PG. Diploma in Natural Language Processing (Part Time)
Programme Code: LLIN14

Curriculum
(For students admitted from the academic year 2019-2020)

Programme Outcomes

<table>
<thead>
<tr>
<th>PO1:</th>
<th>Apply the knowledge of language fundamentals and various literatures in society, computers, psychology, cognitive science and medicine.</th>
</tr>
</thead>
<tbody>
<tr>
<td>PO2:</td>
<td>Formulate, solve and analyze complex problems in variety of domains that constitute the core of language and literature knowledge, including familiarity with diverse questions of interest in the areas of (and interfaces between) structures of language and aesthetics of literature.</td>
</tr>
<tr>
<td>PO3:</td>
<td>Apply the acquired knowledge for analyzing language and writing in appropriate genres and modes for a variety of purposes and audiences and provide solutions to societal and environmental contexts for problems related to language change, policy and planning.</td>
</tr>
<tr>
<td>PO4:</td>
<td>Design and conduct research, analyse and interpret data to provide valid conclusions in the field of literature and in the descriptive as well as applied language studies.</td>
</tr>
<tr>
<td>PO5:</td>
<td>Select and apply appropriate modern theories and techniques including cognitive, psychological, biological, cultural, and social factors for language study and research.</td>
</tr>
<tr>
<td>PO6:</td>
<td>Gain exposure to attain knowledge and understand interdisciplinary and multidisciplinary linguistic and literary approaches.</td>
</tr>
<tr>
<td>PO7:</td>
<td>Acquire professional and intellectual integrity, code of conduct and ethics on communicational practices, understanding responsibilities and norms for sustainable development of society.</td>
</tr>
<tr>
<td>PO8:</td>
<td>Interact with the specific linguistic community and with society at large, through critical conversations and prepare, organize, and deliver their work to the public through speaking and writing.</td>
</tr>
<tr>
<td>PO9:</td>
<td>Understand the aesthetic and scientific concepts of language and demonstrate the knowledge as a skilled person in teams and multidisciplinary tasks in their profession.</td>
</tr>
<tr>
<td>PO10:</td>
<td>Appreciate the need for self-preparation and life-long learning independently in the broadest context of language challenges in the context of multilingualism and globalization.</td>
</tr>
</tbody>
</table>

Programme Specific Outcomes

<table>
<thead>
<tr>
<th>PSO1:</th>
<th>By Studying this course, the students will understand how to incorporate technology in Natural Language Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSO2:</td>
<td>To make the students understand language teaching and learning through technology</td>
</tr>
<tr>
<td>PSO3:</td>
<td>To understand how to compile electronic dictionary</td>
</tr>
<tr>
<td>PSO4:</td>
<td>To understand the concept of artificial intelligence.</td>
</tr>
<tr>
<td>PSO5:</td>
<td>To import the knowledge for instructing computer through programming languages.</td>
</tr>
</tbody>
</table>
PAPER – 1 COMPUTER APPLICATIONS TO LANGUAGE STUDIES

Learning Objectives of the course Computer Applications to Language Studies

❖ To get exposure on fundamentals of computers and the basics of linguistics
❖ To know the applications of computers in linguistics and language studies
❖ To understand various generations of computer,
❖ To have sound knowledge on types and anatomy of computer
❖ To equip programming languages and applications of computer

UNIT – 1 Introduction to Computers:

Computer generations; Types of computers- main frame, mini and personal computers; Hardware- computer anatomy; input and output devices, central processing UNIT -- memories; Rom, Ram; software; programming Languages; Databases.

UNIT -2 Introduction to linguistics:

Definition of Language and Linguistics; Levels in Language analysis- Phonetics, Phonology, Morphology, Syntax, Semantics and Discourse; Theoretical Vs applied.

UNIT -3 Word processing and DTP (Desk Top Publishing):

Text processing; Translation; Lexicography; Language Learning and Teaching; Speech signal processing; Tools for Linguistic analysis- Sorting, Indexing, Frequency counting, KWIC (Key Word In Context).

UNIT – 4 Tools and Techniques for Computer Assisted Text Processing:

Preparing Text for computer analysis; Reusing text and encoding standard; Packages for Description of texts; Databases; Text Retrieval System; Object Oriented DBMS and Hypertext.

UNIT – 5 Introduction to Computational Linguistics:

Natural Language Processing (NLP); Issues and problems in NLP; Application of Natural Language Understanding (NLU); Evaluating Natural Language Understanding Systems; Representation and Understanding; Organization of Natural Language Understanding Systems.

Text Books:

1. Allen, J. Natural Language Understanding, The Benjamin Company. 1995,
3. Rajaraman, D. et al., Computer Primer, New Delhi: Printice Hall of India Pvt. Ltd. 1986,

Supplementary Readings:


Outcome:

After completion of this course, the students will be able to:

CO:1. Describe fundamentals of computers and the basics of linguistics
CO:2. Understand various applications of computers in linguistics and language studies.
CO:3. Explore generations, types and anatomy of computer
CO:4. Explain natural language processing, understanding, and evaluating systems
CO:5. Differentiate natural languages and programming languages

Outcome Mapping

<table>
<thead>
<tr>
<th>Course Code:</th>
<th>CO1</th>
<th>CO2</th>
<th>CO3</th>
<th>CO4</th>
<th>CO5</th>
</tr>
</thead>
<tbody>
<tr>
<td>PO1</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>PO2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PO3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PO4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PO5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PO6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PO7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PO8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PO9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PO10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSO1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSO2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSO3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSO4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSO5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PAPER – 2 CORPUS LINGUISTICS

Learning Objectives

- To quantitative study of language
- To use of language data in machine-readable
- To understand generations of corpus
- To investigate types of corpus,
- To have sound knowledge on text processing techniques

UNIT -1 Goals and Methods of the Corpus-based approach:

Definition of corpus; corpus creation; corpora vs. collection; corpus-based approach – characteristics, association patterns in language use, role of quantitative analysis, comparison with other approaches; Corpus Based Investigations of Language Use; Corpora of written and spoken English; Corpora for Indian Languages;

UNIT – 2 Corpus Design:

Corpus generation; Types of Corpora - Primary, Secondary, Parallel Corpora;
Issues in Corpus design and Corpus Development- Data sampling, Size of Corpus, Method of data collection;
UNIT -3 Corpus Annotation:
Definition; Annotated vs. raw corpora; Levels of annotation – Phonetic/phonemic, grammatical, syntactic, semantic tagging; Criteria for annotation; Methodologies of Annotation – Automatic processing, Interactive processing, Machine-aided manual input; Characteristic of tagged corpora; Tagging of Indian Language corpora.

UNIT – 4 Corpus analyzing Tools:
Different text processing techniques - Frequency count, Concordance, Collocation, Key Word In Context (KWIC), morphological processing, parsing, and Lemmatization; UNIT -of analysis in corpus-based studies

UNIT – 5 Applications of Corpus Research:
Use of corpora in Natural Language Processing, Linguistics, Lexicography, Language teaching, Speech technology, Machine aided Translation

Text Books:

Supplementary Readings:
1. Allen, J. Natural Language Understanding, The Benjamin Company. 1995,
2. Ganesan, M. Scheme for Tagging Large Corpus in Indian Languages” in Language and Technology (Ed.) B.B. Rajapurohit, Mysore: CIIL. 1994,

Outcome:
After completion of this course, the students will be able to:
CO 1: Know the importance of quantitative study of language
CO 2: Explain use of language data in machine-readable form in natural language processing
CO 3: Create a corpus and compare with other approaches
CO 4: Differentiate the types of corpus, text processing techniques and corpus generations
CO 5: Describe the use of corpora in various fields like Linguistics, Lexicography, Language teaching, Speech technology, Machine aided Translation
Outcome Mapping

<table>
<thead>
<tr>
<th>COs</th>
<th>PO1</th>
<th>PO2</th>
<th>PO3</th>
<th>PO4</th>
<th>PO5</th>
<th>PO6</th>
<th>PO7</th>
<th>PO8</th>
<th>PO10</th>
<th>PSO1</th>
<th>PSO2</th>
<th>PSO3</th>
<th>PSO4</th>
<th>PSO5</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO1</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CO2</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CO3</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CO4</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CO5</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PAPER-3 NATURAL LANGUAGE PARSING

Learning Objectives of the course Natural Language Parsing

- To understand the system of language
- To imbibe the knowledge on various levels of language
- To acquire organization of language components internally and externally
- To know how to analyze language components
- To study the parsing of smallest elements in the larger components

UNIT -1 Morphological Analysis:

Introduction to Morphology; morph, morpheme and allomorph; free vs. bound morph; Morphological Processes - Inflectional, Derivational and Compositional Morphology; Automatic Morphological Analysis; Representation of Morphological information; MRD (Machine Readable Dictionary) for stems, suffixes; A Simple Morphological parsing scheme

UNIT – 2 Finite State Automata:

Morphology and Finite State Transducers – Finite State Morphological Parsing; Word structure rules- context free rules- features and categories- word structure formalism; Unification and its varieties; role of word grammar.

UNIT – 3 Grammar and Parsers:

Definition; Parsing in traditional grammar; Parsing in formal Linguistics; Parsing in Artificial Intelligence Grammar and sentence structure; Top – Down Parser; Bottom-up Chart parser; Transition Network Grammars; Top-Down Chart Parsing; Augmented Transition Networks ATN); Recursive Transition Network (RTN).

UNIT - 4 Features and Unification:

Feature Structures; Unification of Feature Structures; Features Structures in the Grammar; Agreement; Head Features; Subcategorization; Long-Distance Dependencies; Implementing Unification.

UNIT – 5 Lexicalized and Probabilistic Parsing:

Probabilistic Context-Free Grammars; Learning PCFC Probabilities; Problems with PCFC; Probabilistic Lexicalized CFGs; Dependency Grammars; Categorical Grammar; Human Parsing.

Text Books:

Supplementary Readings:
1. Allen, J. Natural Language Understanding, The Benjamin Company. 1995,
2. Ganesan, M. Scheme for Tagging Large Corpus in Indian Languages” in Language and Technology (Ed.) B.B. Rajapurohit, Mysore: CIIL. 1994,

Outcome
After completion of this course, the students will be able to:

CO1: Explain different levels of natural language parsing such as morphological, syntactic, semantic and lexical levels
CO2: Differentiate the traditional and computational grammar and its applications
CO3: Create automatic morphological analyzer and parser.
CO4: Describe morphophonemic changes and morphotactic rules and sandhi rules
CO5: Unify the structural features with grammar and implement the unification

Outcome Mapping

<table>
<thead>
<tr>
<th>CO1</th>
<th>PO1</th>
<th>PO2</th>
<th>PO3</th>
<th>PO4</th>
<th>PO5</th>
<th>PO6</th>
<th>PO7</th>
<th>PO8</th>
<th>PO9</th>
<th>PO10</th>
<th>PSO1</th>
<th>PSO2</th>
<th>PSO3</th>
<th>PSO4</th>
<th>PSO5</th>
</tr>
</thead>
<tbody>
<tr>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
</tbody>
</table>

PAPER – 4 ARTIFICIAL INTELLIGENCE

Learning Objectives
❖ To understand the unambiguous Representations
❖ To equip knowledge representations to the systems
❖ To know meaning structure of language
❖ To get exposure on semantic network and frames
❖ To acquire knowledge on lexical relations

UNIT – 1 Representing Meaning:
Computational Desiderata for Representations; Verifiability; Unambiguous Representations; Canonical Form; Inference and Variables; Expressiveness; Meaning Structure of Language; Predicate-Argument Structure; First order Predicate Calculus; Elements of FOPC; The Semantics of FOPC; Variables and Quantifiers; Inference.

UNIT – 2 Some Linguistically Relevant Concepts:
Categories; Events; Representing Time; Aspect; Representing Beliefs; Pitfalls; Related Representational Approaches; Alternative Approaches to Meaning; Meaning as Action; Meaning as Truth; Semantic Networks; Frames

UNIT – 3 Semantic Analysis:
Syntax-Driven Semantic Analysis; Semantic Augmentations to Context-Free Grammar Rules; Quantifier Scooping and the Translation of Complex-Terms; Attachments for a Fragment of English; Sentences; Noun Phrases; Verb
Phrases; Prepositional Phrases; Idioms and Compositionality; Robust Semantic Analysis; Semantic Grammars; Information Extraction.

UNIT – 4 Lexical Semantics:
Relations among Lexemes and their Senses; Homonymy; Polysemy; Synonymy; Hyponymy; WordNet: A Database of Lexical relations; The Internal Structure of Words; Thematic Roles; Selectional Restrictions; Primitive Decomposition: Semantic Fields.

UNIT – 5 Word Sense Disambiguation and Information Retrieval:
Selectional Restriction-Based Disambiguation; Limitations of Selectional Restrictions; Robust Word Sense Disambiguation; Machine Learning Approaches; Dictionary-Based Approaches; Informational Retrieval; The Vector Space Model; Term Weighting; Term Selection and Creation; Homonymy, Polysemy and Creation.

Text books:
2. Jurafsky, Daniel and James H. Martin, 2002, Speech and Language Processing, New Delhi: Pearson Education

Supplementary Readings:

Outcome
After completion of this course, the students will be able to:

CO1: Explain the concept of meanings and their representation
CO2: Differentiate various methods of semantic analysis
CO3: Distinguish the word sense relations and word sense disambiguation
CO4: Describe machine learning approaches; dictionary-based approaches
CO5: Create semantic networks and frames

Outcome Mapping

<table>
<thead>
<tr>
<th>CO6</th>
<th>PO1</th>
<th>PO2</th>
<th>PO3</th>
<th>PO4</th>
<th>PO5</th>
<th>PO6</th>
<th>PO7</th>
<th>PO8</th>
<th>PO9</th>
<th>PO10</th>
<th>PSO1</th>
<th>PSO2</th>
<th>PSO3</th>
<th>PSO4</th>
<th>PSO5</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO1</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>CO2</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>CO3</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>CO4</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>CO5</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

*******