131P15C401: Natural Language processing

Sr. No.	Modules/Units	No of Hour
1	**Classical Approaches to Natural Language Processing**	15
2	Corpus and Statistical Techniques	15
3	Statistical parsing and working with word similarity	15
4	Speech Recognition and Translation along with Applications	15
Total		60

Course Objectives

1) To understand the leading trends and systems in natural language processing.
2) To understand the concepts of morphology, syntax, semantics, and pragmatics of the language.
3) To recognize the significance of pragmatics for natural language understanding.
4) Enable students to be capable to describe the application based on natural language processing and to show the points of syntactic, semantic, and pragmatic processing.

Learning Outcomes

Learner should be able to

1) Apply approaches to syntax and semantics in NLP.
2) Understand approaches to discourse, generation, dialogue, and summarization within NLP.
3) Understand current methods for statistical approaches to machine translation.
4) Implement machine learning techniques used in NLP.

Unit I:

Classical Approaches to Natural Language Processing

Text Preprocessing: Introduction, Challenges of Text Preprocessing, Tokenization, Sentence Segmentation

Lexical Analysis: Introduction, Finite State Morphonology, Finite State Morphology, "Difficult" Morphology and Lexical Analysis, Paradigm-Based Lexical Analysis

Syntactic Parsing: Introduction, parsing as deduction, review of LR Parsing, Constraint based grammars, Issues in Parsing

Semantic analysis: Introduction, Approaches to semantic representation, relational issues

Natural Language Generation: Introduction, Examples of Generated Texts, The Components of a Generator, Approaches to Text Planning, The Linguistic Component

Unit II:

Corpus Creation: Introduction, Corpus Size, Balance, Representativeness, and Sampling Data Capture and Copyright Corpus Markup and Annotation Multilingual Corpora Multimodal Corpora

Treebank Annotation: Introduction, Corpus Annotation Types, Morphosyntactic Annotation, Treebanks: Syntactic, Semantic, and Discourse Annotation, The Process of Building Treebanks, Applications of Treebanks, Searching Treebanks

Fundamental Statistical Techniques: Binary Linear Classification, One-versus-AII Method for Multi-Category Classification, Maximum Likelihood Estimation, Generative and Discriminative Models, Mixture Model and EM, Sequence Prediction Models References

Part-of-Speech Tagging: Introduction, The General Framework, Part-of-Speech Tagging Approaches, Other Statistical and Machine Learning Approaches, POS Tagging in Languages Other Than English

Unit III:

Statistical Parsing: Basic Concepts and Terminology, Probabilistic Context-Free Grammars, Generative Models, Discriminative Models, Beyond Supervised Parsing

Multiword Expressions: Linguistic Properties of MWEs, Types of MWEs, MWE Classification

Normalized Web Distance and Word Similarity: Methods for Word Similarity, Background of the NWD Method, Introduction to Kolmogorov Complexity, Information Distance, Word Similarity: Normalized Web Distance, Applications **Word Sense Disambiguation:** Word Sense Inventories and Problem Characteristics, Applications of Word Sense Disambiguation, Early Approaches to Sense Disambiguation, Supervised Approaches to Sense, Disambiguation, Lightly Supervised Approaches to WSD, Unsupervised WSD and Sense Discovery
Unit IV:
Modern Speech Recognition: Major Architectural Components, Major Historical Developments in Speech Recognition, Speech-Recognition Applications, Technical Challenges **Alignment:** Definitions and Concepts, Sentence Alignment, Character, Word, and Phrase Alignment, Structure and Tree Alignment, Biparsing and ITG Tree Alignment **Statistical Machine Translation:** Approaches, Language Models, Parallel Corpora, Word Alignment, Phrase Library, Translation Models, Search Strategies **Applications:** Information retrieval, Question answering, Report generation, Ontology Construction, Sentiment analysis

Teaching & Learning Process-

Assessment Method-

Class participation, Presentation, Practical, Viva/ test, End Semester Exam

Textbook(s)-

1. Handbook of Natural Language Processing, Indurkhya, N., & Damerau, F. J., CRC Press Taylor and Francis Group, 2nd Edition
2. Speech and Language Processing, Martin, J. H., & Jurafsky, D., Pearson Education India, 2nd Edition, 2013
3. Foundations of Statistical Natural Language Processing, Manning, Christopher, and Heinrich, Schutze, MIT Press
4. Natural Language Processing in Action_ Understanding, analyzing, and generating text with Python, Hobson Lane, Hannes Hapke, Cole Howar, Manning Publications, 2019

References Books –

1. Natural Language Processing with Python, Steven Bird, Edward Lope, O'Reilly Media, 2nd 2016
2. Natural Language Processing, Sybgen Learning, 2020
3. Mastering Natural Language Processing with Python, Deepti Chopra, Nisheeth Joshi, Iti Mathur - Packt Publishing, 2016

PRACTICAL COMPONENT (1 Credit)

Sr. No	Module
1	Program on tokenization, substitution, and correcting tokens.
2	Program on applying similarity measures
3	Program on parts of speech tagging
4	Program on morphology and stemmer
5	Program on sentiment analysis
6	Program on information retrieval
7	Program for analyzing and parsing data
8	Program for statistical language modeling (word frequency and smoothing)

131P15C402: Project

The syllabus proposes project implementation as part of the semester–IV. The project implementation is continuation of the project proposal the students has submitted and evaluated in semester-III. The student is expected to continue with the proposal made and examined in the semester-III and implement the same in the semester–IV. In addition, experimental set up, analysis of results, comparison with results of related works, conclusion and prospects will be part of the project implementation.

A student is expected to make a project implementation report and appear for a project viva. He or she needs to spend around 200 hours for the project implementation, which fetches 5 credits. Student should submit a detailed project implementation report at the time of viva.

Student should submit a detailed project implementation report at the time of viva. A Student should submit project implementation report with details like Title, Implementation details, Experiment setup and results, Analysis of the result, Conclusion, Future enhancement, and Program code.

The project must be atleast 30 Pages (Excluding program code), which needs to be signed by the teacher in charge and head of the Department. Student should submit the signed project implementation report along with evaluated copy of the project proposal documentation (of semester –III) at the time of Project evaluation and viva as part of the University examination.

131P15C403: Internship

Internship with industry the syllabus proposes an internship for about 8 weeks to 12 weeks to be done by a student. It is expected that a student chooses an IT, IT-related, or IT Sector of any industry and formally works as a intern during the period.

The student should subject oneself with an internship evaluation with proper documentation of the attendance and the type of work he or she has done in the chosen organization by the person, to whom the student was reporting, with Organization's seal should be attached as part of the documentation.

A student is expected to make a report based on the internship he or she has done in an organization. It should contain the details like Certificate from organization, Evaluation form, Title, Description of the organization, Activity of the student as an intern, self-assessment of the student.

The internship report may be around 15 pages and this needs to be submitted to the external examiner at the time of university examination.

131P15E405: Data Science

Sr. No.	Modules/Units	No of Hour
1	Data Science, data Modeling and optimization	15
2	Data Exploration and Text Analytics	15
3	Data Analytics	15
4	Deep Learning	15
Total		60

Course Objectives

1) Building the fundamentals of data science.
2) Imparting design thinking capability to build bigdata
3) Gaining practical experience in programming tools for data sciences
4) Empowering students with tools and techniques used in data science

Learning Outcomes

1) Apply data visualization in big-data analytics
2) Utilize EDA, inference, and regression techniques.
3) Utilize Matrix decomposition techniques to perform data analysis.
4) Apply data pre-processing & data analytics techniques.

Unit I:
Data Science: Introduction, Important, Tools for data Science- R and Python. **Data Visualization:** Understanding and Visualizing Data, Data Visualization in Tableau, PowerBI. **Decision Making and Predictive Analysis:** Implementing Scientific Decision Making, Using Predictive Data Analysis, Case Study **Data Modeling and Optimization:** Modeling Uncertainty and Risk, Optimization and Modeling Simultaneous Decisions, Case Study
Unit II:
Data Exploration: Text Analytics & Processing + Text-Based Predictive Modelling Basic Visualizations using Tableau, Data Storytelling, visualization using PowerBI, visualization using Plotly, Case Studies of Visualizations
Unit III:
Data Analytics: Clustering Techniques, Anomaly Detection, Dimensionality Reduction, Association Rule Learning.
Unit IV:
Deep Learning: Tree Models, Model Selection & General ML Techniques, Principal Component Analysis, Advanced Regression, Bagging and Boosting, Time Series Analysis, Convolutional Neural Networks - Introduction and Industry Applications, Recurrent Neural Networks

Teaching & Learning Process-

Assessment Method-

Class participation, Presentation, Practical, Viva/ test, End Semester Exam

References Books –

1. Storytelling with Data: A Data Visualization Guide for Business Professionals, Wiley, 2021

PRACTICAL COMPONENT (1 Credit)

Sr. No	Module
1	Data Visualization and creating story telling using Tableau
2	Data Visualization and creating story telling using PowerBI
3	Data Visualization using plotly
4	Implementing Predictive – Time series analysis
5	Case study on Optimization and Modeling Simultaneous Decisions
6	Association rule learning using case study
7	Text based predictive modeling
8	Dimension reduction case study
9	Implementing Bagging and Boosting
10	Implementing recurrent Neural network

131P15E406: Cryptography and Cryptanalysis

Sr. No.	Modules/Units	No of Hour
1	**Number Theory for Cryptography and Cryptanalysis**	15
2	**Designing Simple Cryptosystems**	15
3	**Popular Cryptosystem**	15
4	**Managing Key Distribution and Key Agreements**	15
Total		60

Course Objectives

1) To develops **a** basic understanding of the algorithms used to protect users online and to understand some of the design choices behind these algorithms.
2) To develop a workable knowledge of the mathematics used in cryptology
3) Make student aware working of popular cryptosystem.

Learning Outcomes: On Completion of this course, students will have the knowledge and skills to:

1) Solve **problems** in elementary number theory
2) Apply elementary number theory to cryptography
3) Develop a deeper conceptual understanding of the theoretical basis of number theory and cryptography
4) Have knowledge and understanding of key distribution and key agreement in Cryptography.

Unit I:

Number Theory for Cryptography and Cryptanalysis: Introduction, estimates for doing arithmetic-divisibility and the Euclidean algorithm,
Congruence: Definitions and properties, linear congruence, residue classes, Euler's phi function, Fermat's Little Theorem, Chinese Reminder Theorem, Applications to factoring, finite fields, Quadratic residues, and reciprocity: Quadratic residues, Legendre symbol, Jacobi Symbol.

Unit II:

Designing Simple Cryptosystems: Shift Cipher, Substitution Cipher, Affine Cipher, Vigenère Cipher, Vermin Cipher, Hill Cipher, Permutation Cipher, Stream Cipher, Cryptanalysis of Affine Cipher, Substitution Cipher, Vigenère Cipher and Hill Cipher, **Designing and Using Block Ciphers:** Introduction, Algorithm Modes, DES, Double DES, Triple DES, Meet-in-Middle Attack, AES, IDEA algorithm.
Cryptographic Hash Functions: Hash Functions and Data Integrity, Security of Hash Functions, Secure Hash Algorithm, Message Authentication Code, Nested MACs, HMAC.

Unit III:

Popular Cryptosystem: The RSA Algorithm, Primarily Testing, Legendre and Jacobi Symbols, The Solvay- Strassen Algorithm, The Miller-Rabin Algorithm, Factoring Algorithm: The pollard p-1 Algorithm, Dixon's Random Squares Algorithm, Attacks on RSA, The Rabin Cryptosystem.
Public Key Cryptosystems: The idea of public key Cryptography, The Diffie-Hellman Key Agreement, El Gamal Cryptosystem, The Pollard Rho Discrete Logarithm Algorithm, Elliptic Curves, Knapsack problem.

Unit IV:

Managing Key Distribution and Key Agreements
Diffie-Hellman Key distribution and Key agreement scheme, Key Distribution Patterns, Mitchell-Piper Key distribution pattern, Station-to-station protocol, MTI Key Agreement scheme.
Public-Key Infrastructure: Understanding PKI, Secure Socket Layer, Certificates, Certificate Life cycle, Trust Models: Strict Hierarchy Model, Networked PKIs, The web browser Model, Pretty Good Privacy.

Teaching & Learning Process-

Assessment Method-

Class participation, Presentation, Practical, Viva/ test, End Semester Exam

Textbook(s):

1. Discrete Mathematics and Its Applications, Kenneth H. Rosen, 7th Edition,
2. McGraw Hill, 2012. Cryptography Theory and Practice, 3rd Edition, Douglas R. Stinson, 2005.

References Books –

1. Cryptography and Network Security: Principles and Practices, William Stalling, Fourth Edition, Prentice Hall, 2013.
2. Introduction to Cryptography with coding theory, second edition, Wade Trappe, Lawrence C. Washington, Pearson, 2005
3. An Introduction to Mathematical Cryptography, Jeffrey Hoffstein , Jill Pipher, Springer,2014
4. Network Security and Cryptography, Atul Kahate, McGraw Hill, 2003.

PRACTICAL COMPONENT (1 Credit)

Sr. No	Module
1	Program to implement (1) Chinese Reminder Theorem (ii) Fermat's Little Theorem
2	Program to implement Cipher methods (i) Affine Cipher (ii) Rail Fence Technique (iii) Simple Columnar Technique (iv) Vermin Cipher (v) Hill Cipher to perform encryption and decryption.
3	Program to implement the RSA Algorithm to perform encryption and decryption.
4	Program to implement the encryption and decryption using (i) Miller-Rabin Algorithm (ii) pollard p-1 Algorithm.
5	Program to implement the El Gamal Cryptosystem.
6	Program to generate symmetric keys using the Diffie-Hellman Key Agreement algorithm.
7	Program to implement hashing techniques using MD5 algorithm to compute the message digest.
8	Program to implement DES algorithm (initial permutation, key generation, and substitution).
9	Program for encrypting and decrypting text using IDEA algorithm.
10	Program to implement HMAC signatures scheme.

Evaluation Pattern
(Project Course)

Internal

Maximum Marks: 50

Project Implementation	Project Conduct	Quality and Relevance	10
		Documentation	10
		Presentation	10
		Project Viva	20
		Total Marks	50

External

Maximum Marks: 100

Project Implementation	Project Conduct	Quality and Relevance	30
		Documentation	20
		Presentation	20
		Project Viva	30
		Total Marks	100

Evaluation Pattern
(Internship Course)
External

Internship	Internship Conduct	Quality and Relevance	40
		Documentation	30
		Presentation	30
		Internship Viva	50
		Total Marks	150