

**M.Tech Thesis Defense**

# **Normalization of Spelling Variations in Code-Mixed Data**

**ON**

**19th August @ 2:30 PM IST**

**Advisors: Dr. Shad Akhtar,  
Dr. Tanmoy Chakraborty**

**Examiners:**

- 1. Dr. Mukesh Mohania,  
Professor, IIIT Delhi, India.**
- 2. Dr. Koteswar Rao Jerripothula,  
Assistant Prof. IIIT Delhi,  
India.**

**Krishna Yadav**

Please visit  
<https://cse.iiitd.ac.in/events-seminars/> for  
more details

## Abstract

Code-mixed text infused with low resource language has always been a challenge for natural language understanding models. A significant problem while understanding such texts is the correlation between the syntactic and semantic arrangement of words. The phonemes of each character in a word dictates the spelling representation of a term in low resource language. However, there is no universal protocol or alphabet mapping for code-mixing. In this paper, we highlight the impact of spelling variations in code-mixed data for training natural language understanding models. We emphasize the impact of using phonetics to neutralize this variation in spelling across different usage of a word with the same semantics. The proposed approach is a computationally inexpensive technique and improves the performances of state-of-the-art models for three dialog system tasks viz. intent classification, slot-filling, and response generation. We propose a data pipeline for normalizing spelling variations irrespective of language.