# Hrishikesh Rajesh Terdalkar

🏠 hrishikeshrt.github.io | 🔗 hrishikeshrt | ⭘ hrishikeshrt

## AFFILIATION

Postdoctoral Researcher
LIRIS Research Lab, Claude Bernard University Lyon 1

## SUMMARY

My research lies in the intersection of *Computational Linguistics*, *Natural Language Processing*, and *Software Engineering* with a particular focus on low-resource languages such as Sanskrit and other Indian languages. My focus is on addressing the unique challenges faced by these languages, enhancing language understanding, and advancing knowledge extraction from unstructured data. I am committed to leveraging my interdisciplinary background to pioneer innovations, with a strong emphasis on creating a real-world impact. My interests also include *Information Retrieval*, *Artificial Intelligence*, *Data Mining*, and *Machine Learning*.

## EDUCATION

**Ph.D. in Computer Science and Engineering**                                    2014 – 2023
*Indian Institute of Technology Kanpur*                          *Kanpur, Uttar Pradesh, India*
CGPA: 8.67

**M.Sc. in Computer Science**                                                    2011 – 2013
*Chennai Mathematical Institute*                                    *Chennai, Tamil Nadu, India*
CGPA: 8.88

**B.Sc. (Hons.) in Mathematics and Computer Science**                            2007 – 2010
*Chennai Mathematical Institute*                                    *Chennai, Tamil Nadu, India*
CGPA: 7.69

**Higher Secondary School Certificate (Standard XII)**                                2007
*Maharashtra State Board of Secondary and Higher Secondary Education*    *Pune, Maharashtra, India*
Percentage: 86.50

**Secondary School Certificate (Standard X)**                                         2005
*Maharashtra State Board of Secondary and Higher Secondary Education*    *Pune, Maharashtra, India*
Percentage: 95.33

## THESES

**Ph.D. Thesis**                                               Supervisor: Arnab Bhattacharya
*Sanskrit Knowledge-based Systems: Annotation and Computational Tools*

- Development of **knowledge-based systems** in **Sanskrit** with the ultimate goal of **question-answering** in Sanskrit. Contributions include a framework for the **automated construction** of **knowledge graphs**; two web-based **annotation tools**: (1) *Sangrahaka*: for **knowledge-graph focused annotation** and **querying**, and (2) *Antarlekhaka*: for general purpose **natural language annotation** of **multiple tasks** in a **sequential** manner; a system to identify **Sanskrit meters** and utilize Sanskrit prosody for **post-OCR correction**; several **web-interfaces** including three ancient **Indian alpha-numeric systems**; **tools** including a **Telegram bot** for Sanskrit grammar, and **software libraries** related to *computational aspects of Sanskrit*.

- All contributions available at **https://sanskrit.iitk.ac.in**

**M.Sc. Thesis**                                               Supervisor: Rajeeva Karandikar
*On Symmetric Key Cryptography*

- Study and implementation of various stream ciphers and block ciphers. (RC4, HC128, RC5, IDEA).

## PUBLICATIONS

- *"Antarlekhaka: A Comprehensive Tool for Multi-task Natural Language Annotation"* . **Hrishikesh Terdalkar**, Arnab Bhattacharya. In Proceedings of the 3rd Workshop for NLP Open Source Software (NLP-OSS) @ EMNLP, Singapore, December 2023. Association for Computational Linguistics (ACL).

- *"Chandojnanam: A Sanskrit Meter Identification and Utilization System"* . **Hrishikesh Terdalkar**, Arnab Bhattacharya. In Proceedings of the Computational Sanskrit & Digital Humanities: Selected papers presented at the 18th World Sanskrit Conference (WSC), pages 113–127, Canberra, Australia, January 2023. Association for Computational Linguistics (ACL).

- *"Semantic Annotation and Querying Framework based on Semi-structured Ayurvedic Text"* . **Hrishikesh Terdalkar**, Arnab Bhattacharya, Madhulika Dubey, Ramamurthy S, Bhavna Naneria Singh. In Proceedings of the Computational Sanskrit & Digital Humanities: Selected papers presented at the 18th World Sanskrit Conference (WSC), pages 155–173, Canberra, Australia, January 2023. Association for Computational Linguistics (ACL).

- *"A Novel Multi-Task Learning Approach for Context-Sensitive Compound Type Identification in Sanskrit"* . Jivnesh Sandhan, Ashish Gupta, **Hrishikesh Terdalkar**, Tushar Sandhan, Suvendu Samanta, Laxmidhar Behera, Pawan Goyal. In Proceedings of the 29th International Conference on Computational Linguistics (COLING), pages 4071–4083, Gyeongju, Republic of Korea, October 2022. International Committee on Computational Linguistics.

- *"Sangrahaka: A Tool for Annotating and Querying Knowledge Graphs"* . **Hrishikesh Terdalkar**, Arnab Bhattacharya. In Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE), pages 1520–1524, Athens, Greece, August 2021. Association for Computing Machinery (ACM).

- *"Framework for Question-Answering in Sanskrit through Automated Construction of Knowledge Graphs"* . **Hrishikesh Terdalkar**, Arnab Bhattacharya. In Proceedings of the 6th International Sanskrit Computational Linguistics Symposium (ISCLS), pages 97–116, IIT Kharagpur, India, October 2019. Association for Computational Linguistics (ACL).

## Talks, Posters and Demonstrations

- *"Vedapāṭhapaddhati: A Tool to Generate Prakṛtipāṭha and Vikṛtipāṭha forms of Sanskrit Text"* . **Hrishikesh Terdalkar**, Arnab Bhattacharya. The 7th International Sanskrit Computational Linguistics Symposium, 2024.

- *"Āyurjñānam: Exploring Āyurveda with Knowledge Graphs"* . **Hrishikesh Terdalkar**, Vishakha Deulgaonkar, Arnab Bhattacharya. National Youth Conference on Indian Knowledge Systems (NYCIKS), 2023.
  **Best Poster Award**.

- *"Vaiyyākaraṇaḥ: A Sanskrit Grammar Bot for Telegram"* . **Hrishikesh Terdalkar**, Mahesh A V S D S, Shubhangi Agarwal, Arnab Bhattacharya. The 18th World Sanskrit Conference (WSC), 2023.

- *"PyCDSL: A Programmatic Interface to Cologne Digital Sanskrit Dictionaries"* . **Hrishikesh Terdalkar**, Arnab Bhattacharya. The 18th World Sanskrit Conference (WSC), 2023.

- *"Jñānasaṅgrahaḥ: A Collection of Computational Applications related to Sanskrit"* . **Hrishikesh Terdalkar**, Arnab Bhattacharya. The 18th World Sanskrit Conference (WSC), 2023.

- *"KaTaPaYadi System"* . **Hrishikesh Terdalkar**, Arnab Bhattacharya. The 6th International Sanskrit Computational Linguistics Symposium (ISCLS), 2019.

## Experience

**Postdoctoral Researcher**                                                                 May 2024 – current
*LIRIS Research Lab, Claud Bernard University Lyon 1*

- Research: Large Language Models for Knowledge Graph Repair
- Project: 'Human Factors in Data-Intensive Applications for Health' (CPJ Data4Health).

**Project Scientist**                                                                             Jan 2024 – Apr 2024
*CSE, IIT Kanpur*

- Project: 'An Interpretable Unified Framework Text-to-text Translation among Indian Languages using Sanskrit-based Interlingua Representation'.
- Funding Agency: Ministry of Electronics and Information Technology (MeitY), Government of India.

**Senior Student Research Associate**                                                 Aug 2022 – Dec 2023
*CSE, IIT Kanpur*

- Project: 'An Interpretable Unified Framework Text-to-text Translation among Indian Languages using Sanskrit-based Interlingua Representation'.
- Funding Agency: Ministry of Electronics and Information Technology (MeitY), Government of India.

**Web Chair**                                                                                          Jan 2023 – Feb 2024
- Created and managed the website and the proceedings for the conference ISCLS 2024.

**Webmaster**                                                                                          Dec 2014 – Jan 2016
- Managed website of CSE Department, IIT Kanpur.

**Project Associate**                                                                                  Jul 2013 – Apr 2014
- Worked on the project 'Localization in Wireless Sensor Networks' at CSE, IIT Kanpur.

## TEACHING

**Senior Tutor**                                                                                       Jan 2017 – Nov 2020

- Managed (Prutor) backend administration for 5 semesters at IIT Kanpur.
- Duties also involved heading the tutor team, setting question papers and grading.

**Teaching Assistant**                                                                                 Aug 2014 – Nov 2020

- CS315: Principles of Database Systems (Jan-Apr 2019, Jan-Apr 2021)
- CS685: Data Mining (Aug-Nov 2018)
- CS774: Optimization Techniques (Aug-Nov 2016)
- ESC101: Introduction to Programming (Aug-Nov 2014)

**Teaching Assistant (MOOC)**                                                                          May 2017 – Sep 2017
*NPTEL*

- NOC17-CS33: Fundamentals of Database Systems

## ACHIEVEMENTS

- **Best Software Award** at the 57th Convocation, Indian Institute of Technology Kanpur, 2024.
- **Best Poster Award** at National Youth Conference on Indian Knowledge Systems 2023 **NYCIKS 2023** (Roorkee).
- **Financial support** from Research-I Foundation to attend **ISCLS 2019** (Kharagpur) and **WSC 2023** (Online).
- **Financial support** under ACM SIGSOFT CAPS Virtual Program to attend **ESEC/FSE 2021**.
- **Travel grants** from ACM India to attend **CoDS-COMAD 2017** (Goa), **CoDS-COMAD 2018** (Kolkata) and **CoDS-COMAD 2019** (Hyderabad).
- **TCS Research Scholar Program** Fellowship (*Cycle 11*), 2015.
- **Kishore Vaigyanik Protsahan Yojana (KVPY)** Fellowship, given by Indian Institute of Sciences (IISc), 2007.
- **National Talent Search (NTS)** Scholarship, given by National Council of Educational Research and Training (NCERT), 2005.

## TECHNICAL SKILLS

| | | |
|---|---|---|
| **Languages** | : | Python, C, C++, Shell Scripting, Haskell, R, PHP, JavaScript, HTML, CSS |
| **Databases** | : | MySQL, Neo4j, MongoDB |
| **Libraries** | : | Flask, SQLAlchemy, Pandas, Scikit-learn, Gensim, NLTK, spaCy, PyTorch, NumPy |
| **Dev Tools** | : | git, Python packaging, Make, docker |
| **Typesetting** | : | LaTeX, Markdown, reStructuredText |

## PROJECTS

**GitHub Profile Statistics**     ★ *140+*   *45+*   *2800+*     hrishikeshrt

**PyPI Profile**     hrishikeshrt

---

**Sanskrit Language Model**     *ongoing*

- Train language models for Sanskrit using a large corpus

**Antarlekhaka**     Antarlekhaka/code

- Annotation of a comprehensive set of computational linguistic tasks.
- General purpose, multi-task, sequential annotation.

**Sangrahaka**     hrishikeshrt/sangrahaka

- Annotation towards creation of knowledge graphs.
- Querying of knowledge graphs using natural language query templates.

**Jnanasangraha**     https://sanskrit.iitk.ac.in/jnanasangraha

- Collection of several web-based Sanskrit computational tools.
- Includes three of ancient Indian alpha-syllabic numeral systems, a Sankrit prosodical meter identification and utilization system and various utility functions related to Sanskrit alphabet.

**Vaiyyakarana**     vyakarana_bot

- Telegram bot for the learners of Sanskrit grammar.

**Sanskrit Audio Alignment**     https://hrishikeshrt.github.io/audio_alignment

- Interface for Sanskrit corpus recitation using forced-alignment to align text and audio.

## EXTRACURRICULAR

- I like to provide support to students who require academic assistance and have volunteered as an academic mentor at CSE, IIT Kanpur.
- I am a frequent blood donor with 45 donations so far and actively promote blood donation. I have also served as the Overall Co-ordinator and PG Wing Head of *Raktarpan*, IIT Kanpur and helped organize more than 20 blood donation camps in and around campus.
- I am interested in Sanskrit language and grammar. I hold a Diploma in Sanskrit Language under the Non-Formal Sanskrit Education programme offered by Central Sanskrit University. I have also completed Kovida, an advanced-level correspondence course offered by Samskrita Bharati.
- I love bicycling and have completed multiple 50km and 100km rides.

## REFERENCES

**Arnab Bhattacharya**     arnabb@iitk.ac.in, arnabb@cse.iitk.ac.in
*Professor, Indian Institute of Technology Kanpur*     +91-512-259-7650
https://cse.iitk.ac.in/users/arnabb/

**Andrea Mauri**     andrea.mauri@liris.cnrs.fr
*Junior Professor, Claude Bernard University Lyon 1*     +33-472-43-1636
https://www.andreamauri.me/

**Pawan Goyal**     pawang@cse.iitkgp.ac.in
*Associate Professor, Indian Institute of Technology Kharagpur*     +91-3222-282370
https://cse.iitkgp.ac.in/~pawang/

**Amba Kulkarni**     apksh@uohyd.ernet.in, ambapradeep@gmail.com
*Professor, University of Hyderabad*     +91-40-23133800
https://sanskrit.uohyd.ac.in/faculty/amba/

**Piyush Rai**     piyush@cse.iitk.ac.in
*Associate Professor, Indian Institute of Technology Kanpur*     +91-512-259-6894
https://cse.iitk.ac.in/users/piyush/