

# Hrishikesh Rajesh Terdalkar

🏠 [hrishikeshrt.github.io](https://hrishikeshrt.github.io) | [in](#) [hrishikeshrt](#) | [G](#) [hrishikeshrt](#)

## CURRENT AFFILIATION

Data Engineer  
BharatGen, Indian Institute of Technology Kanpur, INDIA

## SUMMARY

My research lies at the intersection of *Computational Linguistics*, *Natural Language Processing*, and *Graph Databases*, with a focus on low-resource languages like Sanskrit and other Indian languages. I aim to tackle the unique linguistic and computational challenges they present, advancing language understanding and knowledge extraction from unstructured data. Drawing on my interdisciplinary background, I'm committed to build impactful, real-world solutions. My broader interests include *Databases*, *Software*, *Information Retrieval*, *Artificial Intelligence*, *Human-Computer Interaction*, and *Data Mining*.

## EDUCATION

<b>Ph.D. in Computer Science and Engineering</b> <i>Indian Institute of Technology Kanpur</i> CGPA: 8.67	2014 – 2023 Kanpur, Uttar Pradesh, India
<b>M.Sc. in Computer Science</b> <i>Chennai Mathematical Institute</i> CGPA: 8.88	2011 – 2013 Chennai, Tamil Nadu, India
<b>B.Sc. (Hons.) in Mathematics and Computer Science</b> <i>Chennai Mathematical Institute</i> CGPA: 7.69	2007 – 2010 Chennai, Tamil Nadu, India
<b>Higher Secondary School Certificate (Standard XII)</b> <i>Maharashtra State Board of Secondary and Higher Secondary Education</i> Percentage: 86.50	2007 Maharashtra, India
<b>Secondary School Certificate (Standard X)</b> <i>Maharashtra State Board of Secondary and Higher Secondary Education</i> Percentage: 95.33	2005 Maharashtra, India

## THESES

<b>Ph.D. Thesis</b> <i>Sanskrit Knowledge-based Systems: Annotation and Computational Tools</i>	Supervisor: Arnab Bhattacharya
<ul style="list-style-type: none"><li>Development of <i>knowledge-based systems for Sanskrit</i> with the ultimate goal of question-answering.</li><li>Contributions include: a framework for the automated construction of knowledge graphs; two web-based annotation tools: (1) <i>Sangrahaka</i>: for knowledge-graph focused annotation and querying, and (2) <i>Antarলেখকা</i>: 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.</li><li>All contributions available at <a href="https://sanskrit.iitk.ac.in">https://sanskrit.iitk.ac.in</a></li></ul>	
<b>M.Sc. Thesis</b> <i>On Symmetric Key Cryptography</i>	Supervisor: Rajeeva Karandikar
<ul style="list-style-type: none"><li>Study and implementation of various stream ciphers and block ciphers. (RC4, HC128, RC5, IDEA).</li></ul>	

## PUBLICATIONS

- “Graph Repairs with LLMs: An Empirical Study” . **Hrishikesh Terdalkar**, Angela Bonifati, Andrea Mauri. **TO APPEAR** In Proceedings of the 8th Joint Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics (NDA) @ SIGMOD/PODS 2025, Berlin, Germany, June 2025. Association for Computing Machinery (ACM).

- “*Aganittiyam: Learning Tamil Grammar through Knowledge Graph based Templatized Question Answering*” . Mithilesh K, Amarjit Madhumalararungeethayan, Dharanish Rahul S, Abhijith Balan, C Oswald, **Hrishikesh Terdalkar**. In Proceedings of the 38th Pacific Asia Conference on Language, Information and Computation (PACLIC), Tokyo, Japan, December 2024. Association for Computational Linguistics (ACL).
- “*Samanvaya: An Interlingua for Unity of Indian Languages*” . Chaitali Dangarikar, Arnab Bhattacharya, Karthika N J, **Hrishikesh Terdalkar**, Pramit Bhattacharyya, Annarao Kulkarni, Chaitanya S Lakkundi, Ganesh Ramakrishnan, Shivani V. ISBN: 978-93-48435-21-7. October 2024. Central Sanskrit University. [\[Book\]](#)
- “*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.
- “*Jñānasāgrahaḥ: 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.

## TEACHING

---

### Instructor

Feb 2025 – Apr 2025

École Normale Supérieure de Lyon

- Program: Master 1 Informatique (ICA)
- Course: Databases and Data Mining

### Instructor

Jan 2025 – Mar 2025

Claude Bernard University Lyon 1

- Program: M2: Data and Intelligence for Smart Systems (DISS)
- Course: Natural Language Processing

### 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 (MOOC)

May 2017 – Sep 2017

NPTEL

- NOC17-CS33: Fundamentals of Database Systems

### 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)

### Academic Mentor

Jul 2018 – Apr 2020

- Provided support to students in need of academic assistance

## EXPERIENCE

### Data Engineer

BharatGen

May 2025 – present

### Postdoctoral Researcher

LIRIS Research Lab, Claude Bernard University Lyon 1

May 2024 – Apr 2025

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

### Project Scientist

CSE, IIT Kanpur

Jan 2024 – Apr 2024

- 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

CSE, IIT Kanpur

Aug 2022 – Dec 2023

- 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.

### Project Associate

Jul 2013 – Apr 2014

- Worked on the project 'Localization in Wireless Sensor Networks' at CSE, IIT Kanpur.

## ADMINISTRATIVE EXPERIENCE

### Program Committee Member

2025 – present

- GRADES-NDA Workshop @ SIGMOD/PODS (2025), ICWE (2025), TGD Workshop @ EDBT/ICDT (2025)

### Reviewer

2022 – present

- ARR (2022, 2023, 2024, 2025)

### Web Chair

Jan 2023 – Feb 2024

- Created and managed the website and the proceedings for the conference ISCLS 2024.

### External Reviewer

2020-2023

- CoDS-COMAD (2020, 2021, 2023), CIKM (2021, 2022), AAI (2022), DASFAA (2022)

### Webmaster

Dec 2014 – Jan 2016

- Managed website of CSE Department, IIT Kanpur.

## PROJECTS

### NLTM Interlingua

ongoing

- Interlingua-based Translation among Indian Languages.
- Bhasini Consortium Project. MeitY.

## 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.

## TECHNICAL SKILLS

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

## RESEARCH SOFTWARE AND TOOLS

### GitHub Profile Statistics

★ 150+ 🍴 45+ 🔄 3000+

[hrishikeshrt](#)

### PyPI Profile

[hrishikeshrt](#)

### 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 Sanskrit 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](https://hrishikeshrt.github.io/audio_alignment)

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

## EXTRACURRICULAR

- 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**

*Professor, Indian Institute of Technology Kanpur, India*

🏠 <https://cse.iitk.ac.in/users/arnabb/>

✉ [arnabb@iitk.ac.in](mailto:arnabb@iitk.ac.in), [arnabb@cse.iitk.ac.in](mailto:arnabb@cse.iitk.ac.in)

☎ +91-512-259-7650

### **Angela Bonifati**

*Distinguished Professor, Claude Bernard University Lyon 1, France*

🏠 <https://perso.liris.cnrs.fr/angela.bonifati/>

✉ [angela.bonifati@liris.cnrs.fr](mailto:angela.bonifati@liris.cnrs.fr)

☎ +33-472-44-5824

### **Andrea Mauri**

*Junior Professor, Claude Bernard University Lyon 1, France*

🏠 <https://www.andreamauri.me/>

✉ [andrea.mauri@liris.cnrs.fr](mailto:andrea.mauri@liris.cnrs.fr)

☎ +33-472-43-1636

### **Pawan Goyal**

*Associate Professor, Indian Institute of Technology Kharagpur, India*

🏠 <https://cse.iitkgp.ac.in/~pawang/>

✉ [pawang@cse.iitkgp.ac.in](mailto:pawang@cse.iitkgp.ac.in)

☎ +91-3222-282370

### **Amba Kulkarni**

*Professor, University of Hyderabad, India*

🏠 <https://sanskrit.uohyd.ac.in/faculty/amba/>

✉ [apksh@uohyd.ernet.in](mailto:apksh@uohyd.ernet.in), [ambapradeep@gmail.com](mailto:ambapradeep@gmail.com)

☎ +91-40-23133800