

<b>M. Sc (Information Technology)</b>		<b>Semester – III</b>	
<b>Course Name: Natural Language Processing</b>		<b>Course Code: PSIT201</b>	
<b>Periods per week</b> <b>1 Period is 60 minutes</b>	<b>Lectures</b>	<b>4</b>	
	<b>Credits</b>	<b>4</b>	
		<b>Hours</b>	<b>Marks</b>
<b>Evaluation System</b>	<b>Theory Examination</b>	<b>2½</b>	<b>60</b>
	<b>Theory Internal</b>	<b>--</b>	<b>40</b>

<b>Objectives</b>	<ul style="list-style-type: none"> <li>• The prime objective of this course is to introduce the students to the field of Language Computing and its applications ranging from classical era to modern context.</li> <li>• Course also aims to provide understanding of various NLP tasks and NLP abstractions such as Morphological analysis, POS tagging, concept of syntactic parsing, semantic analysis etc.</li> <li>• Course provide knowledge of different approaches/algorithms for carrying out NLP tasks.</li> <li>• Course also highlights on the concepts of Language grammar and grammar representation in Computational Linguistics.</li> </ul>
-------------------	--

<b>Pre requisites</b>	Knowledge of Algorithms and mathematical foundation
-----------------------	---

<b>Unit</b>	<b>Details</b>	<b>References</b>
<b>I</b>	Introduction to NLP, brief history, NLP applications: Speech to Text(STT), Text to Speech(TTS), Story Understanding, NL Generation, QA system, Machine Translation, Text Summarization, Text classification, Sentiment Analysis, Grammar/Spell Checkers etc., challenges/Open Problems, NLP abstraction levels, Natural Language (NL) Characteristics and NL computing approaches/techniques and steps, NL tasks: Segmentation, Chunking, tagging, NER, Parsing, Word Sense Disambiguation, NL Generation, Web 2.0 Applications : Sentiment Analysis; Text Entailment; Cross Lingual Information Retrieval (CLIR).	<p><a href="https://www.xenonstack.com/blog/evolution-of-nlp/#:~:text=well%20as%20sentences,-.History%20of%20Natural%20Language%20Processing%20(NLP),initial%20period%20of%20the%201960s.">https://www.xenonstack.com/blog/evolution-of-nlp/#:~:text=well%20as%20sentences,-.History%20of%20Natural%20Language%20Processing%20(NLP),initial%20period%20of%20the%201960s.</a></p> <p><b>Chapter 1:</b> Classical Approaches to Natural Language Processing</p> <p><b>Chapter 2:</b> Text Processing</p> <p><b>Chapter 3:</b> Lexical Analysis</p> <p>( Pre-requisite for Unit V)</p> <p>(Reference 1)</p>

		<a href="https://www.codecademy.com/learn/natural-language-processing">https://www.codecademy.com/learn/natural-language-processing</a>
<b>II</b>	Text Processing Challenges, Overview of Language Scripts and their representation on Machines using Character Sets, Language, Corpus and Application Dependence issues, Segmentation: word level(Tokenization), Sentence level. Regular Expression and Automata Morphology, Types, Survey of English and Indian Languages Morphology, Morphological parsing FSA and FST, Porter stemmer, Rule based and Paradigm based Morphology, Human Morphological Processing, Machine Learning approaches.	<p><b>Chapter 4:</b> Syntactic Parsing</p> <p><b>Chapter 5:</b> Semantic Analysis</p> <p><b>Chapter 6:</b> Natural Language Generation</p> <p>(Reference 1)</p> <p><a href="https://www.codecademy.com/learn/natural-language-processing">https://www.codecademy.com/learn/natural-language-processing</a></p>
<b>III</b>	Word Classes and Part-of-Speech tagging(POS), survey of POS tagsets, Rule based approaches (ENGTOWL), Stochastic approaches(Probabilistic, N-gram and HMM), TBL morphology, unknown word handling, evaluation metrics: Precision/Recall/F-measure, error analysis.	<p><b>Chapter 10:</b> Part-of-Speech Tagging</p> <p><b>Chapter 17:</b> Speech tagging</p> <p>(Reference 1)</p> <p><a href="https://www.aclweb.org/anthology/C96-2141.pdf">https://www.aclweb.org/anthology/C96-2141.pdf</a></p>
<b>IV</b>	NL parsing basics, approaches: TopDown, BottomUp, Overview of Grammar Formalisms: constituency and dependency school, Grammar notations CFG, LFG, PCFG, LTAG, Feature-Unification, overview of English CFG, Indian Language Parsing in Paninian Karaka Theory, CFG parsing using Earley's and CYK algorithms, Probabilistic parsing, Dependency Parsing: Covington algorithm, MALT parser, MST parser.	<p><b>Chapter 11:</b> Statistical Parsing</p> <p><b>Chapter 18:</b> Chinese Machine Translation</p> <p>(Reference 1)</p> <p>(Indian Language Parsing in Paninian Karaka Theory)</p> <p><a href="https://dl.acm.org/doi/pdf/10.3115/991146.991151">https://dl.acm.org/doi/pdf/10.3115/991146.991151</a></p> <p>(MALT parser)</p> <p><a href="https://www.researchgate.net/publication/200179363_MaltParser_A_Data-Driven_Parser-Generator_for_Dependency_Parsing/link/004635141fe81139c7000000/download">https://www.researchgate.net/publication/200179363_MaltParser_A_Data-Driven_Parser-Generator_for_Dependency_Parsing/link/004635141fe81139c7000000/download</a></p>

<b>V</b>	Concepts and issues in NL, Theories and approaches for Semantic Analysis, Meaning Representation, word similarity, Lexical Semantics, word senses and relationships, WordNet (English and IndoWordnet), Word Sense Disambiguation: Lesk Algorithm Walker's algorithm, Coreferences Resolution: Anaphora, Cataphora.	<b>Chapter 12:</b> Multiword Expressions <b>Chapter 13:</b> Normalized Web Distance and Word Similarity <b>Chapter 14:</b> Word Sense Disambiguation <b>(Reference 1) Also links with chapter 3</b>

<b>Books and References: (Please mention the unit-wise chapters)</b>					
<b>Sr. No.</b>	<b>Title</b>	<b>Author/s</b>	<b>Publisher</b>	<b>Edition</b>	<b>Year</b>
01	<i>Handbook of Natural Language Processing</i>	Indurkha, N., & Damerau, F. J.	CRC Press Taylor and Francis Group	2 <sup>nd</sup>	2010
02	<i>Speech and Language Processing</i>	Martin, J. H., & Jurafsky, D.	Pearson Education India	2 <sup>nd</sup>	2013
03	Foundations of Statistical Natural Language Processing	Manning, Christopher and Heinrich, Schutze	MIT Press	1 <sup>st</sup>	1997
04	Natural Language Processing With Python	Steven Bird, Edward Loper	O'Reilly Media	2 <sup>nd</sup>	2016
05	Video Links 1. <a href="http://www.nptelvideos.in/2012/11/natural-language-processing.html">http://www.nptelvideos.in/2012/11/natural-language-processing.html</a>				

## Course Outcome

CO1: Students will get idea about know-hows, issues and challenge in Natural Language Processing and NLP applications and their relevance in the classical and modern context.

CO2: Student will get understanding of Computational techniques and approaches for solving NLP problems and develop modules for NLP tasks and tools such as Morph Analyzer, POS tagger, Chunker, Parser, WSD tool etc.

CO3: Students will also be introduced to various grammar formalisms, which they can apply in different fields of study.

CO4: Students can take up project work or work in R&D firms working in NLP and its allied areas.

CO5: Student will be able to understand applications in different sectors

**(Practical)**

<b>M. Sc (Information Technology)</b>		<b>Semester – III</b>	
<b>Course Name: Natural Language Processing Practical</b>		<b>Course Code: PSIT</b>	
<b>Periods per week (1 Period is 60 minutes)</b>		<b>4</b>	
<b>Credits</b>		<b>2</b>	
		<b>Hours</b>	<b>Marks</b>
<b>Evaluation System</b>	<b>Practical Examination</b>	<b>2</b>	<b>50</b>
	<b>Internal</b>	<b>--</b>	<b>-</b>

**List of Practical: (Can be done in Python/ Tensorflow or any imperative language)**

	<b>Two Practical Assignments on each unit of the syllabus.</b>
--	--