First Project Progress Report

Team “Error 501: Not Yet Implemented”

Aaron Anderson, Eric Gan, Rachel Kobayashi

February 25th, 2014

Since our last meeting we have established a general timeline with tasks and goal deadlines.

|  |  |
| --- | --- |
| Look up and research NLP tools. Look for libraries. Research ways to store data to files so we don’t have to build every time. | 2/22/2014 |
| Categorize questions to be answered. Can we string-match the question? Can we look at sentence structure? syntax tree? | 2/22/2014 |
| Try to solve easy questions | 2/25/2014 |
| Try to solve medium questions | 3/1/2014 |
| Begin developing questions: how to create fluent syntactically-correct questions? | 3/8/2014 |
| Try to solve hard questions | 3/15/2014 |
| Try to create hard questions to stump other people. | 3/22/2014 |
| **COMPLETE PROJECT** | 4/1/2014 |

In our current phase we are examining the functionality the of NLTK toolkit. This has provided us a basis for the type of functionality we want and/or will be using. Currently our approach to answering questions relies on inferring some information from the sentence structure of the question. This has led us to examine and attempt to use the part-of-speech tagging included in the NLTK toolkit.

We have also discussed about how we want to process the source Wikipedia article. We decided it would be helpful to break the article up into sentences, since that is how information is natural divided. We also talked about breaking the content down into N-grams to help us efficiently search through data. We have also discussed and are in the process of using the HTML tags to our advantage since the tags provide a hierarchy for how information is ranked.

We have concerns about the efficiently of our project so we are continuing to research other toolkits or algorithms that may be useful to us. Additionally we are not limiting ourselves to using the NLTK toolkit but we thought it would be worthwhile to understand how it works and to use as a starting point for our project.