

Introduction to NLP – Assamese Workshop

Department of Computer Science and Engineering, Tezpur University, Assam

11th to 15th of July, 2011

The workshop on introduction to NLP for Assamese was organized by LDC-IL in collaboration with the Department of Computer Science and Engineering, Tezpur University, Assam. The workshop was organized from the 11th to 15th of July, 2011, at the conference hall of the Department of Computer Science and Engineering. There were three hour sessions daily from 10 a.m. to 1 p.m. with a tea break of 15 minutes in between. There were totally 84 participants excluding the Department staff, participants and helpers. The participants were from various Departments, including Assamese, Engineering, Linguistics and Computer Science.

The first day started off with welcome speech delivered by the Head of the Department, Computer Science and Engineering, Dr.Utpal Sharma. The registrar of Tezpur University Dr. Choudhury, who was the chief guest for the inaugural session gave a speech on importance of NLP and this was followed by an introductory speech delivered by Atreyee Sharma. The workshop sessions started with a talk on Introduction to LDC-IL given by Atreyee Sharma which was followed by a tea-break before the next session on the Introduction to NLP by Amom Nandaraj Meetei.

The second day was devoted to an extensive lecture on Corpus linguistics and its relevance for the field of NLP by Atreyee Sharma. After tea, Assamese and Bodo corpus specific issues were discussed by Atreyee Sharma, Bridul Basumatary and Farson Daimary.

On the third day Atreyee Sharma talked on POS tagging in general where she also discussed the BIS Tag set of LDC-IL. This was followed by an extensive discussion on POS tagging issues in Assamese and Bodo. Atreyee Sharma, Bridul Basumatary and Farson Daimary illustrated and discussed the Assamese and Bodo specific examples with the participants.

The fourth day was devoted to Morphology and Morphological Analysers/Generators by Amom Nandaraj Meetei. After tea break Nandaraj continued with the lecture. The session ended with a demonstration of Apertium and Suffix stripping tools by Farson Daimary.

The fifth and the final day was scheduled for Speech and Speech Processing related work at LDC-IL. Atreyee Sharma handled the first session, where she discussed in details about speech processing in general and the method employed for speech data collection at LDC-IL in particular. Before tea session Bridul Basumatary gave a demonstration of the Speech segmentation and Annotation tools. This was followed by a demonstration of the Corpus related tools - Transliteration tool and KWIC-KWOC Tool etc. by Farson Daimary.

The workshop ended with a valedictory session after tea wherein the Head of the Department of Computer Science and Engineering, Dr.Utpal Sharma congratulated all on the success of the workshop with a proposal of having a refresher course on NLP in the future. Prof. S.K. Sinha, Dean, Tezpur University who was the chief guest for the valedictory session stressed on the importance of including NLP as a subject in the graduation level in Assam. Dr. Sharma, Prof Sinha and Atreyee Sharma distributed the certificates to all the participants. The workshop concluded with a feedback session and unanimous appeal by all the participants to have similar workshops again in the future. Finally, all the presenters of LDC-IL presented the spoke/gave the vote of thanks on behalf of LDC-IL and Padmaja Sarma presented the vote of thanks on behalf of Department of Computer Science and Engineering, Tezpur University.