About the course:
Natural language is available in two modalities: speech and text. This course will be exclusively concerned with text. Also, the course will concentrate to a very large extent on statistical techniques and learning based approaches.

Some common NL tasks are: translate from language L1 to L2; answer questions based on some text; summarize text; retrieve relevant information; tag text for sentiment, relations, entities, categories/key words; construct an ontology; generate NL text.

These NL tasks usually require one or more text processing steps like: part of speech tagging; named entity recognition; anaphora resolution; word sense disambiguation; shallow and deep parsing; morphological processing, chunking; semantic tagging and meaning representation.

There are statistical techniques/algorithms that can be used for realizing these processing steps. Some of these are: n-grams; HMMs; CRFs; vector-space embedding representations; deep network approaches; LSA, PLSA; amongst others.

The course will study the use of the above techniques for carrying out the basic processing steps to ultimately address some of the NL tasks.

Pre-requisites:
Necessary: Esc101N/equivalent; CS210A/CS210/ESO207A/ESO207/ESO211.
Highly desirable: ML course done earlier or concurrently. This will become a pre-requisite next time onwards.

The course will have programming assignments and a project. So, if you are not comfortable writing programs, using libraries etc. then you are strongly urged to not take this course.

References:
1. Dan Jurafsky, JH Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition, 2nd Ed, Pearson 2009. Softcopy of 2nd Ed. available online. Draft of parts of 3rd Ed. available online.
4. Recent papers from journals and conferences.

Note: The changed contents are being sent for necessary approvals.