AMRUTVAANI: An AI-Based Multilingual Platform for Understanding Sanskrit Shlokas and Indian Scriptures

Khushi Mehta[1], Yash Ambale [2], Niharika Haval[3], Shrawani Pawar[4]
[1,2,3,4]*Undergraduate Student,Dr.Bapuji Salunkhe Institute of Engineering & Technology, Kolhapur*

***Abstract*—Indian scriptures contain valuable knowledge, stories, and teachings, but understanding them can be difficult for people who are not familiar with Sanskrit. AMRUTVAANI is a proposed AI-based platform that aims to make Sanskrit shlokas easier to understand through translations, word-by-word meanings, and simple explanations. It is planned to include features such as text recognition from images, multilingual support, audio, and an interactive chatbot. The initial work focuses on a Ganapati Stotra dataset containing Sanskrit verses and their word meanings. A Python prototype has been developed to load the dataset and display a selected verse with its available meanings. Further development will focus on adding AI-based language processing and integrating the features into a web application. The goal is to make learning about Sanskrit and Indian scriptures more accessible and interesting**

***Index Terms*—Artificial Intelligence, Indian Scriptures, Multilingual Learning, Natural Language Processing, Sanskrit, Shloka Understanding.**

I. INTRODUCTION

India has a rich collection of scriptures, including the Bhagavad Gita, Ramayana, Mahabharata, Puranas, and Sanskrit stotras. These texts contain important teachings, philosophical ideas, and cultural knowledge.

However, many young readers find it difficult to understand them because Sanskrit vocabulary and sentence structures can be unfamiliar. Sometimes, understanding a single shloka requires checking different books, dictionaries, or websites.

AMRUTVAANI aims to make this process easier by using Artificial Intelligence (AI) and Natural Language Processing (NLP). The platform is planned to provide translations, individual word meanings, and simple explanations in different languages.

The project begins with Ganapati Stotra and is planned to expand to other Indian scriptures in future development.

II. PROBLEM STATEMENT

Many young readers are interested in Indian scriptures but find it challenging to understand Sanskrit verses. Some can read the original text but cannot understand its meaning, while others depend on translations that may not explain every word clearly.

The major problems identified are:

1. Difficulty understanding Sanskrit because of language differences.
2. Unfamiliarity with Sanskrit vocabulary and grammatical structures.
3. Lack of simple, beginner-friendly explanations in one place.
4. Difficulty converting Sanskrit text from printed pages or photographs into digital text.
5. Learning material being spread across different websites, books, and applications.
6. Limited availability of platforms that combine meanings, explanations, audio, and interactive learning.

AMRUTVAANI aims to address these problems by providing a structured and accessible platform where users can explore Sanskrit shlokas and understand their meanings with the help of modern technology.

III. PROCEDURE FOR PAPER SUBMISSION

The research paper is prepared using the guidelines provided by IJIRT. It covers the main parts of the

project, including the introduction, problem statement, objectives, literature review, proposed system, methodology, implementation, and expected outcomes.

Before submitting the paper, all team members and the project guide review it carefully. They check whether the information is clear, the technical details are correct, the references are properly included, and the formatting follows the IJIRT template.

Once the review and necessary corrections are completed, the paper is saved in PDF format and submitted through the journal's official submission process. If the journal reviewers suggest any changes, the team makes the required revisions and prepares the final version.

IV. UNITS

AMRUTVAANI is a software-based project that focuses on making Sanskrit shlokas and Indian scriptures easier to understand. It mainly involves Sanskrit text processing, word-by-word meanings, translation into different languages, and AI-based learning features. Therefore, the project does not require many physical units of measurement. However, certain standard units are useful for measuring the performance and technical requirements of the system.

During the development and testing of AMRUTVAANI, the processing time of different tasks can be measured in milliseconds (ms) or seconds (s). For example, the time taken to extract Sanskrit text from an uploaded image, process a shloka, or display its meaning can be recorded. These measurements will help us understand how quickly the application responds to user requests.

The storage space required for the dataset, images, audio files, and other application resources can be measured in kilobytes (KB), megabytes (MB), or gigabytes (GB). Similarly, the size of images and documents uploaded by users can be expressed in KB or MB. If online services are used, the amount of data transferred during application usage can also be measured in MB.

Some additional measurements may be required to evaluate the performance of individual modules. For example, OCR performance can be expressed as a percentage by comparing the extracted text with the original Sanskrit text. Translation and text-processing quality can be evaluated using suitable methods. The actual performance values will be collected during the testing phase and included in the final evaluation.

Using standard units will help maintain consistency while recording the results. It will also make it easier to compare different modules, identify areas that need improvement, and understand the overall performance of AMRUTVAANI. All measurements will be reported clearly and accurately, based on the results obtained during development and testing.

V. HELPFUL HINTS

A. Figures and Tables

The AMRUTVAANI research paper mainly presents its information through text-based explanations. The proposed system, its working process, and its features are described in the respective sections. Tables may be used wherever necessary to present dataset details, technologies, and evaluation results in an organized manner. All tables should be numbered using Roman numerals and provided with suitable titles.

B. References

References are included to support the information presented in the research paper. Sources related to Sanskrit language processing, Optical Character Recognition, machine translation, and Retrieval-Augmented Generation are cited wherever required. Citations are numbered in square brackets according to their order of appearance. Complete details of the cited sources are provided at the end of the paper.

C. Abbreviations and Acronyms

Abbreviations are explained when they first appear in the paper. Some terms used in AMRUTVAANI include Artificial Intelligence (AI), Machine Learning (ML), Natural Language Processing (NLP), Optical Character Recognition (OCR), and Retrieval-Augmented Generation (RAG). Their short forms are used in the remaining sections to maintain clarity and avoid repetition.

D. Equations

Mathematical equations are used only when required to explain the working or evaluation of the proposed system. Any equation included in the paper is numbered consecutively, and all symbols are clearly explained. Since AMRUTVAANI mainly focuses on

Sanskrit text processing, multilingual translation, and scripture-based learning, mathematical equations are not necessary for every part of the system.

VI. PUBLICATION PRINCIPLES

The AMRUTVAANI research paper is prepared by keeping academic guidelines and research ethics in mind. Our aim is to explain the project in a clear and honest way, covering its purpose, development, and the technologies involved. The information presented in the paper is based on our project work, dataset study, and research related to Artificial Intelligence and Sanskrit language processing.

Originality and Proper Referencing: The paper focuses on our own project ideas and development work. Whenever information is taken from research papers, books, or online sources, proper citations and references are provided. This gives credit to the original authors and helps maintain the originality of our work.

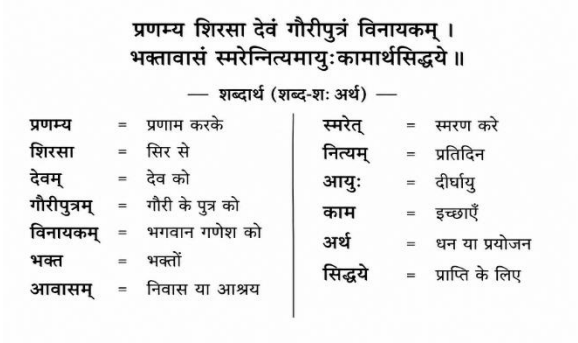
Accuracy of Information: We have taken care to present the project details correctly. Features that are still being developed are described as proposed features rather than completed ones. Any accuracy values or performance results will be included only after proper testing.

Research Ethics: The project follows basic research principles such as honesty, fairness, and transparency. Sanskrit verses, translations, and word meanings used in the dataset need to be checked carefully. Wherever the original source of the content is known, it is properly acknowledged.

Responsible Use of AI: AI-based tools can sometimes provide incorrect translations or explanations. Therefore, their output should be reviewed before being presented to users. AMRUTVAANI is designed as a learning aid to help people understand Sanskrit shlokas, not as a replacement for original scriptures or guidance from knowledgeable teachers.

Clarity and Transparency: The paper explains the project's objectives, proposed system, development approach, limitations, and future plans. The current stage of development is also described honestly so that readers can clearly understand what has been completed and what is planned.

By following these principles, we aim to make the AMRUTVAANI research paper reliable, original, and academically responsible. Through this project, we hope to show how modern technology can be used thoughtfully to make Sanskrit shlokas and Indian scriptures easier for learners to explore and understand.



VII. CONCLUSION

AMRUTVAANI is being developed to make Sanskrit shlokas and Indian scriptures easier to understand with the help of modern technology. The main idea is to combine features such as Sanskrit text recognition, word-by-word meanings, multilingual translation, and simple explanations in one platform. This will help learners understand traditional verses without finding the information from different sources.

The initial work with the Ganapati Stotra dataset has provided a good foundation for the project. It contains Sanskrit verses, transliteration, and word meanings in different languages. Further development will focus on adding OCR, translation, chatbot support, audio features, and a mobile application.

Overall, AMRUTVAANI aims to bring India's traditional knowledge closer to the younger generation by combining it with modern AI technology and creating a simple and engaging learning experience.

APPENDIX

The initial dataset used in AMRUTVAANI is based on the Ganapati Stotra. It contains Sanskrit verses, their transliteration, and word-by-word meanings in English and Hindi. This dataset is stored in an Excel workbook with separate sheets for verses, words, and translations. During the initial development, Python and the Pandas library were used to load and examine the dataset. A basic prototype was also created to display a selected Sanskrit verse along with its corresponding word meanings.

The dataset and prototype will be used as the foundation for further development of the AMRUTVAANI system. Additional scripture content and language support may be included in future versions.

ACKNOWLEDGMENT

We are grateful to our project guide Mrs.S.S.Kamat and faculty members for their constant support, guidance, and helpful suggestions during the development of AMRUTVAANI. Their feedback has helped us understand the project better and improve our work throughout the process.

We would also like to thank Dr. Bapuji Salunkhe Institute of Engineering & Technology, Kolhapur, for providing us with the opportunity and support to work on this project. We sincerely appreciate our department and college for encouraging us throughout the development process. We are also thankful to the researchers and developers whose work on Sanskrit language processing, Optical Character Recognition, machine translation, and Artificial Intelligence has helped us understand the technologies used in our project.

Lastly, we thank all our team members for sharing their ideas, working together, and contributing their efforts to AMRUTVAANI. The cooperation and support of everyone involved have played an important role in this project.

REFERENCES

[1]. R. Smith, "An Overview of the Tesseract OCR Engine," in Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), IEEE Computer Society, 2007, pp. 629–633. doi: 10.1109/ICDAR.2007.4376991.

[2]. J. Gala, P. A. Chitale, A. K. Raghavan, V. Gumma, S. Doddapaneni, A. Kumar, J. A. Nawale, A. Sujatha, R. Puduppully, V. Raghavan, P. Kumar, M. M. Khapra, R. Dabre, and A. Kunchukuttan, "IndicTrans2: Towards High-Quality and Accessible Machine Translation Models for All 22 Scheduled Indian Languages," Transactions on Machine Learning Research, 2023. Available:https://openreview.net/forum?id=vfT4YuzAYA.

[3]. P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W. T. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in Advances in Neural Information Processing Systems (NeurIPS), vol. 33, 2020, pp. 9459–9474. Available:https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html.