MANIPAL UNIVERSITY JAIPUR

School of Computer Science and Engineering

Department of Computer Science & Engineering
Course Hand-out

Natural Language Processing | **CSE3246** | 3 Credits | 3 0 0 3

Session: JAN-MAY 2026 |Dr. Varda Pareek| Mr Virendra Meghwal| Dr. Shikha Mundra |Dr Vivek Sikarwar |Dr Surbhi Sharma Class: VI CSE

Course Coordinator: Dr Surbhi Sharma

A. **Vison**:   To achieve Excellence in Computer Science & Engineering Education for Global Competency with Human Values
B. **Mission**:   M1. Provide innovative Academic & Research Environment to develop competitive Engineers in the field of Computer Science and Engineering
   M2. Develop Problem-solving & Project Management Skills by Student Centric Activities & Industry Collaboration
   M3. Nurture the Students with Social & Ethical Values

C. **PEO1:** Graduates will Demonstrate Professional Skills on Global Platform in Computer Science and Engineering and Integrated Domains
   PEO2: Graduates will Enhance Knowledge and Skills through Higher Studies and Lifelong Learning of New Computing Technologies
   PEO3 Graduates will Provide Innovative Solutions to Drive Societal Advancement through Technology Entrepreneurship and Start-up

D. **Introduction:** This course is offered by Dept. of computer science and engineering as an elective subject, targeting students who wish to pursue development in industries or higher studies in field of Natural Language Processing. This course aims to make the students understand the mentioned points.
   (i)   Apply models, methods, and algorithms of Natural Language Processing for common NLP tasks, such as speech recognition, machine translation, spam filtering, text classification, spell checking etc.
   (ii) Implement probabilistic models, estimate parameters for such models, and run meaningful experiments to validate such language models.
   (iii) Understand linguistic phenomena and explore the linguistic features relevant to each NLP task.
   (iv) Design, implement and analyse algorithms for real-time NLP problems.

E. Course Outcomes: After completion of this course student will be able to:

CSE3246.1 Explain the fundamentals of human language processing, including semantics, pragmatics, challenges, and stages of NLP.
(level 2: Understand)

CSE3246.2 Apply linguistic components such as parts of speech, phrase structure, and morphological processes in natural language analysis. *(level 3: Apply)*

CSE3246.3 Apply statistical foundations, corpus-based methods, and text preprocessing techniques to process and model natural language data.
(level 3: Apply)

CSE3246.4 Analyze language modelling and disambiguation techniques, including n-gram models, collocations, and word sense disambiguation methods.
(level 4: Analyze)

CSE3246.5 Examine sequence labelling, parsing, and advanced NLP techniques, including POS tagging, syntactic parsing, and deep learning models, for real-world applications.
(level 4: Analyze)

F. PROGRAM OUTCOMES AND PROGRAM SPECIFIC OUTCOMES

[PO.1]   **Engineering knowledge**: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems

[PO.2]   **Problem analysis**: Identify, formulate, research literature, and analyze complex engineering problems reaching substantiated conclusions. using first principles of mathematics, natural sciences, and engineering sciences

[PO.3]   **Design/development of solutions**: Design solutions for complex engineering problems and design system components or processes. that meet the specified needs with appropriate consideration for public health and safety, and the cultural, societal, and environmental. considerations

[PO.4]   **Conduct investigations of complex problems**: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

[PO.5]   **Modern tool usage**: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools included. prediction and modeling to complex engineering activities with an understanding of the limitations

[PO.6]   **Engineer and society**: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal, and cultural issues and the consequent responsibilities relevant to professional engineering practice.

[PO.7]   **Environment and sustainability**: Understand the impact of professional engineering solutions on society and environmental. contexts, and demonstrate the knowledge of, and need for sustainable development.

[PO.8]   **Ethics**: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practices

[PO.9]   **Individual and team work**: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings

[PO.10]   **Communication**: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

[PO.11]   **Project management and finance:** Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments

[PO.12]   **Life-long learning**: Recognize the need for and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

[PSO.1]   Will be able to design, develop and implement efficient software for a given real-life problem.

[PSO.2]   Will be able to apply knowledge of AI, Machine Learning and Data Mining in analyzing big data for extracting useful information from it.

and for performing predictive analysis.

[PSO.3]   Will be able to design, manage and secure wired/ wireless computer networks for transfer and sharing of information.

G. Syllabus:

Introduction: Fundamentals of human language processing, semantics and pragmatics in NLP, challenges in NLP (ambiguity and context), stages of NLP, parts of speech (POS), phrase structure, statistical foundations, entropy, perplexity, cross-entropy, character encoding, corpora and morphological analysis (inflectional and derivational), text preprocessing (tokenization, sentence segmentation, etc.); **Corpus Linguistics and Language Modeling**: Corpus design and annotation, statistical approaches to language modeling, n-gram models, collocations, word sense disambiguation (WSD), supervised and unsupervised disambiguation, dictionary and thesaurus-based methods, second-language corpus-based WSD; **POS Tagging and Sequence Labeling**: Markov Models and Hidden Markov Models (HMM), parameter estimation, Viterbi algorithm, multiple observation HMMs, information sources in tagging, applications of POS tagging; **Parsing and Syntax Processing**: Probabilistic Context-Free Grammars (PCFGs), parsing algorithms, inside-outside algorithm, disambiguation using parsing, Treebanks, dependency parsing, phrase structure grammar, lexicalized and derivational parsing models; **Advanced NLP and Deep Learning**: Syntax and semantics in real-world applications, machine translation (word/sentence alignment), information extraction, text mining, question answering systems, sentiment analysis, social network analysis, introduction to deep learning models in NLP – CNN, RNN, LSTM, encoders, and decoders.

H. Textbooks: 1. Steven Bird, Ewan Klein, and Edward Loper, Natural Language Processing with Python, O'Reilly Media, 2009.

2.Jacob Eisenstein, Introduction to Natural Language Processing, MIT Press, 2019.
3.Richard Johansson and Pierre Nugues, Deep Learning in Natural Language Processing, Cambridge University Press, 2020.

I. **References:**

   1. **Christopher D. Manning and Hinrich Schütze,** *Foundations of Statistical Natural Language Processing***, MIT Press, 2003.**
   2. **D. Jurafsky and J. H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, (3e), Upper Saddle River, NJ: Pearson Prentice Hall, 2025.**

J. **Lecture Plan:**

Lec #	Topics	Session Outcome	Mode of Delivery	Corresponding CO	Mode of Assessing the Outcome
1	Fundamentals of Human Language Processing	Understand scope and objectives of NLP	PPT, Discussion	CSE3246.1	Quiz 1 and Sessional-I and End-Sem
2	Human Language Characteristics	Explain ambiguity, variability, and structure of language	PPT, Discussion	CSE3246.1	Quiz 1 and Sessional-I and End-Sem
3	Semantics in NLP	Understand meaning representation	PPT, Examples Discussion	CSE3246.1	Quiz 1 and Sessional-I and End-Sem
4	Pragmatics in NLP	Analyze role of context in interpretation	PPT, Examples Discussion	CSE3246.1	Quiz 1 and Sessional-I and End-Sem
5	Challenges in NLP (Ambiguity & Context)	Identify key NLP challenges	PPT, Marker & Board,	CSE3246.1	Quiz 1 and Sessional-I and End-Sem
6	Stages of NLP	Describe stages of NLP pipeline	Marker & Board, PPT	CSE3246.1	Quiz 1 and Sessional-I and End-Sem

Lec #	Topics	Session Outcome	Mode of Delivery	Corresponding CO	Mode of Assessing the Outcome
7	Parts of Speech (POS)	Identify and classify POS	PPT, Examples, Marker & Board,	CSE3246.2	Quiz 1 and Sessional-I and End-Sem
8	Phrase Structure Grammar	Explain phrase structure rules	Marker & Board, PPT	CSE3246.2	Quiz 1 and Sessional-I and End-Sem
9	Statistical Foundation	Understand language modeling probability concepts.	PPT, Examples	CSE3246.2	Quiz 1 and Sessional-I and End-Sem
10	Entropy and Perplexity	Compute entropies for language models	Marker & Board, PPT	CSE3246.2	Quiz 1 and Sessional-I and End-Sem
11	Cross Entropy, Character Encoding	Evaluate language models and ASCII and unicodes.	Marker & Board, PPT	CSE3246.2	Quiz 1 and Sessional-I and End-Sem
12	Corpora and Corpus Linguistics	Understand types and uses of corpora	PPT and Discussions	CSE3246.2	Quiz 1 and Sessional-I and End-Sem
13	Morphological Analysis	Apply inflectional and derivational analysis	PPT and Discussions	CSE3246.2	Quiz 1 and Sessional-I and End-Sem
14	Text Preprocessing	Apply tokenization and normalization	PPT and Discussions	CSE3246.3	Quiz 1 and Sessional-I and End-Sem
15	Tokenization	Apply tokenization models.	Marker & Board, PPT	CSE3246.3	Quiz 1 and Sessional-I and End-Sem

Lec #	Topics	Session Outcome	Mode of Delivery	Corresponding CO	Mode of Assessing the Outcome
16	Sentence Segmentation	Perform sentence boundary detection	Marker & Board, PPT	CSE3246.3	Quiz 1 and Sessional-I and End-Sem
17	Corpus Design and Annotation	Explain principles of corpus design	Marker & Board, PPT	CSE3246.3	Quiz 1 and Sessional-I and End-Sem
18	Statistical Language Modeling	Explain principles of corpus design	Marker & Board, PPT	CSE3246.3	Quiz 1 and Sessional-I and End-Sem
19	N-gram Models	Build n-gram models	Marker & Board, PPT	CSE3246.4	Quiz 1 and Sessional-I and End-Sem
20	Collocations	Identify and analyze collocations	Marker & Board, PPT	CSE3246.4	Quiz 2 and End-Sem
21	Introduction to WSD	Understand word sense ambiguity	PPT and Discussions	CSE3246.4	Quiz 2 and End-Sem
22	Supervised and Unsupervised WSD	Apply supervised and unsupervised disambiguation	PPT and Discussions	CSE3246.4	Quiz 2 and End-Sem
23	Dictionary-based WSD, Thesaurus-based	Explain thesaurus-based and knowledge-based WSD	Marker & Board, PPT	CSE3246.4	Quiz 2 and End-Sem
24	Second language based WSD and Markov Models	Explain Markov assumptions	Marker & Board, PPT	CSE3246.4	Quiz 2 and End-Sem
25	Hidden Markov Models (HMM)	Understand HMM structure	Marker & Board, PPT	CSE3246.4	Quiz 2 and End-Sem
26	HMM Parameter Estimation	Estimate transition & emission probabilities	Marker & Board, PPT	CSE3246.4	Quiz 2 and End-Sem

Lec #	Topics	Session Outcome	Mode of Delivery	Corresponding CO	Mode of Assessing the Outcome
27	Viterbi Algorithm	Apply Viterbi decoding	Slides / Black Board	CSE3246.4	Quiz 2 and End-Sem
28	Multiple Observation HMMs	Understand extended HMM models	PPT and Discussions	CSE3246.4	Quiz 2 and End-Sem
29	Information sources in tagging, POS Tagging Applications	Analyze applications of POS tagging	PPT and Discussions	CSE3246.4	Quiz 2 and End-Sem
30	Probabilistic Context-Free Grammars (PCFGs), parsing algorithms & Inside–Outside Algorithm	Explain probabilistic parsing	Marker & Board, PPT	CSE3246.5	Quiz 2 and End-Sem
31	Disambiguation using parsing, phrase structure grammar,	Compare parsing techniques	Marker & Board, PPT	CSE3246.5	Quiz 2 and End-Sem
32	Dependency Parsing & Treebanks, lexicalized and derivational parsing models	Analyze dependency structures	Marker & Board, PPT	CSE3246.5	Quiz 2 and End-Sem
33	Syntax and semantics in real-world applications	Analyze Syntax and Semantics in real-world applications.	Marker & Board, PPT	CSE3246.5	End-Sem
34	Machine translation (word/sentence alignment)	Understand Machine Translation.	Marker & Board, PPT	CSE3246.5	End-Sem
35	Information extraction, text mining, question answering systems	Understand concept of text mining.	Marker & Board, PPT	CSE3246.5	End-Sem
36	Sentiment analysis, social network analysis,	Understand concept of NLP.	PPT and Discussions	CSE3246.5	End-Sem
37	Introduction to deep learning models in NLP – CNN, RNN.	Understand CNN and RNN Models.	PPT and Discussions	CSE3246.5	End-Sem

Lec #	Topics	Session Outcome	Mode of Delivery	Corresponding CO	Mode of Assessing the Outcome
38	Introduction to deep learning models in NLP –LSTM, encoders, and decoders.	Understand LSTM, Encoder–Decoder models	PPT and Discussions	CSE3246.5	End-Sem
39	Discussion	Discussion	NA	NA	NA

K. Course Evaluation: As per DOA guidelines

Criteria	Description	Date	Maximum Marks
Internal Assessment (Summative)	Sessional Exam	As per Academic Calendar	30
	1. Assignments by Students: 20 Marks (Submission + Viva) 2. Two Quizzes: 10 Marks (Best one out of two quizzes will be taken for Evaluation)	Monthly	30
End Term Exam (Summative)	End Term Exam	As per Academic Calendar	40
	Total		100
Attendance (Formative)	A minimum of 75% Attendance is required to be maintained by a student to be qualified for taking up the End Semester examination. The allowance of 25% includes all types of leaves including medical leaves.		
Make up Assignments (Formative)	Students who miss a class will have to report to the teacher about the absence. A makeup assignment on the topic taught on the day of absence will be given which has to be submitted within a week from the date of absence. No extensions will be given on this.		
Homework/ Home Assignment/ Activity Assignment	There are situations where a student may have to work in home, especially before a flipped classroom. Although these works are not graded with marks. However, a student is expected		

(Formative)	to participate and perform these assignments with full zeal since the activity/ flipped classroom participation by a student will be assessed and marks will be awarded.

L. **Make-up Policy:** As per University Norms

M. **Chamber Consultation:** online/offline as per Instructor

N. **Notice:** Via email/ Teams/WhatsApp (Use University Microsoft Account)

Course Articulation Matrix: (Mapping of COs with POs)

Course Out-Comes	STATEMENT	CORRELATION WITH PROGRAM OUTCOMES												CORRELATION WITH PROGRAM SPECIFIC OUTCOMES		
		PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CSE3246. 1	Explain the fundamentals of human language processing, including semantics, pragmatics, challenges, and stages of NLP.	2	3	2									2			
CSE3246. 2	Apply linguistic components such as parts of speech, phrase structure, and morphological processes in natural language analysis.	3	2	2									2		2	
CSE3246. 3	Apply statistical foundations, corpus-based methods, and text preprocessing techniques to process and model natural language data.	3	2	3	2				2	2		2	3		3	
CSE3246. 4	*Analyze language modelling and disambiguation techniques, including n-gram models, collocations, and word sense disambiguation methods.*	3	3	2	2				3	2		3	2		2	
CSE3246. 5	Examine sequence labelling, parsing, and advanced NLP techniques, including POS tagging, syntactic parsing, and deep learning models, for real-world applications.	2	3	3	2				3	3		3	3		3	

1- **Low Correlation; 2- Moderate Correlation; 3- Substantial Correlation**