

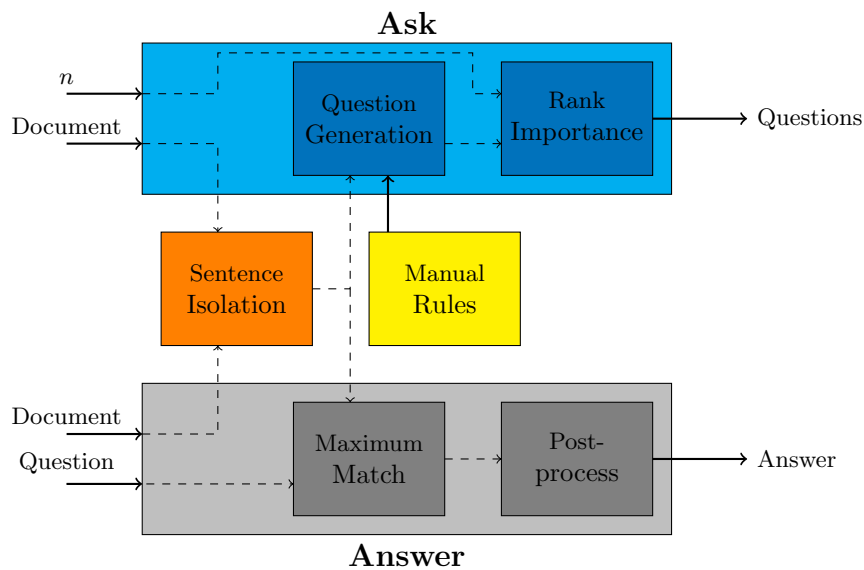
# Initial Progress Report

Haya Al-Kuwari, Akhyar Kamili, Abubaker Omer, Mohammed Nurul Hoque

October 31, 2018

---

## 1 The Pipeline



### 1.1 Sentence Isolation

Reference resolution and sentence segmentation so that each sentence is  $\sim$ independent. This should also remove parenthetical expressions and put them as standalone sentences.

### 1.2 Manual Rules

A set of manual rules for generating a question from a sentence (using parse tree?).

### 1.3 Questions Generation

Apply the rules on sentences to generate the questions (also maybe add random variations of each generated sentence).

### 1.4 Rank Importance

Rank all questions based on the topic and output highest  $n$ .

### 1.5 Maximum Matching

Return sentence that matches the question topic the most.

### 1.6 Post-process

Handle negation, etc.

## 2 Tools and libraries

We checked out NLTK and Stanford NLP core both of which have useful functionalities for our project, e.g. POS tagger, parser, reference resolution. Other tools might be required as we starting implementing and learn the actual requirements of each task

## 3 Team Logistics

We decided to initially divide the work by component as follows:

- Sentence isolation: Haya,
- Question generation: Abubaker,
- Rank importance: Mohammed,
- Maximum match: Akhyar

We will compile a list of maual tools together and decide for post processing later as it is not very well defined yet.