11-611 Natural Language Processing
Initial Plan

Group Members
Siyu Chen, Xiaoqiu Huang, Jiachen Li
`siyuche, xiaoqiuh, jiachenl`

1 Overall Plan

We are going to use mainly Python & Java for system development in this project. The project contains two part: one for information extraction and question generation, another for question understanding and answering. Our goal is to build a back end system, which can extract fluent and reasonable questions from Wikipedia documents and generate fluent, correct answers to questions which are generated by human and program. The toolkits we may use are NLTK library, Berkeley parser, WordNet, Lancaster Stemmer.

2 Detailed Plan

(1) Feb. 6 - Feb. 24
Define the different question types, design the detail of Q&A system, and implement the prototype of the information extraction and question generation part.

(2) Feb. 25 - Mar. 19
Revise and improve the question generation part. Design and implement question understanding and answering part.

(3) Mar.20 - April. 7
Test the whole Q&A pipeline, fix bugs and revise the system design.

(4) April. 8 - April. 14
Optimize the system performance and start to summarize the work.

(5) April. 15 - April. 21
Write and revise final report writing.