WWW.IJRPR.COM

**International Journal of Research Publication and Reviews**

Journal homepage: www.ijrpr.com ISSN 2582-7421

# Nirukti: A Saṁskṛta Natural Language Processing (NLP) System

***MukundHari R[1], Shrinkhala Kumari[2]***

Undergraduate Scholar, Department of BCA, BMS College of Commerce and Management, Bengaluru, Karnataka, India.
Assistant Professor, Department of BCA, BMS College of Commerce and Management, Bengaluru, Karnataka, India.

**A B S T R A C T :**

Nirukti is an intelligent and interactive Sanskrit linguistic platform developed to analyze, explain, and derive classical Saṁskṛta words—specifically focusing on the 108 sacred names of Lord Viṣṇu (Viṣṇu Aṣṭottaraśatanāma). Bridging the gap between traditional Sanskrit grammar (Vyākaraṇa) and modern computational tools, the system integrates a structured dictionary with AI-powered natural language processing using Python, Flask, and spaCy. Unlike traditional resources that are fragmented and non-interactive, Nirukti offers unified access to grammatical attributes such as dhātu (verbal root), vigrahavākya (derivational analysis), artha (meaning), vibhakti (case), liṅga (gender), and vacana (number) for each name. The application employs fuzzy matching algorithms and semantic intent recognition to accurately interpret user queries—even with inconsistent transliterations—and returns linguistically rich outputs. Additional features such as reverse search by meaning, conversational interaction, and voice-based input lay the groundwork for future expansion into mobile applications, grammar tutoring, and multilingual support. Nirukti exemplifies the synergy of tradition and technology, aiming to make Sanskrit accessible, engaging, and intelligible to a global audience of learners, scholars, and enthusiasts.

**Keywords**: Nirukti, Sanskrit, Vishnu, NLP, Dictionary, fuzzy, Python

## Introduction

Nirukti is an intelligent and interactive dictionary application developed to analyze and elucidate Saṁskṛta words, with a specific focus on the 108 primary names of Śrī Viṣṇu (Viṣṇu Aṣṭottaraśatanāma). In the contemporary age of digital linguistics and artificial intelligence, ancient languages such as Saṁskṛta are witnessing a revival through computational tools that transform how scholars, educators, and enthusiasts engage with classical texts.

As one of the most ancient, structurally intricate, and semantically profound languages, Saṁskṛta holds immense significance across spiritual, literary, and philosophical domains. However, its complexity often poses a barrier to learners and researchers, particularly in a digital-first world. To address this challenge, the Nirukti system seeks to bridge traditional grammatical principles with modern technological paradigms, offering a seamless platform for both analysis and learning.

By integrating grammatical parsing, semantic linkage, and AI-assisted interactivity, Nirukti provides an innovative model for exploring the depth of Saṁskṛta linguistics. The system not only aids in accurate interpretation of words but also fosters broader accessibility, aiming to make Saṁskṛta study more engaging, systematic, and intelligible for a diverse global audience.

## Literature Review

Despite the growing interest in computational linguistics and language technologies, there exists no integrated system with an architecture similar to *Nirukti*. The study and interpretation of Saṁskṛta names—particularly those found in sacred texts such as the *Viṣṇu Aṣṭottaraśatanāma*—remain largely dependent on manual methods and dispersed traditional sources. While individual resources such as grammar treatises, lexicons, academic papers, and scriptural commentaries provide valuable insights, they are accompanied by several limitations:

- ***Fragmented Information****:*
  Essential linguistic components such as *dhātu* (verbal root), *vigrahavākya* (analytical derivation), *artha* (meaning), and grammatical parameters like *vibhakti*, *liṅga*, and *vacana* are scattered across multiple texts and platforms. This fragmentation makes it difficult to access a cohesive explanation in a unified format.
- ***Non-Digital and Static Sources****:*
  A significant portion of the available data exists in printed books, scanned documents, or static web pages. These formats lack interactivity, dynamic content generation, and smart data retrieval capabilities.
- ***Lack of Personalization or Querying Mechanisms****:*
  Current tools do not offer users the ability to query specific names and receive detailed, structured linguistic outputs that encapsulate multiple grammatical dimensions in one place.
- ***Absence of AI-Powered Linguistic Analysis****:*

There are no existing systems that utilize natural language processing or artificial intelligence to parse, analyze, or explain Saṁskṛta names. The lack of such intelligent systems limits the pedagogical and research utility of current resources.

- ***High Learning Curve for Beginners****:*
  In the absence of guided or interactive tools, learners must rely on extensive training in traditional grammar (*Vyākaraṇa*) to understand the structure and significance of Saṁskṛta words. This poses a considerable entry barrier, especially for modern students.

The *Nirukti* system is designed to address these gaps by integrating classical linguistic knowledge with modern computational techniques, providing a digital platform that is both accessible and academically rigorous.

## Methodology

The core functionality of the *Nirukti* system is driven by a structured Python-based backend integrated with a conversational interface built using the Flask framework. At the heart of the application lies a meticulously curated Python dictionary containing all 108 names from the *Viṣṇu Aṣṭottaraśatanāma*, each annotated with relevant linguistic and semantic attributes such as *prakriyā* (derivational process), *vigrahavākya* (analytical expression), meaning (*artha*), positional index in the list, and the associated mantra.

The system operates through a simple yet effective pipeline designed to simulate natural language interaction. Users can communicate with *Nirukti* in English through a chat-like interface, where queries are interpreted and processed using a combination of string-matching algorithms and natural language understanding libraries. The core logic involves the following steps:

1. **Input Interpretation and Normalization**
   The user query, typically in English, is parsed to identify potential Saṁskṛta names or semantic intents. The system uses fuzzy matching techniques and transliteration handling to account for alternate spellings and phonetic variations in Saṁskṛta names.
2. **Name Matching and Retrieval**
   The parsed input is matched against the internal dictionary. If an exact match is found, the system retrieves and displays the complete linguistic breakdown. If no exact match exists, the system attempts partial matches based on closest resemblance, and if no significant match is found, it responds gracefully, indicating the absence of the queried name.
3. **Reverse Search Capability**
   The system also supports reverse lookup functionality. Users can provide a meaning or thematic keyword (e.g., "names related to water"), and *Nirukti* will return all names from the database that correspond semantically, along with their individual grammatical and semantic attributes.
4. **Response Generation and Display**
   Based on the retrieval result, the system dynamically generates an informative response in English, presenting structured data that combines both traditional linguistic detail and user-friendly narrative.

This methodology enables *Nirukti* to function as a bridge between classical Saṁskṛta linguistics and modern-day language interfaces, making the exploration of ancient names both intuitive and educational.

## Results & Discussion

The current implementation of *Nirukti – The Saṁskṛta NLP* demonstrates a fully functional, interactive platform that integrates multiple linguistic layers of selected Sanskrit names into a unified system. The application presently focuses on the 108 names of Śrī Viṣṇu (Viṣṇu Aṣṭottaraśatanāma), delivering accurate grammatical and semantic analysis based on user input.

### *5.1 Usability and Interface Design*

The front-end is built using Flask, providing a clean, responsive web interface that seamlessly adjusts across devices and screen sizes. Users can engage with the system using natural English queries, making the platform accessible to both scholars and novices. The interaction mimics a chatbot-like conversation, facilitating intuitive dialogue rather than technical command inputs.

### *5.2 Intelligent Name Recognition and Fuzzy Matching*

To accommodate the diverse spellings and transliterations of Sanskrit names, a custom mapping and fuzzy matching algorithm has been implemented. For example, user inputs such as *Vishnu*, *Visnu*, or *Vihsnu* are all accurately recognized and mapped to *Viṣṇu*. The system returns either a full match, a partial match with confirmation prompts, or a graceful error message if no relevant name is found. This significantly lowers the entry barrier for users unfamiliar with diacritical conventions.

### 5*.3 Natural Language Query Processing*

The application uses SpaCy for natural language processing to analyze English input and extract intent. The backend identifies linguistic cues from expressions such as *"explain the meaning"*, *"derive the name"*, or *"find names related to water"*. This processed input is then used by a generate_response() function to return structured, linguistically rich information, including:

- **Vigrahavākya (grammatical derivation)**
- **Artha (meaning)**
- **Prakriyā (etymology)**
- **Associated Mantra**
- **Positional Index**

### *5.4 System Limitations*

At present, the knowledge base is restricted to the curated dataset of 108 Viṣṇu names. While the responses within this scope are accurate and detailed, the system cannot currently handle words or names beyond this predefined dictionary. Expansion of the database will enhance its utility for broader Sanskrit lexical studies.

### *5.5 Comparative Advantages*

Traditional Sanskrit study often requires consulting multiple disconnected sources—such as dhātu-kośas, grammar treatises, and stotra texts—to understand a single term. *Nirukti* centralizes all this information in a unified view, delivering a consolidated output that combines grammatical, semantic, and scriptural dimensions. This streamlines the learning process and supports deeper linguistic exploration, especially for learners without formal training in Vyākaraṇa.

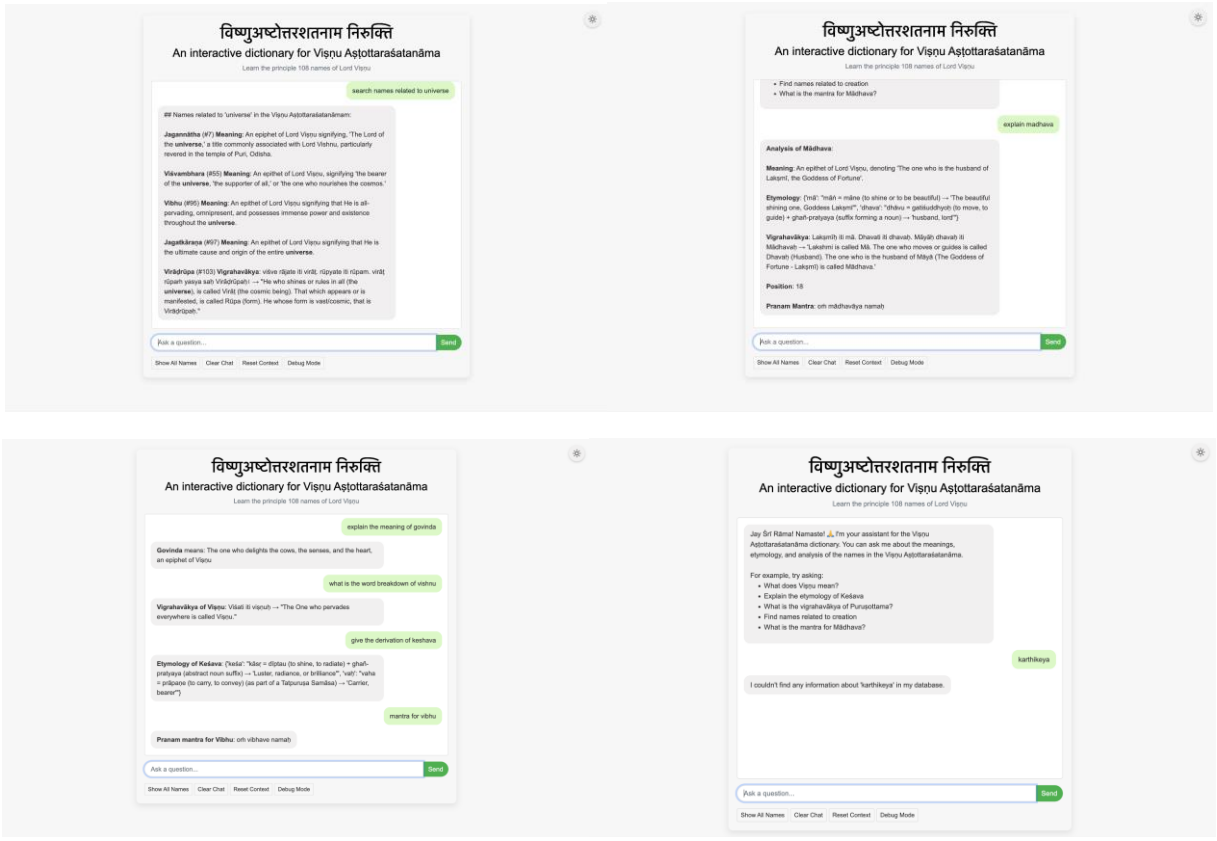


**Fig. 1 - Interface of Nirukti showing different queries**

## 4. Conclusion

Nirukti stands at the intersection of traditional linguistic knowledge and modern technological innovation, offering a transformative platform for the analysis and understanding of Sanskrit vocabulary through etymological, grammatical, and semantic insights. By integrating computational linguistics, structured data representation, and natural language processing, the system delivers a precise, accessible, and pedagogically robust tool for scholars, students, and enthusiasts alike.

The project addresses a significant gap in the digital Sanskrit landscape by unifying fragmented resources into a coherent and interactive system. Through intelligent query interpretation, fuzzy matching, and conversational interaction, Nirukti simplifies the complexity of classical grammar while preserving its depth and authenticity.

Looking ahead, Nirukti is envisioned to grow beyond its current corpus—centered on the Viṣṇu Aṣṭottaraśatanāma—toward broader textual domains within Sanskrit literature and philosophy. The incorporation of AI-powered features, community-driven content expansion, and integration with educational platforms will further strengthen its role as both a research tool and a learning companion.

With a strong commitment to linguistic integrity, usability, and scholarly rigor, Nirukti aspires to contribute meaningfully to the revitalization and global appreciation of Sanskrit—a language whose timeless wisdom continues to inspire intellectual and spiritual inquiry.

## REFERENCES


### Primary Textual Sources

1. Govindacharya, B. (n.d.). *Viṣṇu Sahasranāma* [Commentary].
2. SanskritDocuments.org. (n.d.). Retrieved from https://sanskritdocuments.org
3. Vignanam.org. (n.d.). Retrieved from https://vignanam.org
4. Stotranidhi.com. (n.d.). Retrieved from https://stotranidhi.com
5. DrikPanchang.com. (n.d.). Retrieved from https://drikpanchang.com
6. Ashtadhyayi.com. (n.d.). Dhātu derivations and Pāṇinian grammatical rules. Retrieved from https://ashtadhyayi.com

### Technological Tools and Libraries

1. GitHub. (n.d.). Open-source repositories and version control. Retrieved from https://github.com
2. Explosion AI. (n.d.). *spaCy: Industrial-Strength Natural Language Processing in Python*. Retrieved from https://spacy.io
3. Flask. (n.d.). *Flask Documentation*. Retrieved from https://flask.palletsprojects.com
4. Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media. (for NLTK)
5. OpenAI. (n.d.). *GPT language models*. Retrieved from https://openai.com
6. Google AI. (n.d.). *Gemini Language Models*. Retrieved from https://deepmind.google/technologies/gemini
7. Anthropic. (n.d.). *Claude AI*. Retrieved from https://www.anthropic.com