

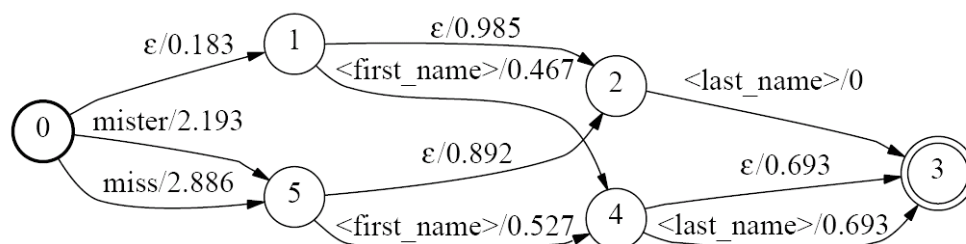
Statistical NLP for the Web

Fall 2012

COMS E6998-4

Time: Wednesday, 4:10 – 6pm

Instructor: Dr. Sameer Maskey



Are you interested in developing a Sentiment Analysis algorithm that uses Twitter fire hose data? Do you want to learn how Hidden Markov Models and Finite State Machines can be used to implement a Spoken Dialog System like Siri? Would you like to understand News clustering algorithm or Maximum Entropy based Question Answering systems? This course will explore topics that juxtapose Statistical/Machine Learning algorithms with real world NLP/Speech tasks that use a large amount of web data. We will study NLP/Speech topics such as Text Mining, Document Classification, Topic Clustering, Summarization and Dialog systems. We will explore Statistical Methods/Machine Learning techniques such as Linear classifiers, Clustering techniques, Inference algorithms, Ranking Methods that are used in addressing some of these NLP/Speech problems. Students will get hands-on experience in implementing some of these techniques efficiently to build an NLP/Speech system that can handle a significant amount of unstructured web data (text, speech and video).

For more information go to course website:
<http://www.cs.columbia.edu/~smaskey/CS6998/>