14M1WCI333: Natural Language Processing

Course Credit: 3
Semester: M.Tech, I

Introduction

Natural language is used by native speakers in day-to-day communications among themselves. Unlike formal computer languages, natural languages are not precise, often ambiguous, evolve with time with a large numbers of variations and nuances. As all the raw information get generated in natural language form, they need to be processed in different contexts with varying perspectives. Retrieval of relevant information, its categorization, summarization, question-answering, machine translation are numerous natural language processing applications that encompass almost all walks of life that involve computer-human interaction. These include education, business & commerce, governance, judiciary, entertainment, criminology etc. It is a real challenge to build such computer systems with human competing performance. In this course you will learn about theory and practice of building such systems. In a multilingual country like ours, the research & development in this area has a direct impact on our economic and social development. The course will have lectures that will be highly interactive. Our presentations will derive examples from English and Hindi (and other Indian languages) and will provide ample opportunity for you to be innovative.

Course Objectives (Post-conditions)
Knowledge objectives:

On completion of the course, you are expected to develop
1. Ability to analyze natural language phenomenon and create models mimicking human behavior.
2. Ability to design and build NLP application systems in real life based on societal needs.
3. Ability to evaluate NLP software systems.
4. Ability to develop software systems that learn from real life data
5. Ability to re-engineer components of a NLP software in a multi-lingual environment and evolving language scenario.
6. Ability to translate a specification into a design, and then realize that design practically, using an appropriate processing tools and methods.

Application objectives:
At the end of the course, it is expected that you will be in a position to develop applications in document processing, OCR, transliteration, translation, speech processing, question-answering systems, chat box, summarization, sentiment analysis, advertizing, entertainment industry, criminology, e- governance, tourism, health, environment, agriculture, etc.

Expected Student Background (Preconditions)
Knowledge of CS's Core and System Courses on Programming, Data Structure, Algorithms, Artificial Intelligence.

Topics Outline:

S NO	Topics	Hrs

1	Introduction, NLP applications, Ambiguities, Role of syntax, semantics, context, world-knowledge & pragmatics, Rule-based approach, vs. Statistical approach to analysis	3
2	Fundamental aspects of English grammar; Contrastive examples from Hindi, Linguistic resources, Morphological Analysis and synthesis,	4
3	Syntactic Analysis, TN, RTN, ATN, CFG, Probabilistic CFG,	7
4	Collocations, Multi-word expressions, Named entities,	4
5	Language modeling, Markov models, Hidden Markov Models, Parts of speech tagging,	8
6	Word Sense Disambiguation, Pronoun reference disambiguation,	5
7	text similarity, Text categorization, Text Summarization,	3
8	Sentiment Analysis,	2
9	Information Retrieval, Question Answering,	3
10	Natural Language Generation, Machine Translation.	3
	Total	42

References

1. Daniel Jurafsky & James H.Martin, " Speech and Language Processing", Pearson Education (Singapore) Pte. Ltd., 2002.
2. James Allen, "Natural Language Understanding", Pearson Education, 2003
3. C.D. Manning and H. Schutze, "Foundations of Statistical Natural Language Processing", MIT Press, 1999.

<u>Evaluation Scheme:</u>

S.No	Examination	Marks
1	T-1	15
2	T-2	25
3	T-3	35
4	*Internal Marks	25

*Internal Marks Breakdown:

Assignments   9 marks (3x3)

Quizzes   12 marks (3x4)

Regularity   4 Marks