbers. As a result, there is a unique value associated with a valid word or a phrase in a system, but there might be many valid representations of a number in the language.

The core interface for each of the system consists of an encoding interface to encode numeric values into a valid text representation a decoding interface to decode any valid text representation into the corresponding numeric value.

Figure ?? shows the encoding and decoding capabilities of Bhūtasaṅkhyā system. The Saṅkhyāpaddhatiḥ system is available at https://sanskrit.iitk.ac.in/jnanasangraha/sankhya/.

### Chandojñānam

Chandojñānam is a web-based system to detect the chanda in Sanskrit text. The system sports a user-friendly interface. The input can be provided as a plain text, a text file, or also in the form of an image. The input may make use of any transliteration scheme supported by the indic-transliteration package.

The system is tolerant towards erroneous texts and helps the user in locating and correcting such errors. The text processing supports two modes, treating the text as (1) a list of individual lines ( $p\bar{a}das$ ), and (2) a collection of verses. The system is able of finding approximate and close matches (termed, fuzzy matches) if a line or a verse does not directly correspond to a meter from the metrical database. Fuzzy matching is performed using edit distance comparison on the metrical signature of the lines ( $p\bar{a}da$ ). The tool also provides informative display showing the steps involved in the meter identification.

https://github.com/indic-transliteration/

Figure 2 shows the image upload interface and Figure 3 shows the fuzzy matching in action. A full version of the work ((Terdalkar and Bhattacharya, 2022)) explains the workings of the system. The system is available at https://sanskrit.iitk.ac.in/jnanasangraha/. The source code in the form of a Python library can be found at https://github.com/hrishikeshrt/chanda/.

#### Varņajñānam

Varṇa (वर्ण) is a phonetic unit of Sanskrit language. The Varṇajñānam system consists of utility functions related to varṇa information and manipulation. These include utilities to (1) form or resolve Pratyāhāra (2) split a word into its component varṇa, (3) display information related to the pronunciation of varṇa, namely uccāraṇa sthāna and prayatna and (4) calculate and display frequency of varṇa as well as pronunciation classes. The system is available at https://sanskrit.iitk.ac.in/jnanasangraha/varna/.

#### References

[Terdalkar and Bhattacharya2019] Hrishikesh Terdalkar and Arnab Bhattacharya. 2019. Katapayadi system. In *Proceedings of the 6th International Sanskrit Computational Linguistics Symposium*, IIT Kharagpur, India, October. Association for Computational Linguistics.

[Terdalkar and Bhattacharya2022] Hrishikesh Terdalkar and Arnab Bhattacharya. 2022. Chandojnanam: A sanskrit meter identification and utilization system.

Bhūtasaṅkhyā									
Saṅkhyā Home	About	Enco	de	Decode	Help	List	Examples		
			Dec	code text					
Encode number			सूर्याश्वाः						
712 Submit									
Submit									
DI - ( )   - E			Е	Bhūtasaṅkh	yā Num	bers			
Bhūtasaṅkhyā Encodings			712						
आदित्यद्वीपाः									
			Decode text						
मासाचला:			राश्यद्रय:						
सूर्याश्वाः			Su	bmit					
अर्कस्वरा:				_ <del></del>					
राश्यद्रयः			Bhūtasaṅkhyā Numbers						
			7	12					

Figure 1: Sańkhyā<br/>paddhatiḥ: Bhūtasańkhyā Encoding and Decoding

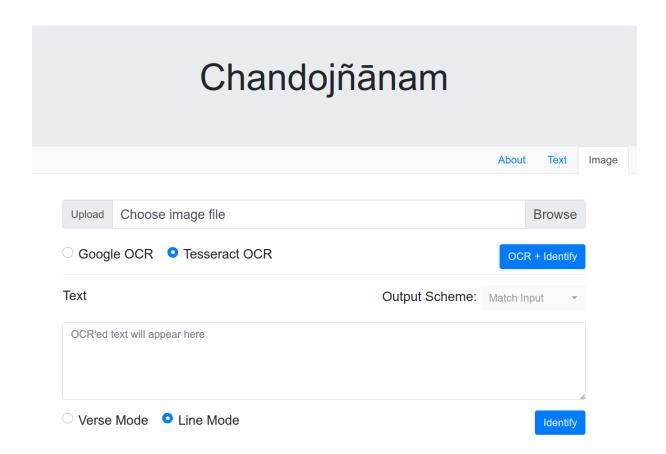


Figure 2: Chandojñānam: Image upload interface

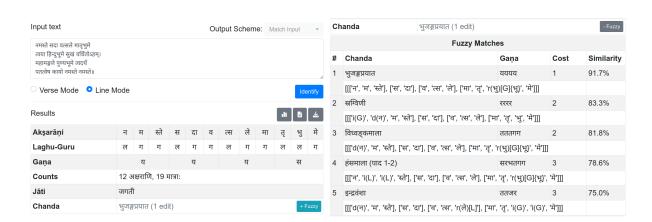


Figure 3: Chandojñānam: Meter identification with fuzzy matching and suggestions

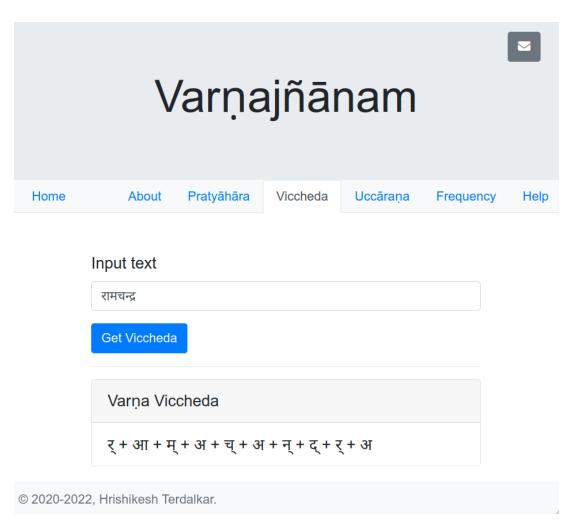


Figure 4: Varṇajñānam: Splitting varṇas (varṇavicchedaḥ)

# Input text रामः Get Uccāraṇa

Varņa	Sthāna	Ābhyantara	Bāhya
र्	मूर्धा	ईषत्स्पृष्टः	संवारः नादः घोषः अल्पप्राणः च
आ	कण्ठ:	विवृत:	संवारः नादः घोषः अल्पप्राणः उदात्तः च
म्	ओष्ठौ नासिका च	स्पृष्ट:	संवारः नादः घोषः अल्पप्राणः च
अ	कण्ठः	संवृत:	संवारः नादः घोषः अल्पप्राणः उदात्तः च
o <b>:</b>	कण्ठः	विवृत:	महाप्राण:

Figure 5: Varṇajñānam: Pronunciation information Uccaraṇasthānayatna