UNIVERSITY of HOUSTON | COMPUTER SCIENCE

COSC 6397: Natural Language Processing

New Course in FALL 2014

MW: 1:00pm - 2:30pm
Classroom: SW 221

Course web page: http://www.cs.uh.edu/~arjun/courses/nlp

Instructor: Dr. Arjun Mukherjee
(arjun at cs.uh.edu)
Department of Computer Science

The tremendous growth of the Web and the surge of **social media** render massive amount of data that present new forms of actionable intelligence. In the domain of Web and social media, unstructured test ranges from scientific articles, to news archives, including social conversations. How to mine latent knowledge hidden in these vast amounts of textual data? How do we make sense of peoples' opinions on various topics? How can we predict gender, mood, age, etc. from language? Being able to solve these problems in Web using language processing has important future implications on big data and informatics. We have seen a glimpse of this gigantic potential of language processing in the famous IBM Watson which employs state-of-the-art techniques.

In this course, students will first learn the fundamentals of natural language processing. Starting with basics of probability, students will learn how computational approaches are used in language modeling and part-of-speech tagging. Students will also be introduced to the machine learning approaches and algorithms for performing important text mining tasks such as text categorization and clustering. The course will also throw light upon sentiment analysis and opinion mining which are popular and trending areas in NLP. Students will also learn how to apply natural language processing techniques to practical tasks as above.

Prerequisites: *The course assumes background in basic math and/or probability and proficiency in any high level programming language (e.g., Java). If you have taken and did well in one or more of the equivalent courses/topics such as Algorithms, Artificial Intelligence, Numerical methods, or have some background in probability/statistics, it will be helpful. The course however reviews and covers required mathematical and statistical foundations.*