Syllabus

Course Description:

Natural language processing (NLP) deals with the application of computational models to text or speech data. Application areas within NLP include automatic (machine) translation between languages; dialogue systems, which allow a human to interact with a machine using natural language; and information extraction, where the goal is to transform unstructured text into structured (database) representations that can be searched and browsed in flexible ways. NLP technologies are having a dramatic impact on the way people interact with computers, on the way people interact with each other through the use of language, and on the way people access the vast amount of linguistic data now in electronic form. From a scientific viewpoint, NLP involves fundamental questions of how to structure formal models (for example statistical models) of natural language phenomena, and of how to design algorithms that implement these models.

In this course you will study mathematical and computational models of language, and the application of these models to key problems in natural language processing. The course has a focus on machine learning methods, which are widely used in modern NLP systems: we will cover formalisms such as hidden Markov models, probabilistic context-free grammars, log-linear models, and statistical models for machine translation. The curriculum closely follows a course currently taught by Professor Collins at Columbia University, and previously taught at MIT.

Problem Sets:

There were will be 3 programming assignments during the class, due roughly every two weeks.

Syllabus:

Topics covered include:

1. Language modeling.
2. Hidden Markov models, and tagging problems.
3. Probabilistic context-free grammars, and the parsing problem.
4. Statistical approaches to machine translation.
5. Log-linear models, and their application to NLP problems.
6. Unsupervised and semi-supervised learning in NLP.

Readings:

Notes for the class will be posted at http://www.cs.columbia.edu/~mcollins.

Created Mon 25 Jun 2012 10:48 AM PDT

Last Modified Mon 11 Feb 2013 9:38 AM PST