11-411 Progress Report #1

Bobbie Chen, Naixin (Natalie) Fan, Brandon Jin, Phapimol (Baseball) Yoovidyha

Planned Solution

There are two parts to our planned solution.

The first part uses keyword recognition and syntax parsing to generate and answer questions. For example, the sentence “The tree fell in the forest” could be converted into the question “Where did the tree fall?” In the other direction, the question “Where did the tree fall?” could be answered by searching through the reference text for the keywords “tree” and “fall” followed by a preposition.

The second part uses information retrieval. First, we build up a database of relationships between nouns in the article, categorizing the nouns into narrow classes similar to named entities. Then, we can ask questions about the relationships using pre-generated formats, and answer questions by finding the relationships between the nouns that were involved in the question.

Progress

We have begun working on a keyword-recognition based question answering system. To narrow the search space, we begin by separating the text into sentences using NLTK. Then, each sentence is converted into a hash set of word stems using the Snowball Stemmer from NLTK. We can then focus our efforts on sentences which contain many of the same stems as the question. The next step (which has not yet been implemented) is to convert questions and high-match sentences to parse trees to look for the answer.