Instructor: Chitta Baral

Description of the course: In this course we will learn the science and engineering behind building a question answering system using natural language understanding and reasoning.

The theoretical content that will be covered in the course includes:

- Knowledge Representation
- Translating natural language to knowledge representation languages
- Probabilistic Combinatorial Categorial grammar
- Lambda Calculus and Inverse Lambda
- Machine learning approaches to natural language processing
- Ontology development
- Building a natural language corpus

There will not be any textbook for the course. The course material will consist of several chapters from books and have several research papers.

As part of the course groups of 3-4 students will build a system that can take natural language text as input and can answer questions about them. The system will be a learning based system that will take as input natural language sentences and their translations to appropriate knowledge representation languages. It will then learn how to do such translations for new unseen sentences. The system will then use such translations to answer questions. We expect our systems to have deeper reasoning ability than the WATSON system developed by IBM that could beat the Jeopardy champions.

Some of the domains that will be offered for student projects are:

- Combinatorial puzzles
- Wordnet glosses
- Framenet annotations
- Planning domains expressed in natural language
- Robocup commands
- Geoquery
- Chapters of high school Biology, Chemistry and Physics books
- Human robot interaction languages
- Policy descriptions in natural language
- Natural language descriptions of databases
- Natural language text in Archaeology
- Natural language text from Pubmed

The grading will be based on a test (40%), class participation and presentation (10%) and on the project (50%).