

CENTRAL UNIVERSITY OF PUNJAB



Certificate Course in Computational Linguistics

Session - 2021-22

School of Languages, Literature and Culture

Graduate Attributes

After attaining this certificate course students will be able to acquire skills and applications of computational linguistics in various fields of Punjabi and Hindi languages.

Course Structure of the Programme

Semester - I

Course Code	Course Title	Course Type	L	T	P	Cr
C.L.201	Linguistics and Languages of India	CC	3	0	2	4
C.L.202	Introduction to Computational Linguistics	SDC	3	0	2	4
C.L.203	Fundamentals of Programming-Python	SDC	3	0	2	4
	Total Credits		9	0	6	12

Semester - II

Course Code	Course Title	Course Type	L	T	P	Cr
C.L 221	Corpora based Language processing	CC	3	0	2	4
C.L 222	Natural Language Processing using Python	SDC	0	0	6	3
C.L. 225	Project	SDC	0	0	12	6
	Total Credits		3	0	20	13

Evaluation Criteria for Theory Courses:

Core, Discipline Elective, Compulsory Foundation, Value Added and Interdisciplinary Courses		
	Marks	Evaluation
Internal Assessment	25	Various
Mid-semester test (MST)	25	Subjective
End-semester test (EST)	50	Subjective (70%) Objective (30%)

The objective type will include one word answers, fill-in the blank, sentence completion, true/false, MCQs', matching, analogies, rating and checklists. The subjective type will include very short answer (1-2 lines), short answer (one paragraph), essay type with restricted response, and essay type with extended response.

Evaluation Criteria for Internal Assessment will be designed by the concerned course coordinator using any two or more of the given methods (Surprise Tests, in-depth interview, unstructured interview, Jigsaw method, Think-Pair Share, Students Teams Achievement Division (STAD), Rubrics, portfolios, case based evaluation, video based evaluation, Kahoot, Padlet, Directed paraphrasing, Approximate analogies, one sentence summary, Pro and con grid, student generated questions, case analysis, simulated problem solving, media assisted evaluation, Application cards, Minute paper, open book techniques, classroom assignments, homework assignments, term paper).

Evaluation Criteria for Project:

Project will be evaluated on the parameters of accuracy

Abbreviations and Explanations

CC: Core Course

SDC: Skill Development Course

L: Lectures

T: Tutorial

P: Practical

Cr: Credits

Certificate Course in Computational Linguistics (Detailed Syllabus)

Semester-1

Course Code: C.L.201

Course Title: Linguistics and Languages of India

L	T	P	Credits
3	0	2	4

Total Hours: 60

Learning Outcomes:

After completion of this paper, students will be able to:

- Understand sound system of particular language
- Work on phonetics
- Work on segmental and supra-segmental phonology
Work on Morphology

Unit – 1

15 Hours

Language and Linguistics: Definitions and Area of study
Indian Languages: a brief introduction

Unit– 2

15 Hours

Introduction to Phonetics and Phonology
Phonetics: Articulatory, International Phonetic Alphabets (IPA),
Acoustics and Auditory Phonetics, Sound System
Phonology: Segmental and Supra-segmental Phonology generative phonology, Paninian
phonology

Unit– 3

15 Hours

Introduction to Morphology and Syntax: Inflectional and Derivational Morphology
Morphological information Affixation, POS, Word order and Karaka in Indian languages, Paninian
morphology and syntax

Unit – 4

15 Hours

Lexicography
History of lexicography, introduction to fundamental concepts, Computational methods,
thesauri, wordnets, corpus based lexicon development and language analysis

Mode of Transaction: Lecture; Group Discussion; Seminar, Self-learning

Tools of Transaction: Youtube, videos, audio

Suggested readings:

1. Fromkin & Rodman, An Introduction to Language (please check latest edition and authors)
2. Hockett, C.F. **A course in Modern Linguistics**. New Delhi: Oxford and IBH Publishing Co. Verma, S.K. and Krishnamurthy 1989 Modern Linguistics. Oxford University Press. 1973.
3. Chomsky, Noam and Morris Halle. 1998. **The sound pattern of English**. New York: Harper and Row.
4. Yamuna Kachru. 1980. **Aspects of Hindi Grammar**.

Course Code: C.L.202

Course Title: Introduction to Computational Linguistics

L	T	P	Credits
3	0	2	4

Total Hours: 60

Learning Outcomes:

After completion of this paper, students will be able to:

- Understand various tools of Computational Linguistics
- Work on Machine Translation
- Work on Corpora
- Work on tools of NLP

Unit – 1

15 Hours

Introduction of Computational Linguistics

Fundamental concepts, Areas of study under Computational Linguistics, theoretical and applied perspectives

Developmental areas, Language vs Speech technologies, Machine Translation- approaches and problems in Machine Translation

Unit– 2

15 Hours

Methods in Computational Linguistics – Rule based vs statistical methods

Unit – 3

15 Hours

Introduction to Corpus, Examples of language resources – LOB, Brown, ILCI etc, Corpora based language processing, Resource creation and standards in Indian languages

Unit – 4

15 Hours

Speech and Language annotation – POS, Chunk, bimodal/multimodal tagging

Mode of Transaction: Lecture; Group Discussion; Seminar, Self-learning

Tools of Transaction: Youtube, videos, audio

Suggested readings:

1. Mitkov Ruslan (ed), Oxford Handbook of Computational Linguistics
2. Chandra Subhash, 2021, *Language Computing*
3. Leech, Geoffrey, **Corpus Annotation Schemes in Literary and Linguistics Computing**,1993
4. Leech, Geoffrey and Steven Fligelstone, **Computers and Corpus Analysis in Computers and Written texts**
5. Dash. N.S 2005, **Corpus Linguistics and Language Technology**. New Delhi: Mittal Publications.
6. Hutchins W. John, **The Evaluation of Machine Translation System in Practical Experience of Machine Translation System**

Course Code: C.L.203

Course Title: Fundamentals of Programming- Python

L	T	P	Credits
3	0	2	4

Total Hours: 60

Learning Outcomes:

After completion of this paper, students will be able to:

- Make their own application tools on NLP and speech processing
- Make e-dictionaries and e-grammars
- Research on particular language

Unit- 1

15 Hours

Introduction to Programming and the Python Language: History of Programming and Python, Strength and Weakness, Different Versions, Installing Python, Setting up in local environment, IDLE, Executing from file, command line from interactive mode, Python Identifiers and reserved key words.

Python syntax: Variables and Variables type, Data types, Data Types Conversion, Operators (Arithmetic, Comparison, Assignment, Bitwise, Logical, Membership, Identity), Operators Precedence, Python Decision making (if, el if, else, nested if), Python loops (while, for, nested loops), Simple Math and Error Checking, Break and continue statements.

Unit- 2

15 Hours

Python Collections or Sequence: Sequence introduction, Number operations, String Operations, List, Tuple, Dictionary, Set.

Python Functions: Function introduction, User defined functions, Functions with parameters, Keywords and optional parameters, Scope of variables (Global and Local)

Python libraries: Numpy, Scipy, Pandas, Matplotlib, Accessing Database with Python

Unit- 3

15 Hours

Python File handling: Sending Output to STDOUT Using the print() Method, Reading Input with the input() Method, Creating File Objects with the open() Method, Controlling File Access Modes, Working with File Object Attributes, Closing File Objects with the close() Method, Reading and Writing to File Objects with read() and write(), Using File Processing Functions from the OS Module.

Unit- 4

15 Hours

Python Modules: Modules, Standard Modules (Sys, Math, Time), Import Statement, from statement, Dir functions.

Regular Expressions: Introduction, Simple Character Matches, Special Characters, Files I/O, Search function, Matching v/s Searching, Modifiers, Patterns.

Introduction to Data Structure: Implementation of Array, Stack and Queue Operations: Traversal, Insertion & Deletion at and from a given location, Implementation of sorting and searching algorithm

Mode of Transaction: Lecture; Group Discussion; Practice, Self-learning.

Suggested Readings:

1. Allen B. Downey, Think Python. O'Reilly Publishers.
2. John M. Zelle, Python Programming: An Introduction to Computer Science, Franklin, Beedle Publishers.
3. Mark Lutz, Python Pocket Reference: Python in Your Pocket, O'Reilly Publishers.
4. Sebastian Raschka & Vahid Mirjalili, Python Machine Learning, Packt Publishing
5. François Chollet, Deep Learning with Python, Manning Publications
6. Steven Bird, Ewan Klein & Edward Loper, Natural Language Processing with Python, O'Reilly

Semester-2

Course Code: C.L.221

Course Title: Corpora based language processing

L	T	P	Credits
---	---	---	---------

3	0	2	4
---	---	---	---

Total Hours: 60

Learning Outcomes:

After completion of this paper, students will be able to:

- Work on all types of dictionaries
- Work on Semantics
- Work on Word processing and DTP
- To language teach

Unit – 1 **15Hours**

Corpora creation, annotation, standards in POS, Chunks, treebanks, speech data

Unit– 2 **15 Hours**

Fundamentals of Machine Learning, Concepts from corpora based training

Unit – 3 **15 Hours**

Lexical Resource creation in India – speech and language data

Survey, evaluation of gaps

Unit – 4 **15 Hours**

Project in Lexical Resource Creation – speech or text data creation

Mode of Transaction: Lecture; Group Discussion; Seminar, Self-learning

Tools of Transaction: Youtube, videos, audio

Suggested readings:

1. John Lyons 1971, **Introduction to theoretical Linguistics**. Cambridge University Press
2. Kunjuni K. Raja 1963, **Indian theories of Meaning**.
3. Akshar Bharati, Vineet Chaitanya and Rajeev Sangal, 1995. **Natural Language Processing: A Paninian Perspective**. New Delhi
4. Grishman, Ralph. 1986. **Computational Linguistics: An Introduction**, London CUP.
5. Meijs, Willem, 1996. **Linguistics Corpora and Lexicography in Annual Review of Applied Linguistics**.

Course Code: C.L.222 (Practical)

Course Title: Natural Language Processing using Python

L	T	P	Credits
0	0	6	3

Learning Outcomes:

After completion of this paper, students will be able to:

- Understand sound system of particular language
 - Work on speech processing
 - Work on Morph Analyzers and POS tagging
- Work on Basic Syntax

Unit-1

Introduction to Python Text Basics: - NLP Introduction, Introduction to packages like NLTK, Computing with Language- Texts and Words, Texts as Lists of Words, Processing Raw Text- Accessing Text from the Web and from Desk, Strings- Text processing at the Lowest level, Regular Expression for Detecting Word Patterns, Normalizing Text, regular Expressions for Tokenizing Text, Formatting: From Lists to Strings. Stemming, Lemmatization, Stop Words.

Unit-2

Part of Speech Tagging: - Introduction to Section on (part of speech) POS, Tokenization, Part of Speech Tagging, Tagging Words- Using a Tagger, mapping Words to Properties Using Python Dictionaries, Automatic Tagging, N-Gram Tagging, Transformation Based Tagging.

Unit-3

Named Entity and Text Classification:- Introduction to Named Entity Recognition (NER), How to determine the category of a word, Visualizing Part of Speech, Named Entity Recognition, Sentence Segmentation, Introduction to Text Classification, Machine Learning Overview, Classification Metrics, Confusion Matrix, Scikit-Learn, Text Feature Extraction, Text Classification models.

Unit-4

Text Analysis and Sentence Structure: - Introduction to Semantics and Sentiment Analysis, Overview of Semantics and Word Vectors, Semantics and Word Vectors, Sentiment Analysis with NLTK. Information extraction from Text, Analysing Sentence Structure using NLTK.

Text Books:

1. Jurafsky, D. & J. Martin, "Speech and Language Processing: An Introduction to Natural Language Processing Computational Linguistics, and Speech Recognition" Prentice Hall.
2. Steven Bird, Ewan Klein, Edward Loper. Natural Language Processing with Python- Analyzing Text with the Natural Language Toolkit, Shroff Publications and Distributors -O'Reilly Media.

Reference Books:

1. Grosz, B.J., Sparck Jones, K. & Webber, B.L. (eds) "Readings in natural language processing", Los Altos, CA. Morgan Kaufmann.
2. Allen, J., "Natural Language Understanding", Redwood City, Benjamin/Cummings.
3. Bharti, Akshar, Chaitanya Vineet, Sangal Rajeev, "Natural Language Processing", Prentice Hall. Palash Goyal, Sumit Pandey, Karan Jain. Deep Learning for Natural Language Processing. Apress.

Mode of Transaction: Lecture; Group Discussion; Seminar, Self-learning

Tools of Transaction: Youtube, videos, audio

Course Code: C.L.225

Course Title: Project Work

L	T	P	Credits
0	0	12	6

Hours: 180

Learning Outcomes:

After completion of this course students will be able to:

Students will complete one project on this course

Mode of Transaction: Self-learning, Project Method, Mentoring

Evaluation Criteria: Evaluation of Project work will be of 100 marks, in which 50% marks will be given by supervisor on basis of his/her daily progress and mid-term evaluation test and 50% will be on the basis of final project work and the evaluation of this work will be done on the average of marks given by supervisor, HoD and senior most faculty member of the department.