Initial Progress Plan

High-level design

(tools to use)
We will base our text parsing on Stanza, using its pipelines for tokenization, POS tagging, lemmatization, etc. We will possibly use spaCy for assistance if necessary. Then we would like to use the English Wordnet to find various appropriate replacements of words to obtain more accurate searching within the Wikipedia articles.

(how to use development data to improve)
We will first target on giving reasonable answers to relatively easy questions. Then we will test our system on more difficult questions to see what the outcome is. Based on the results we will add more features and details into our system.

(relationship between asking and answering)
We would like to first implement an answering system independent from the asking system. If efficiency requires, we will also try incorporating inside the answering system a simple question-generating system with answers stored, so when asked an easy-level question, it may directly search for the answer before going into the parsed article information.

(technical approach)
Due to the size of packages and capability of CPUs, we want to edit and run our projects on the Virtual Machines for most of the time. We might have to seek help from course staff or departmental staff on this part for more technical support (in setting up environments, installing packages, etc.) Besides, for the sake of time, simple functions will also be tested directly on local computers.

Cooperations

(how to share code date)
We will create and share a git repository. We shall push commits after every major edit.

(coordinate)
We are not dividing up the work into disjoint subsets and assign each to one person. Instead, we will proceed through every step collectively and take turn to do the editing job. We will keep in contact through WeChat group and zoom meeting.