1st Workshop on Challenges in Processing South Asian Languages (CHiPSAL 2025)

Abu Dhabi, United Arab Emirates
19 January 2025

ISBN: 979-8-3313-1360-9

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2025) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2026)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

A Brief Overview of the First Workshop on Challenges in Processing South Asian Languages (CHiPSAL)
Kengatharaiyer Sarveswaran, Surendrabikram Thapa, Sana Shams, Ashwini Vaidya and Bal Krishna Bal 1

Development of Pre-Trained Transformer-based Models for the Nepali Language
Prajwal Thapa, Jinu Nyachhyon, Mridul Sharma and Bal Krishna Bal 9

Benchmarking the Performance of Pre-trained LLMs across Urdu NLP Tasks
Munief Hassan Tahir, Sana Shams, Layba Fiaz, Farah Adeeba and Sarmad Hussain 17

Bengali ChartSumm: A Benchmark Dataset and study on feasibility of Large Language Models on Bengali Chart to Text Summarization
Nahida Akter Tanjila, Afrin Sultana Poushi, Sazid Abdullah Farhan, Abu Raihan Mostofa Kamal, Md. Azam Hossain and Md. Hamjajul Ashmafee 35

DweshVaani: An LLM for Detecting Religious Hate Speech in Code-Mixed Hindi-English
Varad Srivastava 46

Improving Accuracy of Low-resource ASR using Rule-Based Character Constituency Loss (RBCCL)
Rupak Raj Ghimire, Prakash Poudyal and Bal Krishna Bal 61

SiTa - Sinhala and Tamil Speaker Diarization Dataset in the Wild
Uthayasanker Thayasivam, Thulasithan Gnanenthiram, Shamila Jeewantha and Upeksha Jayawickrama 71

A Dual Contrastive Learning Framework for Enhanced Hate Speech Detection in Low-Resource Languages
Krishan Chavinda and Uthayasanker Thayasivam 81

Abstractive Summarization of Low resourced Nepali language using Multilingual Transformers
Prakash Dhakal and Daya Sagar Baral 90

Structured Information Extraction from Nepali Scanned Documents using Layout Transformer and LLMs
Aayush Neupane, Aayush Lamichhane, Ankit Paudel and Aman Shakya 100

Domain-adaptative Continual Learning for Low-resource Tasks: Evaluation on Nepali
Sharad Duwal, Suraj Prasai and Suresh Manandhar 110

POS-Aware Neural Approaches for Word Alignment in Dravidian Languages
Antony Alexander James and Parameswari Krishnamurthy 120

neDIOM: Dataset and Analysis of Nepali Idioms
Rhitabrat Pokharel and Ameeta Agrawal 126

Bridging the Bandwidth Gap: A Mixed Band Telephonic Urdu ASR Approach with Domain Adaptation for Banking Applications
Ayesha Khalid, Farah Adeeba, Najm Ul Sehar and Sarmad Hussain 138

Impacts of Vocoder Selection on Tacotron-based Nepali Text-To-Speech Synthesis
Ganesh Dhakal Chhetri and Prakash Poudyal 151

EmoTa: A Tamil Emotional Speech Dataset
Jubeerathan Thevakumar, Luxshan Thavarasa, Thanikan Sivatheepan, Sajeev Kugarajah and Uthayasanker Thayasivam 159

Benchmarking Whisper for Low-Resource Speech Recognition: An N-Shot Evaluation on Pashto, Punjabi, and Urdu
Najm Ul Sehar, Ayesha Khalid, Farah Adeeba and Sarmad Hussain 168

Leveraging Machine-Generated Data for Joint Intent Detection and Slot Filling in Bangla: A Resource-Efficient Approach
A H M Rezaul Karim and Ozlem Uzuner 174

Challenges in Adapting Multilingual LLMs to Low-Resource Languages using LoRA PEFT Tuning
Omkar Khade, Shruti Jagdale, Abhishek Phaltankar, Gauri Takalikar and Raviraj Joshi 183

Natural Language Understanding of Devanagari Script Languages: Language Identification, Hate Speech and its Target Detection
Surendrabikram Thapa, Kritesh Rauniyar, Farhan Ahmad Jafri, Surabhi Adhikari, Kengatharaiyer Sarveswaran, Bal Krishna Bal, Hariram Veeramani and Usman Naseem 189

Sandhi Splitting in Tamil and Telugu: A Sequence-to-Sequence Approach Leveraging Transformer Models
Priyanka Dasari, Mupparapu Sohan Gupta, Nagaraju Vuppala, Pruthwik Mishra and Parameswari Krishnamurthy 201

Bridge the GAP: Multi-lingual Models For Ambiguous Pronominal Coreference Resolution in South Asian Languages
Rahothvarman P, Adith John Rajeev, Kaveri Anuranjana and Radhika Mamidi 212

1-800-SHARED-TASKS@NLU of Devanagari Script Languages 2025: Detection of Language, Hate Speech, and Targets using LLMs
Jebish Purbey, Siddartha Pullakhandam, Kanwal Mehreen, Muhammad Arham, Drishti Sharma, Ashay Srivastava and Ram Mohan Rao Kadiyala 223

AniSan@NLU of Devanagari Script Languages 2025: Optimizing Language Identification with Ensemble Learning
Anik Mahmud Shanto, Mst. Sanjida Jamal Priya and Mohammad Shamsul Arefin 236

byteSizedLLM@NLU of Devanagari Script Languages 2025: Hate Speech Detection and Target Identification Using Customized Attention BiLSTM and XLM-RoBERTa Base Embeddings
Rohith Gowtham Kodali, Durga Prasad Manukonda and Daniel Iglesias 243

byteSizedLLM@NLU of Devanagari Script Languages 2025: Language Identification Using Customized Attention BiLSTM and XLM-RoBERTa base Embeddings
Durga Prasad Manukonda and Rohith Gowtham Kodali 249

CUET_Big_O@NLU of Devanagari Script Languages 2025: Identifying Script Language and Detecting Hate Speech Using Deep Learning and Transformer Model
Md. Refaj Hossan, Nazmus Sakib, Md. Alam Miah, Jawad Hossain and Mohammed Moshiul Hoque 254

CUET_HateShield@NLU of Devanagari Script Languages 2025: Transformer-Based Hate Speech Detection in Devanagari Script Languages
Sumaiya Rahman Aodhora, Shawly Ahsan and Mohammed Moshiul Hoque 261

CUET_INSights@NLU of Devanagari Script Languages 2025: Leveraging Transformer-based Models for Target Identification in Hate Speech
Farjana Alam Tofa, Lorin Tasnim Zeba, Md Osama and Ashim Dey 268

CUFE@NLU of Devanagari Script Languages 2025: Language Identification using fastText
Michael Ibrahim 274

Dll5143A@NLU of Devanagari Script Languages 2025: Detection of Hate Speech and Targets Using Hierarchical Attention Network
Ashok Yadav and Vrijendra Singh 279

DSLNLP@NLU of Devanagari Script Languages 2025: Leveraging BERT-based Architectures for Language Identification, Hate Speech Detection and Target Classification
Shraddha Chauhan and Abhinav Kumar 290

IITR-CIOL@NLU of Devanagari Script Languages 2025: Multilingual Hate Speech Detection and Target Identification in Devanagari-Scripted Languages
Siddhant Gupta, Siddh Singhal and Azmine Toushik Wasi 296

LLMsAgainstHate@NLU of Devanagari Script Languages 2025: Hate Speech Detection and Target Identification in Devanagari Languages via Parameter Efficient Fine-Tuning of LLMs
Rushendra Sidibomma, Pransh Patwa, Parth Patwa, Aman Chadha, Vinija Jain and Amitava Das 302

MDSBots@NLU of Devanagari Script Languages 2025: Detection of Language, Hate Speech, and Targets using MURTweet
Prabhat Ale, Anish Thapaliya and Suman Paudel 309

Nepali Transformers@NLU of Devanagari Script Languages 2025: Detection of Language, Hate Speech and Targets
Pilot Khadka, Ankit BK, Ashish Acharya, Bikram K.C., Sandesh Shrestha and Rabin Thapa . . 315

NLPineers@ NLU of Devanagari Script Languages 2025: Hate Speech Detection using Ensembling of BERT-based models
Nadika Poudel, Anmol Guragain, Rajesh Piryani and Bishesh Khanal 321

One_by_zero@ NLU of Devanagari Script Languages 2025: Target Identification for Hate Speech Leveraging Transformer-based Approach
Dola Chakraborty, Jawad Hossain and Mohammed Moshiul Hoque 328

Paramananda@NLU of Devanagari Script Languages 2025: Detection of Language, Hate Speech and Targets using FastText and BERT
Darwin Acharya, Sundeep Dawadi, Shivram Saud and Sunil Regmi 335

SKPD Emergency @ NLU of Devanagari Script Languages 2025: Devanagari Script Classification using CBOW Embeddings with Attention-Enhanced BiLSTM
Shubham Shakya, Saral Sainju, Subham Krishna Shrestha, Prekshya Dawadi and Shreya Khatiwada 340