

# START: Sanskrit Teaching, Annotation, and Research Tool - Bridging Tradition and Technology in Scholarly Exploration

<sup>1</sup>Anil Kumar, <sup>2</sup>N. Shailaja, <sup>3</sup>Amba Kulkarni

Department of Sanskrit Studies

University of Hyderabad

{<sup>1</sup>anil.lalit22, <sup>2</sup>shailajanakkawar}@gmail.com,

<sup>3</sup>ambakulkarni@uohyd.ac.in

## Abstract

Sanskrit, a language renowned for its profound literature and philosophical insights, has remained a cornerstone of ancient wisdom. In this digital age, the fusion of tradition and technology has led to the emergence of transformative tools such as the Sanskrit Teaching, Annotation, and Research Tool (START), being developed by the Department of Sanskrit Studies, University of Hyderabad. This research paper delves into the intricate features, methodologies, and implications of START in reshaping the landscape of Sanskrit research and teaching. By exploring its advanced annotation capabilities, collaborative potential and broader impact on the digital humanities, this paper demonstrates how START is redefining the boundaries of scholarly exploration and analysis.

## 1 Introduction

The timeless wisdom present within Sanskrit texts has fascinated scholars for generations. Therefore, in spite of the literature being several centuries old, there is a continuously increasing population interested in understanding these texts. Several online computational tools exist that assist the learners in understanding the Sanskrit texts and enhancing their grammar. The traditional commentaries in the form of a book are being gradually replaced by E-readers. These E-readers were developed semi-automatically for popular texts such as Sankṣepa Rāmāyaṇam, ŚrīmadBhagvad Gītā and Śīsupālavadham.<sup>1</sup> Recently there was a similar effort to transform the Kramapāṭha reader of the Rāmopākhyāna to an E-reader with a search facility (Scharf and Chauhan, 2023), which was inspired by from its printed version. These E-readers are not only useful for the readers, but also are useful for building and enhancing NLP tools, since they provide an annotated gold data for ML techniques.

Samśādhanī and the Sanskrit Heritage Platform, though being used extensively for learning and teaching Sanskrit, and also for developing annotated gold data for various tasks such as segmentation, morph analysis, dependency parsing, etc. a need was felt to develop a better user interface and management of various tools and resources to cater to the diverse needs of the students, teachers, indology researchers, and computational linguists.

With the penetration of Digital Technology in every sphere of life, no wonder that the teaching and learning is affected the most. Sanskrit is no exception. During COVID-19, there was a surge in the learners of Sanskrit and this gave a boost to several online teaching programmes. The online computational tools for Sanskrit also saw an upward surge in its usage since many online courses either recommended these tools or the students in search of online help found them to be supplementary to their online courses.

One such course was also offered by Indian Institute of Technology, Roorkee in collaboration with Sanskrit Bharati where the online Sanskrit Computational Tools Samśādhanī and Sanskrit Heritage Platform were introduced to the students pedagogically. While the teachers followed the traditional methods of teaching, the students were encouraged to use the tools for completing the exercises. Pāṇini's grammar is useful for generating the word forms, joining the words following

---

<sup>1</sup><https://sanskrit.uohyd.ac.in/scl/e-readers>

the sandhi rules, or deciding the appropriate case suffix for a noun, etc.. But for analysis and understanding the text, it is not directly useful, since now the rules are to be used in a reverse way which may also lead to non-determinism. It is here the students found these online platforms useful. However there were several practical difficulties students faced while using these tools. The current platform was not mobile friendly, while most of the students prefer to use the tools on smartphones rather than on a laptop or a desktop. Many of them did not have any access to computing devices other than a mobile. Second major problem was due to the multiple solutions provided by the machine. Readers would like to see only one output rather than many, in the beginning.

Similar course was offered again by the University of Hyderabad in 2022 to teach Sanskrit from scratch using these computational tools. The aim of this course was to see how complete these tools are from a pedagogical perspective. This course, followed by a workshop for Sanskrit teachers at the University of Hyderabad in 2023 helped us to identify the features teachers and students would like to have in the display.

Another use of these tools is to develop a dependency tree bank for Sanskrit. The Sanskrit parser produced several possible dependency trees. While in most of the cases only one or two edges of the graph would go wrong, the interface for correcting the solution was not user friendly and hence annotators faced problems. Problems were also due to the compatibility issues among the devices, browsers, operating systems etc.

Yet another need was of the Indology researchers and computational linguists working with knowledge base or grammar based approaches. They needed a search facility over the annotated data. The existing interface for E-readers is developed targeting the readers, and hence does not provide a good search facility to the researchers.

These limitations of the tools from user's perspective, and the requirement of a large size corpus for the use of Machine Learning approaches led us to think of a platform that can be used simultaneously for learning, teaching, corpus annotation and research.

In the next section, we survey various efforts in these areas. In the third section, we describe the platform START which is developed to cater to the needs described above, describing various utilities available. Finally we describe the architecture of the platform.

## 2 Related work

To the best of our knowledge there does not exist any single platform that caters to the needs of diverse users such as students, teachers, computational linguists and Indologists. However, there are platforms that are designed towards serving some of them. We brief a few important ones.

- **Perseus Digital Library:**<sup>2</sup> This is an open-source project mainly focussing on Greek, Latin and Arabic providing a suite of services for interacting with textual collections. It provides an integrated reading environment with support of linguistic tools such as lemmatizer and morphological analyser with a link to the dictionaries that helps a user in contextual reading. Sanskrit Library<sup>3</sup> and Heritage Platform<sup>4</sup> started with a similar services. Recently two websites Vedaweb<sup>5</sup> and Ambuda<sup>6</sup> were developed which provide similar services for Sanskrit as are provided by the Perseus website. The first one is focussed on Rigveda and the second one on classical Sanskrit texts.
- **The Sanskrit Heritage Platform:**<sup>7</sup> This platform provides various Sanskrit computational tools such as segmentation, morphological analysis and generation and a shallow

---

<sup>2</sup><http://www.perseus.tufts.edu/hopper>

<sup>3</sup><https://sanskritlibrary.org>

<sup>4</sup><https://sanskrit.inria.fr>

<sup>5</sup><http://vedaweb.uni-koeln.de/rigveda>

<sup>6</sup><http://ambuda.org>

<sup>7</sup><http://sanskrit.inria.fr>

parser with lexicon linked to the bilingual Sanskrit-French and Sanskrit-English dictionaries. In addition, it has a well curated semi-automatically segmented corpus with lexicon driven and linked morphological analysis. It also hosts various short lessons targeted towards the new learners of Sanskrit, with complete word level analysis linked to the dependency parser of University of Hyderabad.

- **Sanisādhani**:<sup>8</sup> This platform developed at the University of Hyderabad offers various computational linguistic tools such as sandhi joiner, sandhi splitter, morphological analyser and generator, sentential parser, and Sanskrit-Hindi Translation, and lexical resources such as Amarakośa and Dhātuvṛttis. This platform is linked to the services of Sanskrit Heritage Platform, and provides an integrated environment for analysis of Sanskrit texts complementing the services.
- **DCS**:<sup>9</sup> The Digital Corpus for Sanskrit website is designed for text historical research in Sanskrit linguistics and philology. Users can search for lexical units and their collocations in a huge corpus of about 650,000 lines which is sandhi split as well as morphologically tagged.
- **SanskritShala**:<sup>10</sup> It is the recent addition to this list which provides a Neural Sanskrit NLP Toolkit with Web-Based Interface for Pedagogical and Annotation Purposes(Sandhan et al., 2023).
- **Sangrahaka**: This is an open-source tool that supports collaborative annotation and is tailored for Indian languages, enhancing regional NLP efforts(Terdalkar and Bhattacharya, 2021).

There are numerous platforms available for text annotations such as GATE, BRAT, INCEPTION, WebAnno, Bella, etc. As we will see below, these platforms are not geared towards teaching and research. Their sole aim is to facilitate annotation.

- **GATE**<sup>11</sup>: An open-source platform for text engineering that supports various types of annotations, such as named entities, relations, events, opinions, and coreference. GATE also provides a graphical user interface, a scripting environment, and a plugin architecture for extending its functionality.
- **BRAT**<sup>12</sup>: A web-based tool for collaborative text annotation that can handle span entities, relations, and attributes. BRAT also allows for the configuration of custom annotation schemes and visualizations.
- **Doccano**: A web-based tool for text annotation that supports sequence labeling, text classification, and sequence-to-sequence tasks. Doccano also offers features such as user management, project management, and data import/export(Nakayama et al., 2018).
- **INCEPTION**<sup>13</sup>: A web-based tool for intelligent text annotation that integrates machine learning and active learning techniques to assist human annotators. INCEPTION supports various types of annotations, such as named entities, relations, events, concepts, and sentiments.

---

<sup>8</sup><https://sanskrit.uohyd.ac.in/scl>

<sup>9</sup><http://www.sanskrit-linguistics.org/dcs>

<sup>10</sup><http://cnerg.iitkgp.ac.in/sanskritshala/>

<sup>11</sup><https://gate.ac.uk/>

<sup>12</sup><https://brat.nlplab.org/index.html>

<sup>13</sup><https://inception-project.github.io/>

- **WebAnno**<sup>14</sup>: WebAnno is an open-source web-based tool for annotating text data in natural language processing (NLP) tasks. It offers customizable schemas, collaborative annotation, and integration with NLP pipelines for efficient labeling.
- **Bella**<sup>15</sup>: A JavaScript-based tool for natural language processing tagging that supports part-of-speech tagging, chunking, named entity recognition, and dependency parsing. Bella also allows for the creation of custom tag sets and rules.

## 2.1 Why a new Platform?

While there are several of these platforms that may be extended/adapted towards our goals, we decided to go for a totally new platform due to the following considerations.

- None of these platforms is really geared towards teaching, and building E-readers.
- At the backend we use rule based tools which produce all grammatically possible solutions, which are prioritized using some heuristics or probabilistic models. We would like to provide access to all these analyses should an annotator finds the displayed analysis to be wrong.
- We also would like to use the annotated data to develop E-readers automatically. For languages like Sanskrit, where one needs a good support of lexical resources and grammar tools to understand any Sanskrit text, the annotated texts come handy for the readers with all necessary information at one place.

These reasons motivated us to develop this platform.

## 3 START: Sanskrit Teaching, Annotation and Research Tool

START is a web based tool that provides

### 1. A dashboard for learning and teaching Sanskrit

With the help of various online computational tools such as sandhi joiner and sandhi splitter, morphological analyser and generator, sentential analyser and generator, etc. available on Samsādhani Platform integrated with Sanskrit Heritage Platform, this dashboard provides an environment for the learning as well as teaching. The user interfaces for each of these tools are developed following the pedagogical requirement of a teacher. We describe them in section 3.2 in detail.

### 2. A dashboard for annotation and E-reader builder

This dashboard facilitates an annotator to annotate any Sanskrit text of his choice. The Samsādhani and Sanskrit Heritage Platforms provide the back-end support. Detailed description of this is provided in section 3.1.

### 3. User management System (UMS)

START provides a User Management System (UMS) which emerges as a fundamental tool that empowers START to seamlessly control user access, streamline administrative tasks, and ensure a smooth user experience. It provides User Registration and Onboarding, Role-Based Access Control, Password Management and Activity Monitoring and Auditing. Table 1 explains the roles.

### 4. Encoding Management

The character encoding plays a crucial role in ensuring accurate representation and seamless exchange of textual information, especially in the case of Sanskrit, since it is written in several Indian scripts as well as specially designed Roman script, and also in various

<sup>14</sup><https://webanno.github.io/webanno/>

<sup>15</sup><https://github.com/dennybritz/bella>

Roles	Description
Admin	User Account Creation and Deletion, User Role and Permissions Management, User Support, Monitoring and Reporting
Editor	Group Creation and Management, Monitoring and Reporting
Annotator	Annotation Creation, Viewing Content, Monitoring and Reporting
Viewer	Viewing Content, Read-only Access, Monitoring and Reporting, Limited Interaction

Table 1: Different Roles in User Management System

transliteration schemes, especially in electronic format. The Encoding Management of START provides users to access the data in different Roman transliteration schemes such as WX, IAST, Velthuis, Harvard Kyoto, and also various Indian scripts such as Devanagari, Telugu, Oriya, etc. The encoding management ensures that the data gets stored uniformly in WX notation and is displayed in the script or transliteration scheme of user’s choice, remembering the user’s choice across sessions. (See Fig 1).

## 5. Content Management

START also provides a robust Content Management system which is evolved to not only organise and display content but also provide users with the ability to add, delete, and edit data. With this facility, one can import the texts from existing websites, and then edit according to the version one is following.

## 6. A dashboard for Indologists<sup>16</sup>

The data in the content management system and also the one annotated will also be accessible to the Indology researchers through this dashboard. This dashboard will have an advanced search facility with grammatical and lexical queries. Further, this data can also be exported to any of the standard formats such as University of Hyderabad tagged data format, or CONLLU format etc. for the use of Machine Learning algorithms.

The Mind-map of START is shown in Fig 2.

Now we describe the two dashboards in detail.

### 3.1 A dashboard for annotation and E-reader builder

This dashboard is mainly for annotating the data. Typically the Sanskrit texts are found in sandhied form, and need segmentation before we proceed for any annotation. Majority of popular texts are also in poetic/verse form. This also imposes another requirement for processing that the input text be marked with sentential boundaries providing a meaningful unit for dependency parsing. It is also necessary to get the prose form of the input text so that the translation into other languages becomes easier. The dashboard caters to all these needs. The dashboard provides access to various tools such as

- **Segmenter:** The Sanskrit Heritage segmenter is plugged in which provides the best possible segmentation(Sriram et al., 2023). The annotator can edit the text if the segmentation is wrong at any place (See Fig 3).
- **Word and Sentence analyser:** The Samsāadhanī sentential parser is used to get the dependency parse of the segmented text. The best solution is provided in the form of a table with collapsible rows. For each word, the best morphological analysis in the context

<sup>16</sup>This is a part of the design, but is still not ready.

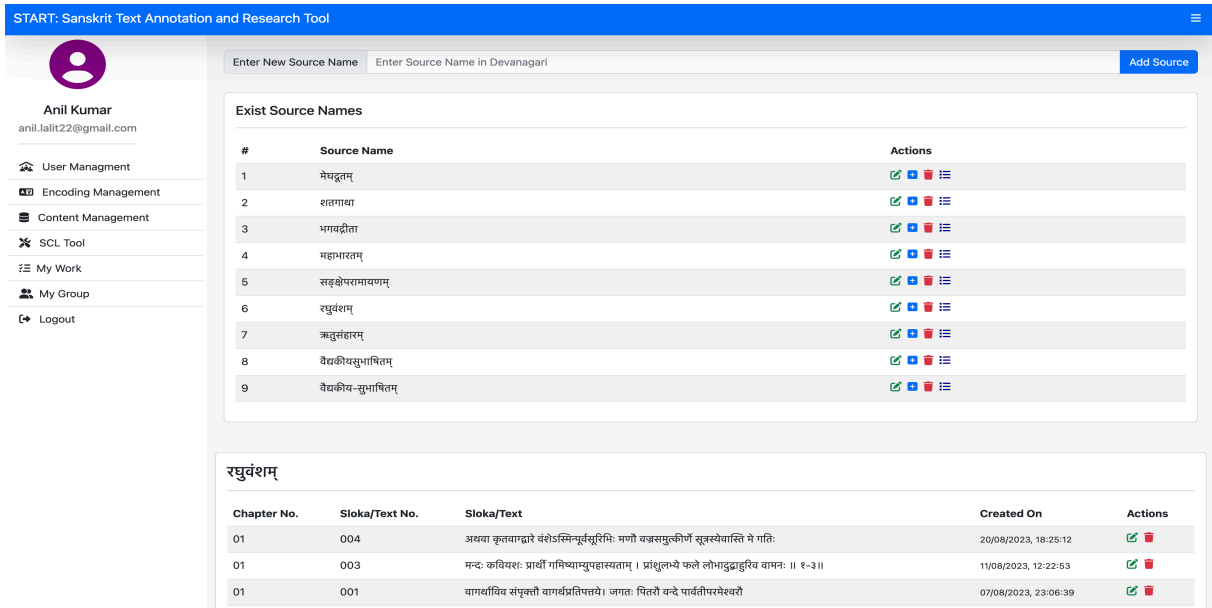


Figure 1: Content Management

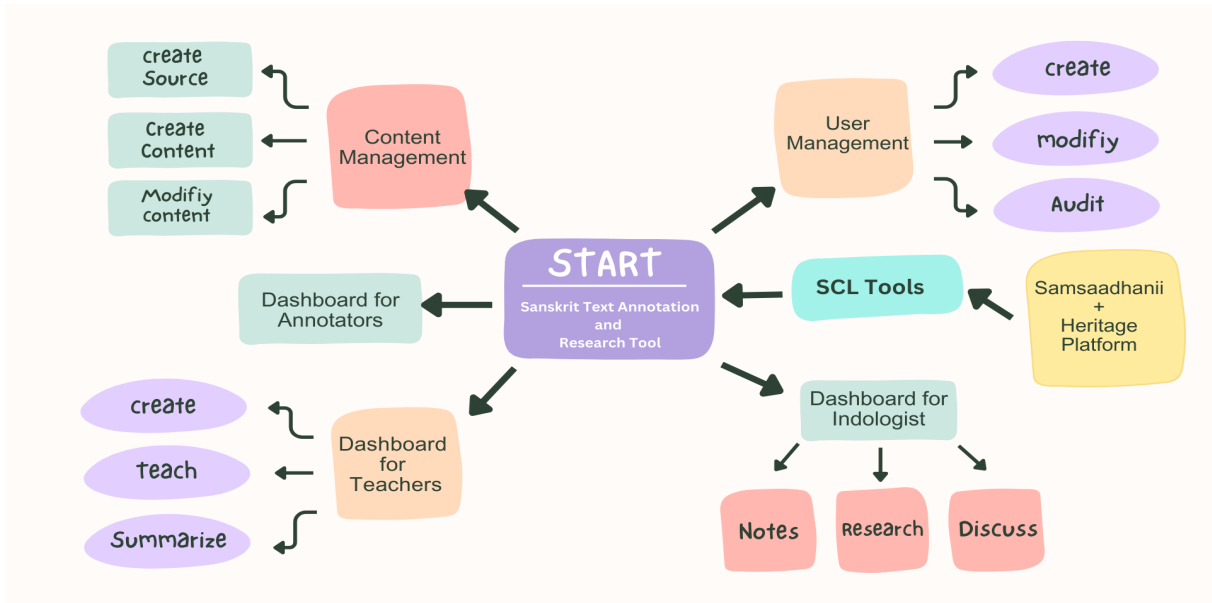


Figure 2: Mind-map of START

and its relation to other word in a sentence are provided, along with a link to the dictionary entry. The analysis generated at each stage is manually editable (See Figure 4). In the case of morphological analysis and dependency parsing, the annotator is provided with alternative possible analyses in the drop down menu from which s/he can select the correct one if the machine produced analysis is wrong. The dependency graph is displayed in the form of a tree, with colored nodes, color reflecting the grammatical category of the word. This way, the annotator can visualise the analyses and confirm her/his annotation (See Figs 4 and 5).

The dashboard shows the source text: "वागर्थाविव संपूक्तौ वागर्थप्रतिपत्तये। जगतः पितरौ वन्दे पार्वतीपरमेश्वरौ" and its analysis. The analysis is divided into three sections: Original Sloka, Sandhi Splitter, and Anvaya. The Sandhi Splitter shows the text split into "वाक्-अर्थो इव सम्पूक्तौ वाक्-अर्थ-प्रतिपत्तये जगतः पितरौ वन्दे पार्वती-परमेश्वरौ". The Anvaya section shows the text as "वागर्थाविव संपूक्तौ वागर्थप्रतिपत्तये। जगतः पितरौ वन्दे पार्वतीपरमेश्वरौ".

Figure 3: E-Reader Dashboard

वाक्-अर्थो इव सम्पूक्तौ वाक्-अर्थ-प्रतिपत्तये जगतः पितरौ वन्दे पार्वती-परम-ईश्वरौ

Index	1.1	1.2	2.1	3.1	4.1	4.2	4.3	5.1
Word	वाक्-	अर्थो	इव	सम्पूक्तौ	वाक्-	अर्थ-	प्रतिपत्तये	जगतः
Sandhi Word	वाक्-अर्थो	--	इव	सम्पूक्तौ	वाक्-अर्थ-प्रतिपत्तये	--	--	जगतः
Morph Analysis	वाक्	अर्थ(पुं,1,दि);अर्थ(पुं,2,दि);अर्थ	इव(अव्य)	सम्पूक्त(पुं,1,दि);सम्पूक्त(पुं,2,दि)	वाक्	अर्थ	प्रतिपत्ति(स्त्री,4,एक);प्रतिपत्ति(पुं,4,एक)	जगत्(पुं,5,एक)/जग
Morph In Context	वाक्	अर्थ(पुं,1,दि)	इव(अव्य)	सम्पूक्त(पुं,2,दि)	वाक्	अर्थ	प्रतिपत्ति(स्त्री,4,एक)	जगत्(पुं,6,एक)
Kaaraka Sambandha	समासः,1.2	उपमानम्,8.3	उपमान_श्रोतकः,1.2	विशेषणम्,6.1	समासः,4.2	समासः,4.3	प्रयोजनम्,7.1	षष्ठीसम्बन्धः,4.3
	Update	Update	Update	Possible Relations विशेषणम्,8.3 विशेषणम्,6.1 विशेषणम्,1.2 कर्म,7.1	Update	Update	Update	Update

Figure 4: Grammatical analysis of the given sloka

- **Prose Order Generator:** Finally we also provide a tool to generate the prose word order from the dependency graph. The output of this tool is also editable, in case the user is not happy with the machine generated word order. In the current version, we have not yet plugged in the prose word order generator.
- **Translator:** We plan to plug-in the Sanskrit-Hindi/Telugu/Marathi translation systems, once they are ready for deployment. The tool also allows one to edit the translation, or provide a manual translation of the text into other languages.

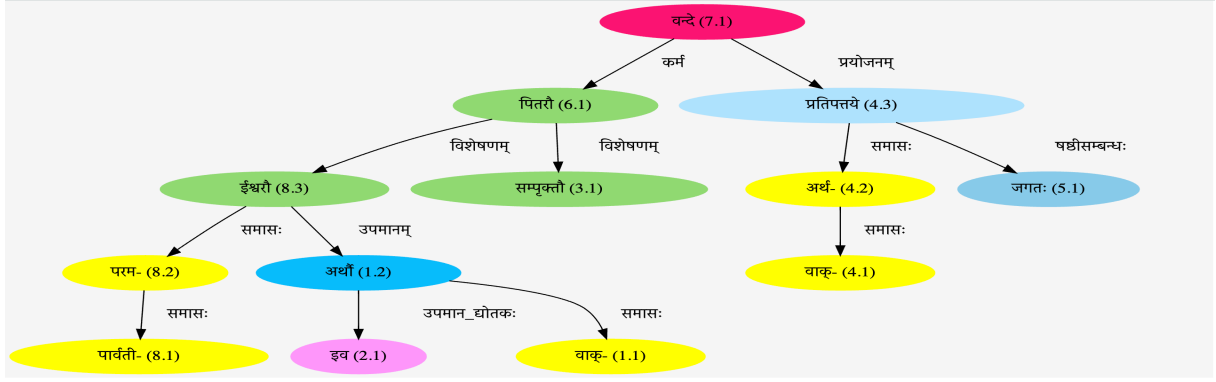


Figure 5: Dependency Tree

- **E-reader generator:** Once the analysed data is saved, the E-reader generator generates a page corresponding to the input dynamically showing analyses at all levels in a well structured manner with collapsible rows (See Fig 6). It is very important that the annotator herself visualises this data graphically to ensure that there are no mistakes (See Fig 7). E-reader builder provides this facility. Further, once such annotated data is certified by the editor, it is also made available to the general public, with a consent of the annotator and editor, on the E-reader platform, so that readers interested in reading such texts get benefited.

1.1	1.2	2.1	3.1
वाक्-	अर्थो	इव	सम्भक्तौ
वाक्-अर्थो	--	इव	सम्भक्तौ
वाक्	अर्थ(रु;1;दि)/अर्थ(रु;2;दि)/अर्थ	इव(अव्य)	सम्भक्त(रु;1;दि)/सम्भक्त(रु;2;दि)
वाक्	अर्थ(रु;1;दि)	इव(अव्य)	सम्भक्त(रु;2;दि)
षष्ठीसम्बन्धः,1,2	उपमानम्,5.1	उपमान_द्योतकः,1,2	विशेषणम्,6.1
4.1	4.2	4.3	5.1
वाक्-	अर्थ-	प्रतिपत्तये	जगतः
वाक्-अर्थ-प्रतिपत्तये	--	--	जगतः
वाक्	अर्थ	प्रतिपत्ति(स्त्री;4;एक)/प्रतिपत्ति(पुं;4;एक)	जगत्(रु;5;एक)/जगत्(रु;6;1)
वाक्	अर्थ	प्रतिपत्ति(स्त्री;4;एक)	जगत्(रु;1;षट्)
षष्ठीसम्बन्धः,4,2	षष्ठीसम्बन्धः,4,3	प्रयोजनम्,7.1	षष्ठीसम्बन्धः,6.1

Figure 6: e-Reader

### 3.2 A Dashboard for learning and teaching Sanskrit

Under this mode, several computational tools such as sandhi joiner and splitter, morphological analyser and generator, sentential analyser and generator, a web interface for Amarakośa, and a concordance of important dhātuvṛttis are available (See Fig. 8). We describe these tools in brief highlighting the pedagogical aspect.

- **Sandhi** : Sandhi is an enhanced version of Sanisāadhanī's sandhi Joiner having 3 teaching modes such as Basic, Intermediate and Advanced. The basic mode can be used for teaching sandhi rules in the schools and also its very useful for beginners. The Intermediate and



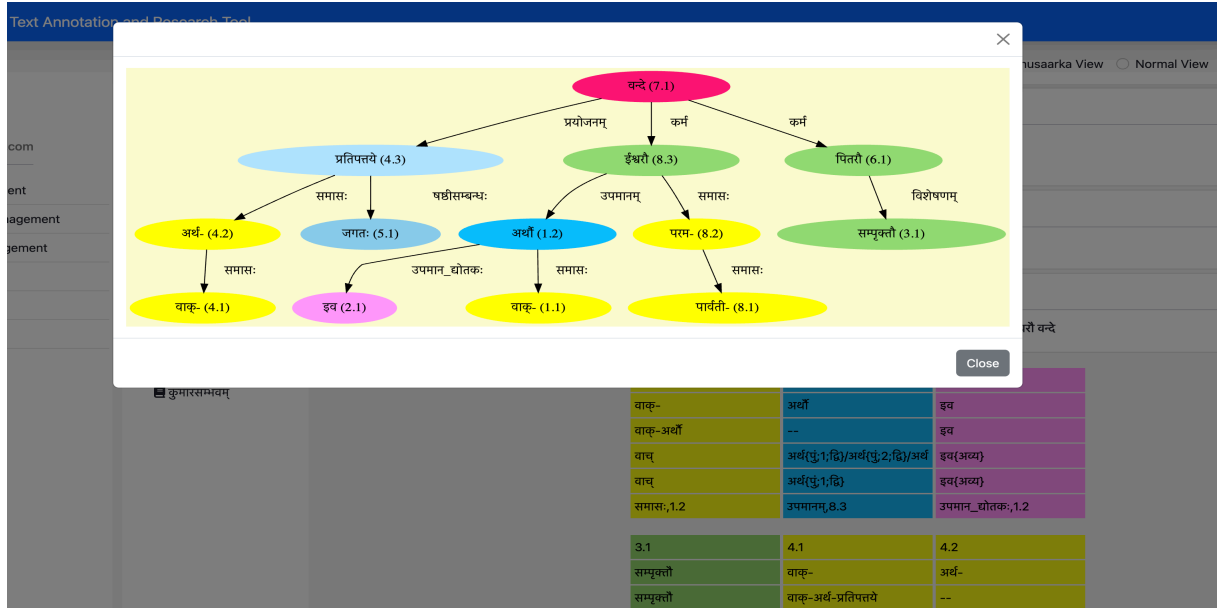


Figure 7: e-Reader with Parser

Figure 8: List of SCL tools

Advanced modes are for Sanskrit scholars or lovers who want to learn sandhi with Pāṇinian grammar. (See Fig 9)

- **Sandhi-Splitter** : Sandhi-splitter automatically splits or separates words that have undergone sandhi changes back into their original forms. It can be used for splitting single word or a full sentence or a paragraph. At the back-end, the Heritage Sanskrit Engine that prioritizes the solutions based on probability models is used.
- **Noun-form Generator** : It generates inflected noun forms of a given nominal stem (prātipadikam) in all the seven cases, and three numbers.
- **Verb-form Generator**: It generates the conjugational verb-forms of a verbal root (dhātu) with various tense, mood, voice, person, number, and one or more prefixes.
- **Morphological Analyser** : It provides all possible analyses of a given word. This is very much useful for a reader who has doubts about the morphological analysis of any word.

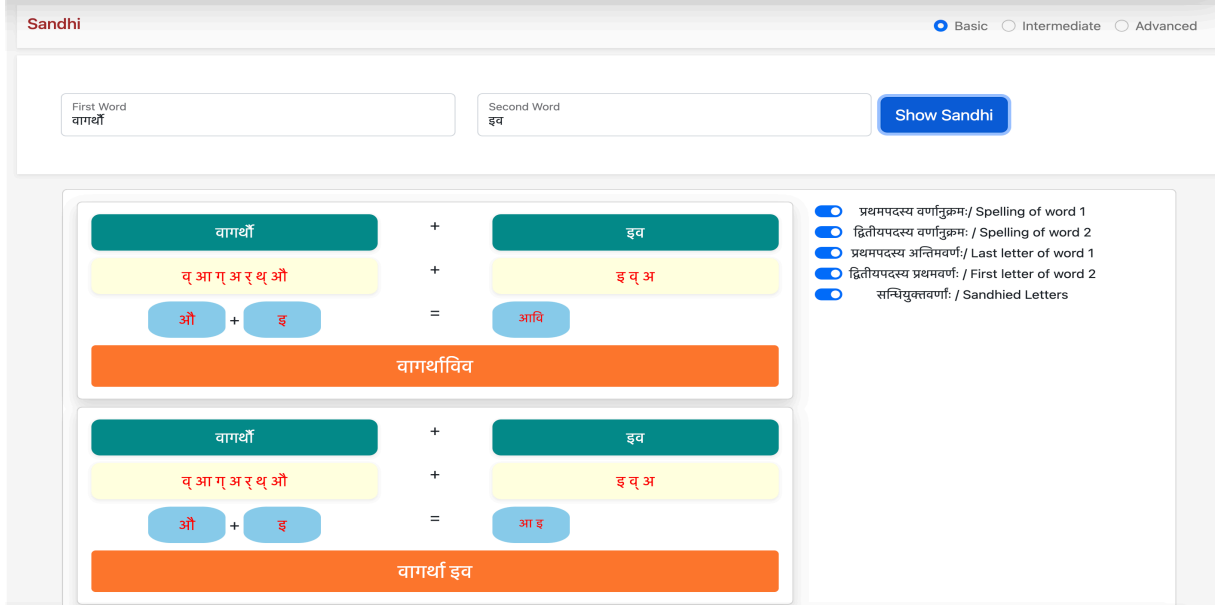


Figure 9: Sandhi

- **Concordance of Pāṇinian Dhātuvṛttis** : The current tool provides the comparison of Pāṇinian Dhātuvṛttis. The three vṛttis chosen are Mādhaviyā Dhātuvṛtti by Sāyaṇa, Kṣīratarāṅgiṇī by Kṣīrasvāmin and Dhātupradīpa by Maitreyarākṣita. This resource is very much useful for the Sanskrit scholars and students.
- **Sentential Parser**: This is part of the E-reader Dashboard.
- **Sentence Generator, Compound Word Generator and Amarakośa**: The interfaces for these are being developed. The original tools are available under Saṁsādhanī.

### 3.3 User Profile

Under user profile, each user can track his/her own work. The facility to form a group is useful for the team leader to monitor the progress of each member. A teacher can use this feature to design different assignments for different classes. If the answers are provided by the teachers in advance, the system can provide instant feedback to the students on their performance.

## 4 Technological Stack in Development

The combined utilization of Angular 13, Bootstrap 5, CGI, MongoDB 4, Python3 Flask, and Graphviz forms a powerful technological stack that empowers START with dynamic frontend interfaces, robust backend functionality, and insightful data visualization capabilities. This comprehensive toolkit lays the foundation for a feature-rich, responsive, and user-centric application.

- **Angular**:<sup>17</sup> Angular is a robust open-source framework for creating dynamic, single-page web applications, known for its component-based architecture and two-way data binding.
- **Bootstrap**:<sup>18</sup> Bootstrap is a widely-used open-source CSS framework that offers a collection of pre-designed templates and components, streamlining the process of building responsive and visually appealing websites.
- **CGI**: CGI (Common Gateway Interface) is a standard protocol for web servers to communicate with external programs, enabling dynamic content generation and interaction between web servers and applications.

<sup>17</sup><https://angular.io/>

<sup>18</sup><https://getbootstrap.com/>

- **MongoDB:**<sup>19</sup> MongoDB is a NoSQL database system that employs a document-oriented model for data storage, providing scalability and flexibility for managing large volumes of unstructured data.
- **Python Flask:**<sup>20</sup> Flask is a lightweight and flexible web framework for Python, simplifying web application development by offering essential tools and libraries, making it suitable for small to medium-sized projects.
- **Graphviz:**<sup>21</sup> Graphviz is an open-source graph visualization software that facilitates the creation of diagrams and graphs for visualizing complex relationships and structures in data or information systems.

## 5 Impact on Digital Humanities

START’s innovation transcends the realm of Sanskrit studies, influencing the broader landscape of digital humanities. Its robust annotation framework, collaborative ethos, and analytical tools serve as a blueprint for similar initiatives in other languages and disciplines. As the digital humanities continue to evolve, START paves the way for a more interconnected and enriched scholarly community.

## 6 Future Horizons

The future of START holds tantalizing prospects. Integration with machine learning algorithms, natural language processing, and visualization tools could further amplify the platform’s analytical capabilities. Additionally, efforts to expand the corpus of annotated Sanskrit texts could yield a more comprehensive understanding of the language’s nuances, intricacies, and cultural significance.

## 7 Conclusion

The Sanskrit Teaching, Annotation and Research Tool (START) stands as a testimony to the harmonious convergence of ancient wisdom and modern technology. Its transformative annotation system and broader impact on the digital humanities signal a new era of scholarly exploration and inquiry. In embracing START, researchers embark on a journey that not only uncovers the depths of Sanskrit literature but also redefines the very essence of interdisciplinary and intertemporal scholarly engagement.

## Acknowledgements

This research was made possible through the generous financial support of the IoE Directorate, University of Hyderabad for the project ‘An Integrated Digital Platform for Language Learning, Teaching and Computational Linguistics’ (2021-23).

## References

- The Text Encoding Initiative Consortium. 2021. Text encoding and interchange (TEI): <https://tei-c.org/guidelines/>.
- Pawan Goyal and Gerard Huet. 2016. Design and analysis of a lean interface for Sanskrit corpus annotation. *Journal of Language Modelling*, 4(2):145–182, Oct.
- Pawan Goyal, Gérard Huet, Amba Kulkarni, Peter Scharf, and Ralph Bunker. 2012. A distributed platform for Sanskrit processing. In *Proceedings of COLING 2012*, pages 1011–1028, Mumbai, India, December. The COLING 2012 Organizing Committee.

<sup>19</sup><https://www.mongodb.com/>

<sup>20</sup><https://flask.palletsprojects.com/en/2.3.x/>

<sup>21</sup><https://graphviz.org/>

- Oliver Hellwig and Sebastian Nehrlich. 2018. Sanskrit word segmentation using character-level recurrent and convolutional neural networks. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2754–2763, Brussels, Belgium, October–November. Association for Computational Linguistics.
- G erard Huet. 2004. Design of a lexical database for Sanskrit. In *Proceedings of the Workshop on Enhancing and Using Electronic Dictionaries*, pages 8–14, Geneva, Switzerland, August 29th. COLING.
- Amrith Krishna, Shiv Vidhyut, Dilpreet Chawla, Sruti Sambhavi, and Pawan Goyal. 2020. SHR++: An interface for morpho-syntactic annotation of Sanskrit corpora. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 7069–7076, Marseille, France, May. European Language Resources Association.
- Amba Kulkarni and Devanand Shukl. 2009. Sanskrit morphological analyser: Some issues. *Indian Linguistics*, 70(1-4):169–177.
- Anil Kumar, Vipul Mittal, and Amba Kulkarni. 2010. Sanskrit compound processor. In Girish Nath Jha, editor, *Sanskrit Computational Linguistics*, pages 57–69, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Hiroki Nakayama, Takahiro Kubo, Junya Kamura, Yasufumi Taniguchi, and Xu Liang. 2018. doccano: Text annotation tool for human <https://github.com/doccano/doccano> software available from <https://github.com/doccano/doccano>.
- Jivnesh Sandhan, Anshul Agarwal, Laxmidhar Behera, Tushar Sandhan, and Pawan Goyal. 2023. Sanskritshala. In *A neural sanskrit nlp toolkit with web-based interface for pedagogical and annotation purposes*.
- Peter M Scharf and Dhruv Chauhan. 2023. R amop akhy ana. In *A web-based reader and index. In Proceedings of the Computational Sanskrit & Digital Humanities: Selected papers presented at the 18th World Sanskrit Conference, pages 146–154, Canberra, Australia (Online mode), January. Association for Computational Linguistics*.
- Krishnan Sriram, Amba Kulkarni, and G erard Huet. 2023. Validation and normalization of DCS corpus and development of the Sanskrit heritage engine’s segmenter. In *Proceedings of the Computational Sanskrit and Digital Humanities: Selected papers presented at the 18th World Sanskrit Conference*.
- Hrishikesh Terdalkar and Arnab Bhattacharya. 2021. Sangrahaka: A tool for annotating and querying knowledge graphs. In *Proceedings of the 29th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE’21)*.