

---

Seminar Title	: Sequence Labeling Tasks for Odia Language
Speaker	: Tusarkanta Dalai ( Rollno : 518cs1008)
Supervisor	: Dr. Tapas Kumar Mishra
Venue	: Convention Hall, CS Department
Date and Time	: 11 May 2024 (10:30)
Abstract	: The overarching objective is to contribute to the enhancement of NLP applications for Odia, particularly in the domains of Part-of-Speech (POS) tagging, Named Entity Recognition (NER), and chunking. The first objective of the research involves an in-depth investigation into the construction of annotated datasets tailored for sequence labeling tasks. Given the scarcity of linguistic resources and annotated data in low-resource languages like Odia, the methodology adopted for dataset creation is meticulous and resourceful. This corpus, precisely curated and annotated, spans various domains, text types, and linguistic nuances. Its creation involved extensive data collection efforts, linguistic analysis, and annotation by domain experts, resulting in a valuable resource for Odia language research. This stage serves as the foundational building block for subsequent developments, ensuring the availability of high-quality annotated data for training and evaluation purposes. The second phase of the thesis focuses on the development of systems for sequence labeling tasks, commencing with POS tagging. A baseline model is established using Conditional Random Fields (CRF) for the development of Odia POS tagger. Subsequently, the thesis explores advanced modeling techniques, incorporating Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM), and transformer models to refine and elevate the accuracy of the POS tagger. Each model is meticulously fine-tuned to the unique linguistic characteristics of the Odia language, offering a nuanced understanding of context and semantics. The third phase of the research extends the developed methodologies to the creation of a phrase chunking system. Leveraging the foundation laid by the annotated dataset and the insights gained from the POS tagging system, the chunking model is designed to capture syntactic structures and linguistic nuances specific to Odia. Similar to the POS tagging phase, CRF, CNN, LSTM, and transformer models are employed to iteratively enhance the results and adaptability of the chunking system. The final stage of the research culminates in the development of a Named Entity Recognition (NER) system. Drawing on the knowledge gained from the preceding phases, the NER system is crafted to identify and categorize named entities within Odia text. The utilization of diverse modeling approaches, including CRF, CNN, LSTM, and transformer models, ensures a comprehensive and nuanced understanding of named entities in the context of the Odia language.