

Urdu NLP Workshop Report (Linguistic Department, Lucknow University, Lucknow) 21st February to 02nd March, 2012

The 10 day NLP orientation programme was organized by LDC-IL in collaboration with Department of Linguistics, Lucknow University (LU), Lucknow, UP from 21st Feb. to 02nd March. The venue was the conference hall of the department. The programme coordinator from LU was Prof. Kavita Rastogi, the present H.O.D & the coordinator from LDC-IL was Shahid Mushtaq Bhat, Lecturer/RP LDC-IL. Different experts & resource persons/research scholars who delivered various lectures regarding Core Linguistics, Corpus Linguistics & NLP were Prof. Kavita Rastogi, Lucknow University (2 lectures), Prof. Sri Kumar, Lucknow University (2 lectures), Dr M.K. Koul, UTRC Lucknow (1 lecture), Dr. Anil Kumar Singh, School of Computer Engineering, KIIT University Bhubaneswar (6 lectures), Mr. Bharat Raju, LDC-IL (2 lectures), Mrs. Manasa, LDC-IL (2 lectures), Mr. Shahid Mushtaq Bhat, LDC-IL (8 lectures), Mr. Mansoor Khan, LDC-IL (2 lectures), Mr. Aju Samuel Thomas, LDC-IL (5 lectures), Dr. Shahnawaz Alam, LDC-IL (1 lecture) & Dr. Satyaendra Awasthi, LDC-IL (1 lecture). Besides, there were five special lectures delivered by a team of research scholars from Kashmir University.

Further, there were 60 participants from five different universities, namely- Lucknow University (LU), English & Foreign Languages University (FLU), Aligarh Muslim University (AMU), Banaras Hindu University (BHU), Central University of Himachal Pradesh & IGNOU.

Day-1 (21st Feb 2012)

The Orientation Programme began with an inaugural function which was somewhat informal of just 20 minutes in which Prof. Kavita Rastogi delivered a welcome speech. It was immediately followed by a reading out of "Orientation Programme goals" & introduction to LDC-IL. Then after tea session there was a lecture on "Language & Linguistics" by Prof. Kavita Rastogi. It was followed by a lecture on "Syntax: Phrases, Clauses, Sentences & their types" by Prof. Sri Kumar. In after lunch session, there were three lectures, one on "Morphology: Inflectional & Derivational" again by Prof. Kavita Rastogi, second one on "Semantics: Theories of Meaning" by Prof. Sri Kumar. Finally, after tea session ends with a lecture on "Corpus Linguistics & Corpus Annotation" by Shahid Mushtaq Bhat.

Day-2 (22nd Feb 2012)

The second day began with the lectures on "Introduction to Artificial Intelligence & Natural Language Processing" & "Approaches to NLP" by Shahid. It was followed by Dr. Anil Kumar Singh's lecture "How to Develop an NLP System" in post tea session. In post lunch sessions there were two special lectures; first one was "Psycholinguistics: Models of Spoken Word Recognition" by Tanveer Habib (Research Scholar Kashmir University). Second one was on "Writing Systems" by Irshad Ahmad (Research Scholar Kashmir University).

Day-3 (23rd Feb 2012)

The third day started with Dr. Anil Kumar Singh's lecture on "Basic Statistics: Frequency, Probability, Regression, Classification, Clustering, etc". After clearing the basic concepts of Statistics, he delivered two lectures on "Machine Learning Approaches: Decision Trees, HMM, SVM & Neural Networks" in post tea session. Finally he delivered a lecture on "Machine Translation" in post lunch session.

Day-4 (24th Feb 2012)

In fourth day, there was one lecture on "Concept of font: ASCII, ISCII, Unicode & Text Encoding: XML, SGML" by Bharat Raju. It was followed by a lecture on "Finite State Automata Theory: FSA & FST" by Dr. Anil Kumar Singh in post tea session. After this, Manasa delivered a lecture on "Spell Checker & Grammar Checker" & then she demonstrated "LDC-IL Tools" in post lunch session.

Day-5 (25th Feb 2012)

Proceedings of the fifth day started with the lecture of Dr. M. K. Koul on "Urdu Phonetics & Phonology." It was immediately followed by a talk on "Multi-Word-Expressions" by Neelofer Wani (Research Scholar Kashmir University). In post lunch session, there were two lectures. One on "Sociolinguistics: Dialect Survey" by Parvaiz Ahmad (Research Scholar Kashmir University) and other by Mansoor Khan on "Development of Urdu Speech Corpus"

Day- (26th Feb 2012)

Sunday was off (Holiday)

Day-6 (27th Feb 2012)

The first presentation of the day was by Dr. Satyaendra Awasthi on "Speech Segmentation & Annotation." The lecture in the post tea session was on "Acoustic Phonetics: Physical Properties of Speech Signal" by Aju Samuel Thomas. Aju Samuel Thomas also delivered next two lectures in post lunch session; first one was on "Automatic Speech Recognition (ASR)" & the second one was on "Text to Speech (TTS) System."

Day-7 (28th Feb 2012)

The seventh day started with the lecture on "Sign Language: ISL Corpus & ASLR" by Aju Samuel Thomas. It was followed by a lecture & demonstration of "Urdu Morph-Analyzer" by Dr. Shahnawaz Alam. In post lunch session one lecture on "POS Tagset: Concept, Applications, Development & Issues" was delivered by Shahid Mushtaq Bhat & other one on "Transliteration" was delivered by Mansoor Khan.

Day-8 (29th Feb 2012)

The day began with a lecture on "Language Pathology & Pathological Corpus" by Aju Thomas. It was a quite long lecture, covering both pre & post tea sessions. Post lunch sessions began with a lecture on "Shallow Parsing, Deep Parsing & Tree-banking" by Shahid Mushtaq Bhat. It was immediately followed by a talk/discussion on the activities/tasks given in the "NLP Workbook" by Shahid Mushtaq Bhat. The participants were supposed to perform the given activities next day.

Day-9 (01st Feb 2012)

The day nine was totally dedicated to NLP activities given in the "Workbook" which included Activity 1: Word Ordering, Activity 2: Morphological Analysis, Activity 2.1: Paradigm Formation, Activity 3: POS & NER, Activity 4: Word Sense Disambiguation, Activity 5: Named Entity Recognition Rules, Activity 6: Sentence Parsing, Activity 7: UNL Tags, Activity 8: Semantic Role Labeling, Activity 9: Text Categorization, Activity 10.1 a) Identify a title of the text. b) Summarize the text in one sentence. c) Summarize the text in three sentences. d) Identify top 5 keywords in the text e) Identify 10 most frequent words in the text. Activity 10.2 a) POS Tagging: Annotate the text with POS Tags b) Chunking: Group the words in Chunks & name each chunk. c) Parsing: Mark the relation between Noun chunks & the Finite Verb Chunks in each sentence. d) Discourse Analysis: Find the linkers between sentences. e) Tagset: Develop Tagset for the above passage. Activity 11: Arrange the NLP modules in Logical order and draw a labeled diagram.

Day-10 (02th Feb 2012)

In the valedictory function Prof. Sri Kumar extended vote of thanks to LDC-IL team, various experts & participants. It was followed certificate distribution by Prof. Kavita Rastogi. Participants were given chance to share their experiences & views on the success or fulfillment of objectives & goals of NLP Orientation Programme. Each participant shared his views. Finally, there was a small cultural programme by a couple of students.