

Jñānasaṅgrahaḥ: A Collection of Computational Applications related to Sanskrit

Hrishikesh Terdalkar Arnab Bhattacharya
hrishirt@cse.iitk.ac.in 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.

Saṅkhyāpaddhatiḥ

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 Bhūtasāṅkhyā.

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 Bhūtasāṅkhyā 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`¹ 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 (pādas), 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 (pāda). The tool also provides informative display showing the steps involved in the meter identification.

¹<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/>.

Varṇajñānam

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.



Bhūtaśaṅkhyā

[Śaṅkhyā Home](#)[About](#)[Encode](#)[Decode](#)[Help](#)[List](#)[Examples](#)

Encode number

712

Submit

Bhūtaśaṅkhyā Encodings

आदित्यद्वीपाः
मासाचलाः
सूर्याश्वाः
अर्कस्वराः
राश्यद्रयः

Decode text

सूर्याश्वाः

Submit

Bhūtaśaṅkhyā Numbers

712

Decode text

राश्यद्रयः

Submit

Bhūtaśaṅkhyā Numbers

712

Figure 1: Śaṅkhyāpaddhatiḥ: Bhūtaśaṅkhyā Encoding and Decoding

Chandojñānam

[About](#)[Text](#)[Image](#)[Upload](#)[Choose image file](#)[Browse](#)☐ Google OCR ☒ Tesseract OCR[OCR + Identify](#)[Text](#)

Output Scheme:

[Match Input](#)

OCR'ed text will appear here

☐ Verse Mode ☒ Line Mode[Identify](#)

Figure 2: Chandojñānam: Image upload interface

Input text

नमस्ते सदा वत्सले मानुषुमे
त्वया हिन्दुभूमे सुखं वर्धितोऽहम्।
महामङ्गले पुण्याभूमे त्वदर्धे
पतलेष कायो नमस्ते नमस्ते॥

☐ Verse Mode ☒ Line Mode

[Identify](#)

Results

Akṣarāṇi	न	म	स्ते	स	दा	व	त्स	ले	मा	तृ	भु	मे										
Laghu-Guru	ल	ग	ग	ल	ग	ग	ल	ग	ग	ल	ल	ग										
Gaṇa	य		य		य		स															
Counts	12 अक्षराणि, 19 मात्राः																					
Jāti	जगती																					
Chanda	भुजङ्गप्रयात (1 edit)																					


Chanda

भुजङ्गप्रयात (1 edit)

[Fuzzy](#)

Fuzzy Matches				
#	Chanda	Gaṇa	Cost	Similarity
1	भुजङ्गप्रयात [[['न', 'म', 'स्ते'], ['स', 'दा'], ['व', 'त्स', 'ले'], ['मा', 'तृ', 'r(भु)(G)(भु)', 'मे']]]	यययय	1	91.7%
2	सन्धिणी [[['(G)', 'd(न)', 'म', 'स्ते'], ['स', 'दा'], ['व', 'त्स', 'ले'], ['मा', 'तृ', 'भु', 'मे']]]	रररर	2	83.3%
3	विध्वङ्कमाला [[['d(न)', 'म', 'स्ते'], ['स', 'दा'], ['व', 'त्स', 'ले'], ['मा', 'तृ', 'r(भु)(G)(भु)', 'मे']]]	तततगग	2	81.8%
4	हंसमाला (पाद 1-2) [[['न', 'l(L)', 'l(L)', 'स्ते'], ['स', 'दा'], ['व', 'त्स', 'ले'], ['मा', 'तृ', 'r(भु)(G)(भु)', 'मे']]]	सरभतगग	3	78.6%
5	इन्द्रवंशा [[['d(न)', 'म', 'स्ते'], ['स', 'दा'], ['व', 'त्स', 'r(ले)(L)], ['मा', 'तृ', 'l(G)', 'l(G)', 'मे']]]	ततजर	3	75.0%

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



Varṇajñānam

[Home](#)[About](#)[Pratyāhāra](#)[Viccheda](#)[Uccāraṇa](#)[Frequency](#)[Help](#)

Input text

रामचन्द्र

Get Viccheda

Varṇa Viccheda

र् + आ + म् + अ + च् + अ + न् + द् + र् + अ

© 2020-2022, Hrishikesh Terdalkar.

Figure 4: Varṇajñānam: Splitting varṇas (varṇavicchedaḥ)

Input text

रामः

Get Uccāraṇa

Varṇa	Sthāna	Ābhyantara	Bāhya
र्	मूर्धा	ईषत्स्पृष्टः	संवारः नादः घोषः अल्पप्राणः च
आ	कण्ठः	विवृतः	संवारः नादः घोषः अल्पप्राणः उदात्तः च
म्	ओष्ठौ नासिका च	स्पृष्टः	संवारः नादः घोषः अल्पप्राणः च
अ	कण्ठः	संवृतः	संवारः नादः घोषः अल्पप्राणः उदात्तः च
ः	कण्ठः	विवृतः	महाप्राणः

Figure 5: Varṇajñānam: Pronunciation information Uccaraṇasthānayātna