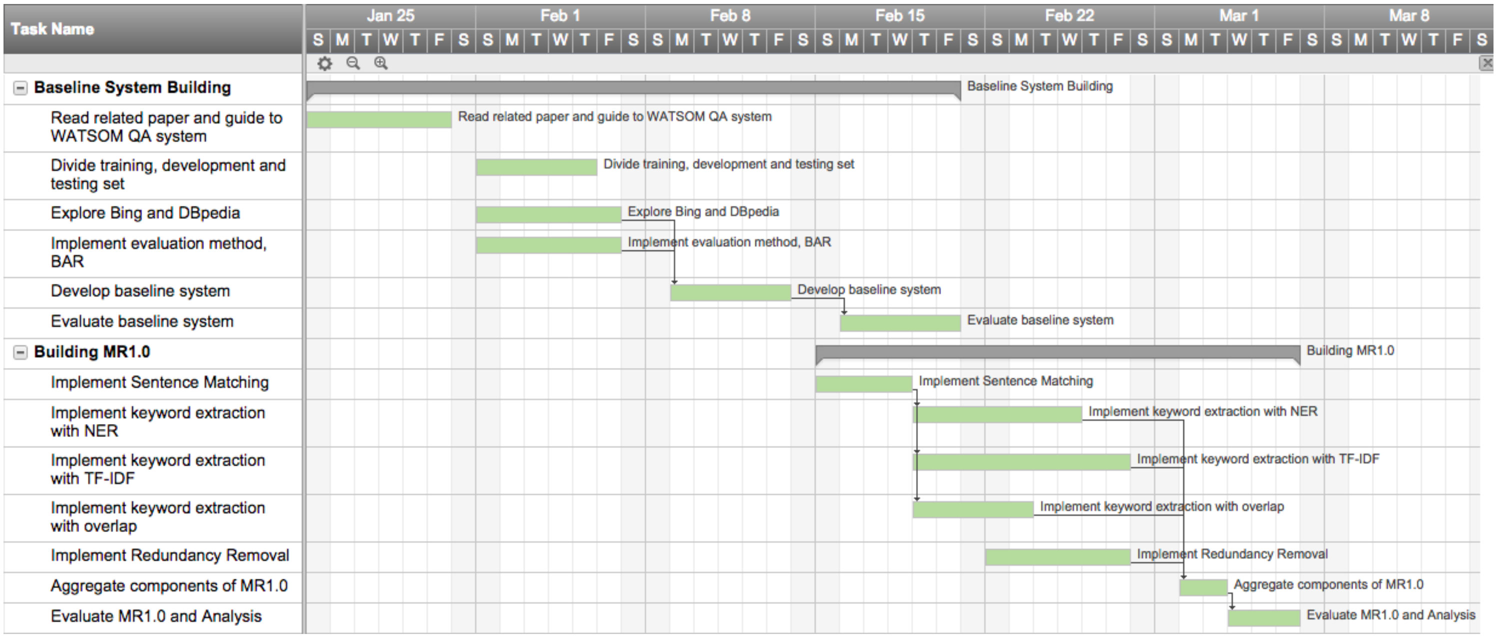
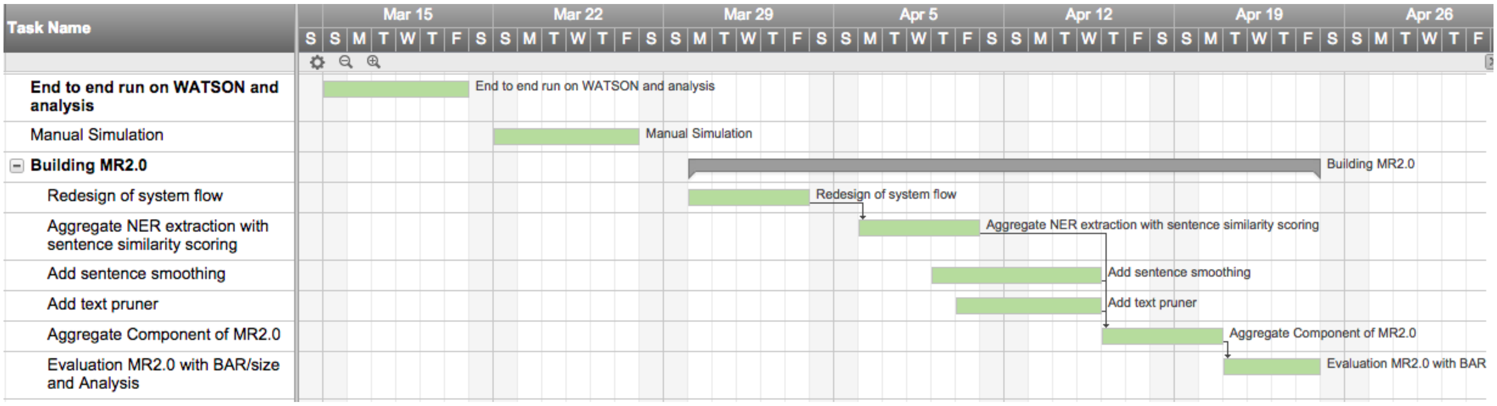
Plan Document

Here’s the Gantt Chart of our plan:



Task Name
Baseline System Building
Read related paper and guide to WATSOM QA system
Divide training, development and testing set
Explore Bing and DBpedia
Implement evaluation method, BAR
Develop baseline system
Evaluate baseline system
Building MR1.0
Implement Sentence Matching
Implement keyword extraction with NER
Implement keyword extraction with TF-IDF
Implement keyword extraction with overlap
Implement Redundancy Removal
Aggregate components of MR1.0
Evaluate MR1.0 and Analysis



Task Name
End to end run on WATSON and analysis
Manual Simulation
Building MR2.0
Redesign of system flow
Aggregate NER extraction with sentence similarity scoring
Add sentence smoothing
Add text pruner
Aggregate Component of MR2.0
Evaluation MR2.0 with BAR/size and Analysis

Work Decomposition:

0. Explore relate work and read Waston APIs:
Date : Jan 25th, 2015 - Jan 31st, 2015
Description: Read papers related to Machine Reading and guide to WATSON.

1. Divide the dataset evenly into train, test and dev:
Date : Feb 1st, 2015 - Feb 7th, 2015
Description : We manually divided the question dataset into train, test and development sets, and made sure that each of the sets had evenly addressed different types of questions - Who,

What, When, Why etc. We then ran all our experiments again on these finely defined datasets.

2. Interfacing with DBPedia and entity extraction:
Date: Feb 1st, 2015 - Feb 7th, 2015
Description: Set up code that can interface with DBpedia to get information about entities. Developed the part of the current system that reads the data from the training questions and bing results. It then spots and annotates the entities from DBPedia and retrieves those entities to add as text to the corpus to be sent to Watson.

3. Set up the initial end to end baseline system:
Date: - March 1st, 2015 - March 7th, 2015
Description: We set up the baseline system, that could get document snippets from the Bing search engine for the given training queries of “Terrorism” dataset. We then split the gold-standard answers into informative unigrams and computed the train, dev and test set Binary Answer Recall scores.

4. Retrieve full content and evaluate BAR on answer phrases:
Date : March 1st, 2015 - March 7th, 2015
Description : We now retrieved the full document from the search engine instead of relevant text snippets, for the given training set questions and computed BAR for the gold standard answer phrases (no preprocessing) .

5. Implementation of Source Expansion(actually it’s querying search engines and get raw corpus)
Date : Feb 15th, 2015 - Feb 28th, 2015
Description: Set up code for querying Bing search engine with example questions.

7. Implementation of NER extraction
Date : Feb 15th, 2015 - Feb 28th, 2015
Description: Use NER extraction to extract named entities in the corpus. The named entities can then be used to form the query for the next iteration of our source expansion system.

8. Extract keywords using overlap:
Date : Feb 15th, 2015 - Feb 28th, 2015
Description: Implement an overlap keywords extraction method and do experiment with it.Tested it on full document retrieval and abstract retrieval situation.

9. Extract keywords using TF-IDF:
Date: Feb 15th, 2015 - Feb 28th, 2015
Description: Implement TF-IDF score calculating method to extract keywords from retrieved corpus. In extracting keywords using TF-IDF method, the terms with highest TF-IDF scores are returned as queries for next iteration of source expansion.

10. Remove Redundancy of corpus:
Date: Feb 15th, 2015 - Feb 28th, 2015
Description: Built a module to remove exact duplicate and near duplicate copies of sentences and documents from the corpus.

11. Aggregation of components for MR1.0
Date : March 1st, 2015 - March 7th, 2015
Description: Aggregate several modules into a whole system as MR1.0: Source Expansion, Evaluation and Pseudo Document Generation, etc. Made the whole system easier to development and test.

12.End to end run with WATSON
Date : March 4th, 2015 - April 15th, 2015
Description : Developed module to save the corpus in a format that can be ingested by WATSON. Manually label the answers to the training questions within the corpus. Test WATSON QA system with BAR metric.

13. Question Analysis
Date : March 4th, 2015 - April 15th, 2015
Description : Leverage the first step, question understanding, of Question Answer System in source expansion part. The initial queries or seeds have great influence on following steps in source expansion. So question analysis becomes important at the beginning of this process. In terms of question analysis, we want to achieve following goals:
1). Find the topics of given queries. 2) Extract the keywords and relationship between them for generating more new queries that in different format or context.

14. Standardize the data
Date : March 4th, 2015 - March 25th, 2015
Description : Standardize the date time and name entities format in both query and corpus. We need to come up with an efficient way to perform answer string matching.

15. Redesign of system flow
Date : March 4th, 2015 - March 25th, 2015
Description: Add more component to our system based on the manually simulation. Redesign our system to cope the new components and new evaluation metric.

16. Use noun, verbs and Named Entities from documents to find relevant sentences.
Date: April 29th - May 11th
Names of individuals: Kavya Srinet and Xiaoqiu Huang
Description: Based on what we found from the task above , we look for sentences that have all the Named Entities as mentioned in the query, the nouns in the query and the verb in the

query to look for answer bearing sentences. We then use these sentences and extract Named Entities from these to expand further on. We used the Stanford parser to do the above.

17. Implement Text Smoothing
Date: April 29th - May 11th
Description: Return a summary of the paragraph that contains answer or elements in the evidence list. The summary contains a smoothing window of the target sentences, the first sentence and last sentence of the paragraph in which the target sentence showed up.

18. Experiment with BAR against corpus size
Date: April 19th - April 26th
Description: We plan to experiment with a lot of other evaluation metrics and the tradeoff between the corpus size, the number of keywords extracted, the number of iterations for expansion, the number of documents retrieved from the search engine, the size of development set and test set, combination of Binary Answer Recall and corpus size.