SYLLABUS

PG CERTIFICATE COURSE IN COMPUTATIONAL HUMANITIES

Centre for Translation and Digital Humanities
Ravenshaw University
Cuttack-3, Odisha

About the course

Humanities scholars are normally trained in close reading of texts. But close reading, as a research practice, is solitary in nature and is based on subjective analysis. This skill-oriented course has been planned in the context of the recent widespread phenomenon of large digital libraries constituted by varied kinds of texts in several languages, and is specially designed to train students to make the best use of their core skills in close reading to address linguistic and literary problems based on objective analysis with the aid of computational tools. The PG Certificate course on Computational Humanities includes introductory modules on the core activities involved in three major disciplines namely language, literature, and computation. For practical assignments in each unit of this course students will be taught to use digital texts (in English / Hindi / Odia / Hindi / Sanskrit language) for data mining. The assignments have been included in theorder of the complexities involved as the course advances.

Objectives of the course

(i)   To train students in computer-based methods to visualise and analyseinformation in the textual universe

(ii)   (ii) To train students in creating digital objects / texts usingcomputation; such digital artifacts are to be used for further research

Learning outcomes

By the end of the course, students are expected to develop skills in naturallanguage processing, translation and language technology, skills that companies actually hire for.

CourseStructure for PG Certificate Course in Computational Humanities

Semester	Course Code	Course name	Credits	Total Marks
I	CCH-1.1.1 CCH-1.1.1 Practical	Introduction to Natural Language Introduction to Natural Language (Lab)	4 2	100
	CCH-1.1.2 CCH-1.1.2 Practical	Natural Language Computing – I Natural Language Computing – I (Lab)	4 2	100
	CCH-1.1.3 CCH-1.1.3 Practical	Digital Library Digital Library (Lab)	2 2	80
II	CCH-1.2.4 CCH-1.2.4 Practical	Corpus Linguistics Corpus Linguistics (Lab)	4 2	100
	CCH-1.2.5 CCH-1.2.5 Practical	Natural Language Computing – II Natural Language Computing – II (Lab)	4 2	100
	CCH-1.2.6 CCH-1.2.6 Practical	Natural Language Processing (Diss.) Natural Language Processing (Lab)	2 2	80
		TOTAL	32	560

PG CERTIFICATE COURSE IN COMPUTATIONAL HUMANITIES
All papers are compulsory

- **Mark distribution for papers CCH-1.1.1, CCH-1.1.2, CCH-1.2.4, CCH-1.2.5 -**
  Midterm: 20 marks, End term: 30 marks, Practical - 50 marks; Total – 100 marks

Credit per paper: 6; Teaching hours per paper – 40 hours Theory classes + 20 hours Practical classes

- **Mark distribution for papers CCH-1.1.3, CCH-1.2.6 -** Midterm: 20 marks, End term:
  30 marks, Practical - 30 marks; Total – 80 marks

Credit per paper:4; Teaching hours per paper – 30 hours Theory classes + 10 hours Practical classes

Pre-course reading list: These are exploratory readings recommended to prepare oneselffor the full-time course.
1. Nyhan, Julianne & Flinn, Andrew, *Computation and the Humanities: Towards an Oral History of Digital Humanities*, Springer, 2016.
2. Grishman,Ralph, *Computational Linguistics: An Introduction*, Cambridge University Press, 1986.

SEMESTER I

CCH-1.1.1

Natural Language

Learning objectives: Develop an understanding of the nature of linguistic inquiry

Learning outcomes: Use logical methods to analyse the structure of natural language

Unit 1: Study of Language
- Definitions of language; Key properties of human language
- Introduction to Language families and the Indo-European language family
- ModernLinguistics: Revival of Panini – Charles Wilkins, William Jones, Franz Bopp, Saussure,Chomsky
- Paninian grammar with special reference to the life history of Maharshi Panini, Maharshi Katyayana, Maharshi Patanjali, Jayaditya and Bamana, Shrimad Bhattoji Dixit, Bhartruhari, Nagesa Bhatta
- Types of Linguistics: Theoretical linguistics, Descriptive linguistics, Historicallinguistics, Introduction to Semiotics

Unit 2: Structural subfields of Linguistics I
- Phonetics – IPA chart and description and classification of phonemes (consonant and vowel sounds)

- Phonology – minimal pairs, phones and allophones, contrastive distribution, complementary distribution, free variation
- Phonological rules: (i) Assimilation: regressive assimilation, gemination (ii) Dissimilation (iii) Epenthesis: prothesis, anaptyxis, paragoge, excrescence (iv) Deletion: aphaeresis, syncope, apocope (v) Metathesis (vi) Lenition (vii) Palatalization
- Writing rules, Syllable structure
- Morphology – free and bound morphemes, inflectional and derivational morphemes, drawing morphology trees; Morphophonology: morphophonemic alternation; Morphosyntax:Parts of speech (POS), morphosyntactic analysis
- Terminologies of Paninian grammar – sutra, vartika, bhashya, pada, vakya, pratyahara, vacyaparivartana, varna samkhya; An illustration: structural forms of noun-word "Ram" in 7 cases, verbal forms of root "bhu" in 'latlakar' (present tense) with respective sutras

Unit 3: Structural subfields of Linguistics II
- Syntax – Syntactic parsing: Immediate Constituent analysis, X-bar, Phrase structure tree, Dependency tree
- Semantics – Semantic roles: Deep roles, Thematic roles; Ambiguity: syntactic lexical ambiguity, semantic lexical ambiguity (word sense ambiguity), structural ambiguity
- Pragmatics – Beyond sentences: Discourse analysis, Implicature

Practical (Labwork):Treebank (Odia/English/Hindi/Sanskrit)

Recommended Text Books:
1. vyākaraṇa-praveśaḥ: An introduction to traditional Sanskrit grammar
2. Sanskrit for Beginners: A simple and comprehensive guide
3. Tripathy, Kunja Bihari, *Odia Bhashatattva 'o' LipiraBikasha*, The Orissa State Bureau of Textbook Preparation and Production, 1977.
4. Mahapatra, Narayan & Das, Sreedhar, *SarbasaraByakarana*, NSS, 1952.
4. Rowe, Bruce M. & Levine, Diane P. *A Concise Introduction to Linguistics*, Routledge, 5th ed. 2018.
5. Williams, Raymond, *Culture and Society: 1780-1950*, Anchor Books, 1960. (Keyword annotation)

Reference:
1. Dhall, G. B. *Aspiration in Oriya*, Utkal University, 1966.
2. Mohapatra, Dhaneswar. *Aama Odia Byakarana*, vols. 1 & 2. Aama Odisha, 2014.
3. Allen, W. S. *Phonetics in Ancient India*, London Oriental Series, vol. 1, 1953.
4. Mohanty, Bansidhar, *Odia bhasharautpatti 'o' kramabikasha*, Friends' Publishers, 1970.
5. Mishra, Binayak, *Odia bhasharaitihasa*, Utkal Sahitya Press, 1927.
6. Das, Chittaranjan, *KSHA ru A': Eka byaktigataabhidhana*, Shikshasandhan, 2005. (Keyword annotation)

CCH-1.1.2

Natural Language Computing-I

Learning objectives: Develop an understanding of how the machine makes sense of natural language

Learning outcomes: Develop an understanding of the challenges and opportunities in machine learning

Unit 1: Basics of coding

Introducing programming language; Introducing source and object code management, Basic knowledge of Data Structure (Array, Linked List, Stack, Queue, Searching and Sorting algorithms)

Unit 2: Object-oriented programming language

Introducing PYTHON

Unit 3: Introducing NLP tools and techniques

Unicode, Review of NLTKs, POS tagging, Lemmatiser-Types of lemmatiser (Wordnet, spaCy, TextBlob), Stemmer-Types of stemmer in NLTK (Porter, Snowball, Lancaster), Morphological analysis, Syntactic analysis, Semantic analysis

Practical (Labwork): Programs based on PYTHON including a Mathematical) Program, and Practical based on Unit-3 Developing morphological analyzer, POS tagger, Stemmer, Lemmatizer

Recommended Text Books:
1. Rao, Nagesh B. *Learning Python*, Cyberplus Infotech, 2017.
2. Siddiqui, Tanveer &Tiwary, U. S., *Natural Language Processing and Information Retrieval*, Oxford Higher Education, 2008.

Reference:
1. Bhattacharyya, Pushpak, *Machine Translation*, CRC Press, 2015.

CCH-1.1.3

Digital library

Learning objectives: To introduce basic concepts and characteristics of digital libraries

Learning outcomes: Develop awareness about the terminologies and technologies related to digital libraries

Unit 1: Introduction to Digital library

Evolution of Digital library, components of digital library, architecture of digital library

Unit 2:Digitisation-I

Definition, Needs of Digitization, Selection of Material for Digitization, Steps in the Process of Digitization:Scanning, Indexing, Store, Retrieve;Digitization-Input and Output Options: Scanned as Image Only, Optical Character Recognition (OCR); Technology of Digitization: Bit Depth or Dynamic Range, Resolution, Threshold, Image Enhancement; Compression: Lossless Compression, Lossy Compression, Compression Protocols, TIFF-G4, JPEG (Joint Photographic Expert Group); File Formats and Media Types

Unit 3:Digitisation-II

Equipment for digitization and implementation: Scanners, scanning software, Image editing applications; Organising Digital images: organize, name, describe

Practical (Labwork):Digitising manuscripts, rare books/records and preparing a special collection along with a digital catalogue of the digitized materials

Recommended Text Books:
E-PG Pathshala: CDL-01, 02, 03

SEMESTER II

CCH-1.2.4

Corpus Linguistics

Learning objectives: Understand the use of language in context

Learning outcomes: Understand how and why corpus is useful for research in language use

Unit 1:The Corpus approach in Linguistics
Techniques and tools, Types of Corpus (First generation Corpus, specialised Corpora, Corpus for advanced research), Features of Corpus

Unit 2: Corpus processing
Frequency count, word sorting, concordance, lexical collocation, lexical word grouping, morphological processing, part of speech tagging, word sense tagging, lemmatization, annotation, parsing

Unit 3: Utility of Corpus
Use of Corpus, Users of Corpus, Limitations of Corpus

Practical (Labwork):Corpus annotation

Recommended Text Books:
1. Dash, Niladri S. *Corpus Linguistics: An Introduction*, Pearson Education India, 2008.
2. Weisser, Martin. *Practical Corpus Linguistics: An Introduction to Corpus-based Language Analysis*, Cambridge University Press, 2018.

Reference:
1. Kuebler, Sandra &Zinsmeister, Heike, *Corpus Linguistics and Linguistically Annotated Corpora*, Bloomsbury, 2014.
2. Bennett, Gena R., "Principles of Corpus Linguistics" in *Using Corpora in the Language Learning Classroom: Corpus Linguistics for Teachers*, Michigan ELT, 2010.

CCH-1.2.5

Natural Language Computing-II

Learning objectives: Understand the computational properties of natural language and how these properties can be used to train the computer to process linguistic information

Learning outcomes: Understand key concepts in NLP

Unit 1: Text analysis
Data Visualization, Basic Text Analysis, Machine Translation, Sentiment analysis

Unit 2:Text mining
Information retrieval, Text extraction

Unit 3:Machine learning
Regular expressions, Parsing, Language modeling, Text summarisation, text-processing approaches

Practical (Labwork): Project and practical based on all the three units

Recommended Text Books:
1. Jurafsky Daniel & Martin James H., *Speech and Language Processing*, Pearson, 2017.
2. Siddiqui, Tanveer &Tiwary, U. S., *Natural Language Processing and Information Retrieval*, Oxford Higher Education, 2008.
3. Chain, Jamie, *Machine Learning With Python*, Learn Coding Fast, 2018

Reference:
1. Bhattacharyya, Pushpak, *Machine Translation*, CRC Press, 2015.
2. Rao, Nagesh B. *Learning Python*, Cyberplus Infotech, 2017.

CCH-1.2.6

Natural Language Processing (NLP)Project

Learning objectives: Understand the context in which the well-known NLTK libraries were created

Learning outcomes: Understand the use of existing NLTK libraries

The NLP project, it is expected, will train students in the use of NLTK libraries as well as help them explore the possibilities for developing NLTK libraries for Indian languages

1) Developing a POS tagger / Morphological Analyser / Lemmatiser / Stemmer / Translator / Sentiment Analyser

OR

2) Designing innovative models related to language and technology

Recommended Text Books:
1. Jurafsky, Daniel & Martin, James H. Martin, *Speech and Language Processing*, Pearson, 2017.
2. Siddiqui, Tanveer &Tiwary, U. S., *Natural Language Processing and Information Retrieval*, Oxford Higher Education, 2008.
3. Brown, Martin C., *Complete Reference Python*, Mc Graw Hill Education, 2018
4. Chain, Jamie, *Machine Learning With Python*, Learn Coding Fast, 2018

Reference:
1. Bhattacharyya, Pushpak, *Machine Translation*, CRC Press, 2015.
2. Rao, Nagesh B. *Learning Python*, Cyberplus Infotech, 2017.