

Natural Language Processing

Lecture-0

Welcome!

- Course Code: 15CSE358
- Title: Natural Language Processing
- Faculty: Manu Madhavan (m_manu@cb.amrita.edu, 7736735850)
- Lectures:
 - Monday slot 1
 - Tuesday slot 2
 - Tuesday slot 1

Some intuition



How old Obama



All

News

Images

Videos

Maps

More

Settings

Tools

About 23,60,00,000 results (1.05 seconds)

Barack Obama / Age

59 years

4 August 1961

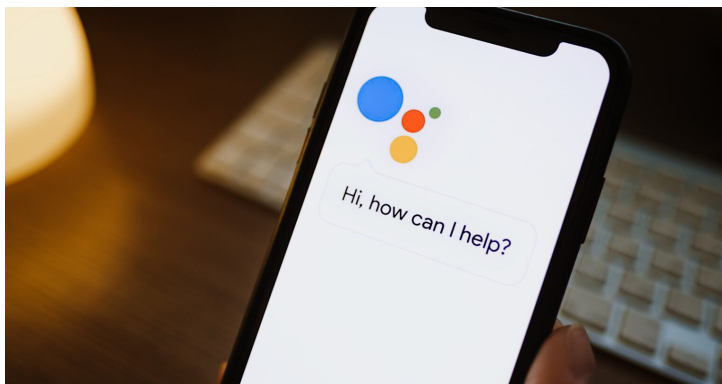


Barack Obama



44th U.S. President

Barack Hussein Obama II is an American politician and attorney who served as the 44th president of the United States from 2009 to 2017. A member of the



English – detected



Tamil

How are you



Both houses of the Indian Parliament use the Simultaneous Interpretation System for instant translation of a speech and its simultaneous delivery. The system was initially used for translating from Hindi to English and vice-versa in 1964 (for both houses). Currently, this system has been expanded for translation across the following languages - Lok Sabha: Kannada, Marathi, Tamil, Punjabi, Malayalam, Manipuri, Sanskrit, Telugu and Oriya. Rajya Sabha: Assamese, Bengali, Tamil, Telugu, Kannada, Gujarati, Malayalam, Marathi, Urdu, Oriya.

எப்படி இருக்கிறீர்கள்
Eṭṭ

Malayalam

എന്താക്കെയുണ്ട്

Hindi

क्या हाल है

kya haal hai

- Automatic reply suggestion
- Sign board detection
- Spell correction
- Word suggestion
- Sentence generation
- Text Mining

Natural Language Processing

- Field focuses on communication between computers and humans in natural language and NLP is all about making computers understand and generate human language
- Languages are the highest form of human cognition
- Try to model language properties using computational model
 - Probability of next word?
 - Most probable answer
 - Most similar word (translation)
- NLU and NLG

Focuses of the course

- Foundation of Natural Language Processing
- Applications
- Language Modeling
- NLP at various stages

Key Features

- Mix of theory and application
- Latest trends in the research/industry will discuss
- Practical sessions
- Research oriented
- Talks and workshops

Course Outcome

COs	Course Outcome	Bloom's Taxonomy Level
15CSE458.CO 1	Understand and represent the linguistic phenomena at morphological, word, syntactic and semantic levels of Natural language processing	L2 - Understand
15CSE458.CO 2	Design the linguistic phenomena using formal language and probabilistic model	L3 - Apply
15CSE458.CO 3	Understand the ambiguity in language at morphological, word, syntactic and semantic levels and apply appropriate methods to avoid them	L4 - Analyze
15CSE458.CO 4	Understand the probabilistic parsing and statistical alignment.	L2 - Understand

Major Topics

- Introduction to NLP
- Probability: Recap
- Linguistic essentials
- Statistical NLP
- Applications- NER, WSD, Topic models
- Language Models
- Information Retrieval
- Machine translation
- Text mining

Syllabus

Unit 1

Introduction: Words – Morphology and Finite State transducers - Computational Phonology and Pronunciation, Modelling - Probabilistic models of pronunciation and spelling – Ngram Models of syntax - Hidden markov models and Speech recognition - Word classes and Part of Speech Tagging.

Unit 2

Context free Grammars for English – Parsing with Context free Grammar – Features and unification -Lexicalized and Probabilistic Parsing -Language and Complexity. Semantics: Representing meaning -Semantic analysis - Lexical semantics - Word sense disambiguation and Information retrieval

Unit 3

Pragmatics: Discourse - Dialog and Conversational agents - Natural language generation, Statistical alignment and Machine translation: Text alignment – word alignment – statistical machine translation.

Evaluation Pattern

Component	Online	Viva	Total
Mid-term assessment	10	20	30
Continuous Assessment (Programming Assignments (2), Case study, Quizzes (5))	20 (Programming-7 Case study-7 Quizzes-6)	-	20
External Assessment - End Semester	20	30	50
Total			100 marks

Textbooks/References

1. C.D. Manning et al, “**Foundations of Statistical Natural Language Processing**,” MIT Press. 1999. isbn: 9780262133609.
2. Daniel Jurafsky and James H. Martin ”**Speech and Language Processing: An Introduction to Natural Language Processing**, Computational Linguistics, and Speech Recognition,” 1st. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2000. isbn: 0130950696.
3. James Allen, “**Natural Language Processing with Python**”, O’Reilly Media, July 2009.
4. Ian Goodfellow, Yoshua Bengio, and Aaron Courville, **Deep Learning**, <http://www.deeplearningbook.org>. MIT Press, 2016.
5. Stephan Raaijmakers, **Deep Learning for Natural Language Processing**, Manning Publications, 2020.
6. Noah A. Smith, ”Linguistic Structure Prediction. Synthesis Lectures on Human Language Technologies,” Morgan and Claypool, May 2011.

Additional References

- [Applied Natural Language Processing - Course](#) (CMI, NPTEL)
- [CS224N: Natural Language Processing with Deep Learning](#) (MIT winter course)
- [NLP Specialization, coursera](#) (Deeplearning.ai)

Any Queries?

Next

- Formal introduction