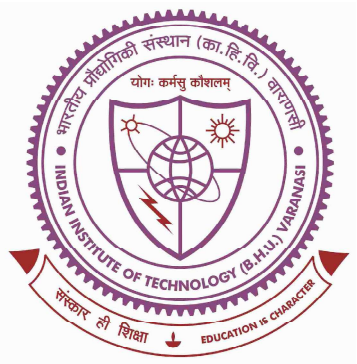
Pre-processing in Indian language IR



Thesis submitted in partial fulfilment
for the award of degree

Doctor of Philosophy

by

Siba Sankar Sahu

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
Indian Institute of Technology
(Banaras Hindu University)
Varanasi

Roll No: 17071012

Year of Submission: 2024

DECLARATION

I, **Siba Sankar Sahu**, certify that the work embodied in this thesis is my own bona fide work and carried out by me under the supervision of **Dr. Sukomal Pal** from July-2017 to April-2023, at the Department of Computer Science and Engineering, Indian Institute of Technology (BHU), Varanasi. The matter embodied in this thesis has not been submitted for the award of any other degree/diploma. I declare that I have faithfully acknowledged and given credits to the research workers wherever their works have been cited in my work in this thesis. I further declare that I have not willfully copied any other's work, paragraphs, text, data, results, etc., reported in journals, books, magazines, reports dissertations, theses, etc., or available at websites and have not included them in this thesis and have not cited as my own work.

Date:

Place: Varanasi

S. Sankar Sahu

Siba Sankar Sahu

CERTIFICATE BY THE SUPERVISOR

It is certified that the above statement made by the student is correct to the best of my/our knowledge.

S Pal

Dr. Sukomal Pal
पर्यवेक्षक/Supervisor
IIT(BHU), Varanasi
Department of Computer Sc. & Engg
भारतीय प्रौद्योगिकी संस्थान
Indian Institute of Technology
(काशी हिन्दू विश्वविद्यालय)
(Banaras Hindu University)
वाराणसी/Varanasi-221005

Signature of Head of Department/Coordinator of School

आचार्य व विभागाध्यक्ष
Professor & Head
संगणक विज्ञान एवं अभियांत्रिकी विभाग
Department of Computer Sc. & Engg
भारतीय प्रौद्योगिकी संस्थान
Indian Institute of Technology
(बनारस हिन्दू यूनिवर्सिटी)
(Banaras Hindu University)
Varanasi-221005

To

The

Supreme Power God

and

My beloved family

CERTIFICATE

It is certified that the work contained in the thesis titled **Pre-Processing in Indian Language IR** by **Siba Sankar Sahu** has been carried out under my supervision and that this work has not been submitted elsewhere for a degree. It is further certified that the student has fulfilled all the requirements of Comprehensive Examination, Candidacy and SOTA for the award of Ph.D. Degree.

SPal 7/6/24

Supervisor

Dr. Sukomal Pal

Associate Professor

Department of Computer Science and Engineering

Indian Institute of Technology (BHU) Varanasi

Varanasi, INDIA, 221005

CERTIFICATE

This is to certify that the revised thesis titled **Pre-Processing in Indian Language IR** is being submitted by **Siba Sankar Sahu** in partial fulfilment for the award of Ph.D. in Department of Computer Science and Engineering, IIT (BHU), Varanasi is a record of bonafide work carried out by him.

Date of Submission: 16/03/2024

SPal 7/6/24

Supervisor

Dr. Sukomal Pal

Associate Professor

Department of Computer Science and Engineering

Indian Institute of Technology (BHU) Varanasi

Varanasi, INDIA, 221005

COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis: **Pre-Processing in Indian Language IR**
Name of Student: **Siba Sankar Sahu**

Copyright Transfer

The undersigned hereby assigns to the Institute of Technology (Banaras Hindu University) Varanasi all rights under copyright that may exist in and for the above thesis submitted for the award of the Doctor of Philosphy.

Date: 16/03/2024
Place: Varanasi

S. S. Sahu
Siba Sankar Sahu

Note: However, the author may reproduce or authorize others to reproduce material extracted verbatim from the thesis or derivative of the thesis for author's personal use provided that the source and the Institute's copyright notice are indicated.

Acknowledgments

Though only my name appears on the cover of this dissertation, so many great people have contributed to its production. I owe my gratitude to all who have made this thesis possible and because of whom my post-graduate experience has been one I will cherish forever.

I take this opportunity to express my profound gratitude and deep regards to my supervisor Dr Sukomal Pal, Associate Professor, Department of Computer Science and Engineering, IIT(BHU), Varanasi, for his exemplary guidance, monitoring and constant encouragement throughout this dissertation.

I am obliged to the faculty members of the Department of Computer Science and Engineering, IIT(BHU), Varanasi, in particular, Prof. Sanjay Kumar Singh and staff members Prof. Anil Kumar Tripathi, Dr Anil Kumar Singh, Dr Pratik Chattopadhyay and Dr Sandip Ghosh, Associate Professor, Department of Electrical Engineering, IIT(BHU), Varanasi. They have provided constant support during my research program.

A very special gratitude goes out to my colleagues and friends, with special mention to Dr Jayraj Singh, Dr Tribikram Pradhan, Dr Pradeepika Verma, Dr Chintoo Kumar, Dr Anita Saroj, Mr Supriya Chanda, Mr Preetam Pal, Mr Biswajit Parmanik, Mr Dinesh Pargai, and Mrs Akanksha Mishra. In addition, I am grateful to Mr Sushil Kulkarni, who has provided me with moral and emotional support during my research program.

I extend special thanks to the non-teaching staff in the department, particularly Mr Ravi Kumar Bharti, Mr Shubham Pandey, Mr Prakhar Kumar, Mr Manoj Kumar Rai, Mr Viplav Biswas, and Mr Akhilesh Kumar Pal.

I thank especially my parents, Mr Budhia Sahu and Mrs Bideshi Sahu, my uncle and aunt, Mr Ananda Sahu and Mrs Santilata Sahu, my sister Priyadarshini and my brother Subrat, Subhasish, Debasish for their constant support and encouragement, without which this assignment would not have been completed at all. I am grateful to my other family

members who have supported me.

I take this opportunity to sincerely acknowledge the Ministry of Human Resource Development, Government of India, for providing a PhD Fellowship for financial assistance. It is said that it is no coincidence whom you meet on the journey. I have had the fortune to meet many wonderful people on the journey, capable and ready to guide, help, and advise me. I sincerely thank all of them who contributed to helping me to see the light at the end of every scary tunnel during my PhD.

Finally, I bow in great reverence to almighty God, the most gracious, the most merciful, whose bounteous blessings enabled me to accomplish this thesis.

Date: _____ Siba Sankar Sahu

List of Tables

List of Figures