Department of Computer Science & Engineering
Tezpur University

Natural Language Processing Laboratory

Natural Language Processing Laboratory (NLP-Lab), Department of Computer Science & Engineering at Tezpur University focuses on computational study of North-East Indian Languages. The North-East (seven-sister's state) India is linguistically very rich. There are languages belonging to all the four major language families spoken in the subcontinent: (a) Indo-European languages represented by Assamese, Bengali, Hindi and Nepali; (b) Tibeto-Burman languages represented by Bodo, Rabha, Mising and Deuri; (c) Austric language represented by Khasi; (d) a few number of Dravidian languages in the tea garden area.

Our goal is to use NLP to support both education and the application field. Our group currently focuses on syntactic, and semantic study of Assamese, Bengali, Bishnupriya Manipuri, Manipuri (Meitei) and Bodo languages, and phonetic study of Assamese.

This lab is assisted by SAP.

a. Equipment(s)

Sl. No.	Equipment (with specification) and Year of Purchase	Quantity
1	PCs	10
2	Workstation	1
3	Server	1
4	UPS	1
5	Switch Rack	1

b. Types of Experiments Conducted

Sl. No.	Experiments Conducted/Performed
1	Incorporating dialect features in synthesized speech, particularly in some Indian languages.
2	Development of speech corpus in Bengali, Assamese, and Mizo languages, for various speech technologies.
3	Study on speech prosody and handcrafting speech prosody features for languages like Assamese and Bengali.
4	Developing classification tools for Mizo folk music using acoustic features and creating a framework for Retrieval system.
5	Automatic phonetic transcription of Assamese speech.

6	Computational modelling of morphology and syntax of some Indian languages such as Assamese, Bishnupriya Manipuri and Manipuri(Meitei) languages.
7	Information extraction (IE) from Assamese texts.
8	Text Aggregation for Assamese Text.
9	Aspect Term Extraction and Sentiment Analysis for Hindi Language.

c. Open-source tools/software used:

Name of tool/software	Utility
Praat	Speech Annotation and Phonetic Analysis
Audacity	Speech recording, editing, analysis, and enhancement
Matlab(Trail)	Speech Signal Analysis and Processing (including feature extraction and ML tasks)
Python	Speech Signal Analysis and Processing (including feature extraction and ML tasks)
Kaldi	Development of ASR System and Speaker Diarization
Terrier	Text search and information retrieval platform.

d. People working in the Laboratory:

Research scholars and students doing projects in NLP work here.