



Orientation cum Training Programme on Natural Language Processing

31st January -10th February, 2012



Organised by

Linguistic Data Consortium for Indian Languages (LDC-IL),
Central Institute of Indian Languages (CIIL), Mysore



in collaboration with
Department of Punjabi,
Regional Campus (G.N.D.U.), Jalandhar





CENTRAL INSTITUTE OF INDIAN LANGUAGES

Department Of Higher Education

Ministry of Human Resource Development, Government of India

Manasagangotri, Mysore - 570 006



Linguistic Data Consortium for Indian Languages (LDC-IL)

Ten day

Orientation cum Training Programme on Natural Language Processing

in collaboration with

Department of Punjabi, Regional Campus (G.N.D.U.), Jalandhar
31st Jan.–10th Feb., 2012

The Linguistic Data Consortium for Indian Languages was established by the Department of Higher Education, Ministry of Human Resource Development, and Government of India during XIth Plan period in the Central Institute of Indian Languages, Mysore. It has a mission to provide "Annotated, quality language data (both-text & speech) and tools in Indian Languages to Individuals, Institutions and Industry for Research & Development - Created in house, through outsourcing and acquisition". The following are the objectives and areas of work:

Objectives:

- A repository of linguistic resources in all Indian languages in the form of text, speech and lexical corpora.
- Facilitating creation of such databases by different organizations.
- Setting standards for data collection and storage of corpora for different research and development activities.
- Supporting development and sharing of tools for data collection and management.
- Facilitating training through workshops, seminars etc. in technical as well as process related issues.
- Creating and maintaining the LDC-IL website that would be the primary gateway for accessing LDC-IL resources.
- Designing or providing help in creation of appropriate language technology for mass use.
- Providing the necessary linkages between academic institutions, individual researchers and the masses.

Major areas:

- Natural Language Processing
- Creation of different kinds of Corpora like Text, Speech and Sign Language.
- Speech Recognition and Synthesis
- Character Recognition
- By-products like, lexicons of different kind, thesauri, usage compilations etc.

Workshop Goals:

The goals of this "10-day NLP Orientation Cum Training Programme" are:

- To disseminate the knowledge of NLP among student community
- To equip students to work or pursue research in language technology
- To promote technology development in Indian Languages

Instructions:

- Aspirants (From Linguistic, Computer Science & Literature background) need to apply to the department of Punjabi. Preference will be given to research scholars (Linguistics) / junior lecturers.
- Only 25-30 selected candidates can participate in the programme.
- Selected candidates need to attend all the days otherwise they will not be entitled for certificates.
- The department is requested to submit the applications & resumes of all the applicants and attendance of the selected candidates to LDC-IL for processing.

Syllabus: (Total Lectures = 26/28)

Unit.1:- Core Linguistics [6 lectures]

- Morphology (Basics)
- Syntax (Basics)
- Phonetics (Acoustics; Physical Properties of Speech Signal)
- Semantics (Theories of Meaning & Componential Analysis (Ontology))
- Phonology (Basics)

Unit.2:- AI & NLP [4 lectures]

- Artificial Intelligence (AI) & NLP/CL
 - Basic Concept of AI & NLP/CL,
 - Importance of Linguistic Knowledge
- Approaches to NLP
 - Computational Grammar Approach,
 - Data Driven/Inductive Approach
- Machine Learning: Statistical Approaches (Decision Trees & Decision Lists, HMM, SVM, Neural Networks & Genetic Algorithms)

Unit.3:- Corpus Linguistics [6 lectures]

- Corpora (Text, Speech & Sign): Concept & Classification
- Encoding (Concept of Font & Encoding; ASCII, ISCII & Unicode)
- Balanced Corpus: Concept, Development & Challenges
- Linguistic knowledge & Corpus: Annotation & Extraction
- Corpus Utilities & Corpus analysis tools (Transliteration, Frequency, N-gram, KWIC-KWOC, Concordances, etc)

Unit.4:- Linguistic Analysis & NLP Modules [10 lectures]

- Morphological analysis (Developing Morph-Analyzer, Models; IA, IP & WP)
- POS annotation/tagging (Developing POS-Tagger)
- Local Word Grouping & Chunking (Developing Chunker/Shallow Parser)
- Syntactic Parsing (Deep Parser: Concept & Applications, Development & Challenges)
- Tree-banking (Issues in Design & Development)
- ASR: Concept, Applications, & Challenges
- TTS: Concept & Applications, Development & Challenges

Unit.5:- System Development and the Challenges [4 lectures]

- MT: Concept & Applications, Development & Challenges
- Word-Net: Concept & Applications, Development & Challenges
- OCR: Concept & Applications, Development & Challenges
- Spell-checker: Concept & Applications, Development & Challenges

Venue : Department of Punjabi, Regional Campus (G.N.D.U.), Jalandhar

Date & Time : 31st Jan.–10th Feb., 2012

Time : 10.00 am to 5.00 pm

Contact Persons:

1. *Local Coordinator* - **Prof. Sukhvinder Singh Sangha**
Head & Associate Dean
Department of Punjabi
Regional Campus, Jalandhar 144001
Mb. No. - +91 9417105225

2. *LDC-IL Coordinator* - **Poonam Dhillon**
LDC-IL, CIIL, Mysore.
E-mail: ldc-poonam@ciil.stpmy.soft.net,
Mb. No. - +91 8088227994



CENTRAL INSTITUTE OF INDIAN LANGUAGES

Department Of Higher Education
Ministry of Human Resource Development, Government of India
Manasagangotri, Mysore - 570 006



Linguistic Data Consortium for Indian Languages (LDC-IL)

WORKSHOP PROFORMA

Ten day Orientation cum Training Programme on Natural Language Processing

1.	Name	:	
2.	Date of Birth	:	
3.	Address	:	
4.	Contact Number	:	
5.	E-mail address	:	
6.	Current Designation	:	
7.	Details of current work (e.g. Projects/Thesis topic etc.,)	:	
8.	Details of Educational Qualification:		
	Degree	Year	Department
			University
9.	Familiarity with computers i. Have you used computers at all : Yes / No ii. If yes, what have you done using them	:	
10.	Areas of Interest	:	
11.	Any other relevant information	:	

Date:

(Signature of applicant)

Duly completed pro-forma may please be submitted to:

Prof. Sukhvinder Singh Sangha
Head & Associate Dean
Dept. of Punjabi
Regional Campus (G.N.D.U.)
Jalandhar.
Email: essangha@yahoo.com

PS: Hostel accommodation is available only to the female participants but they should have to bring winter bedding with them. Filling a form does not guarantee to be selected for attending the workshop.