

Milestone #1: Q4MRE Initial Write-up

Version: 0.1

Fall 2013 11-791

11/4/2013

Jeffrey Gee

Zhiyul Li

Laleh Roostapour

Adam Rosini

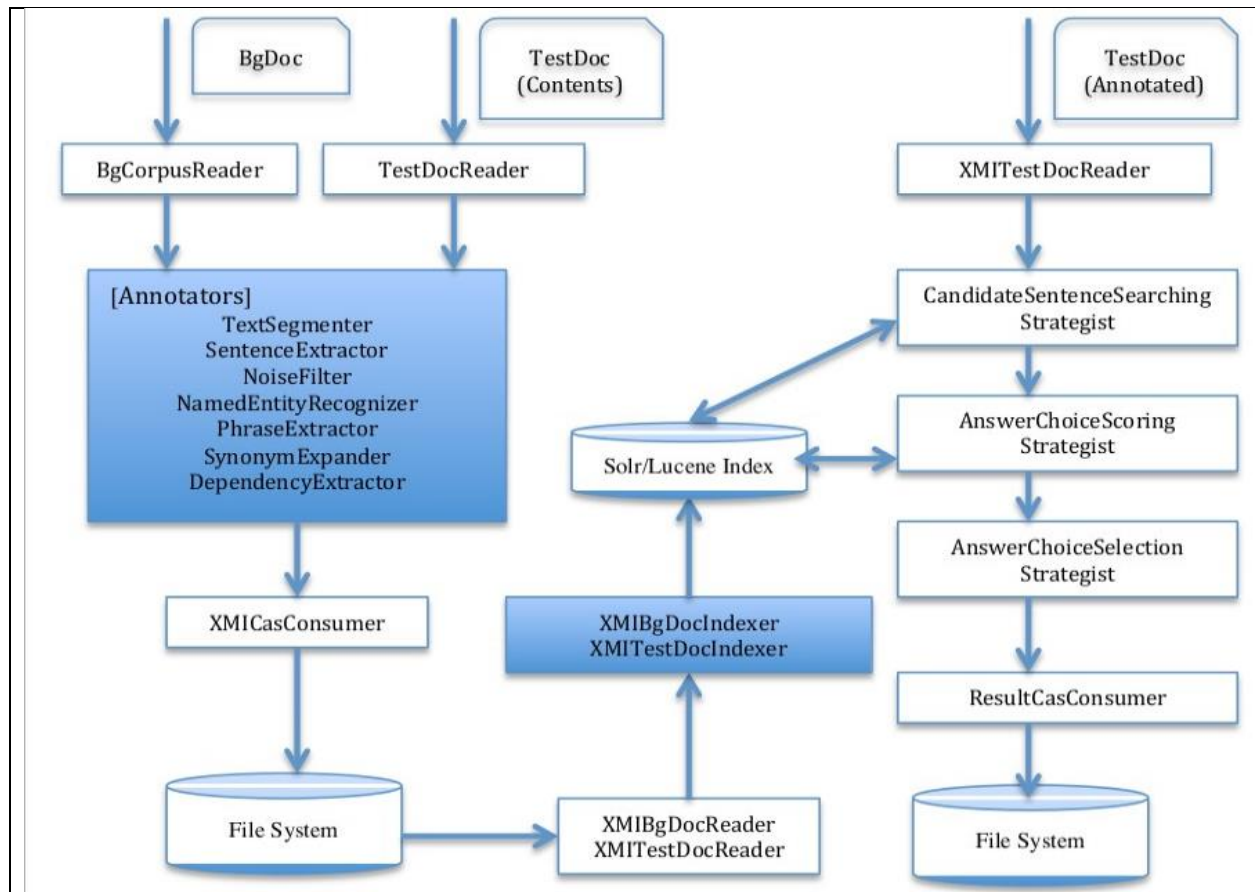
Zhengxiong Zhang

Contents

1	Initial Pipeline.....	3
2	Initial Type-System Design	4
3	Annotators	5
4	Initial Division of Work.....	5
4.1	Mini-Iteration 1: Investigation Phase.....	5
4.2	Areas of Focus	6

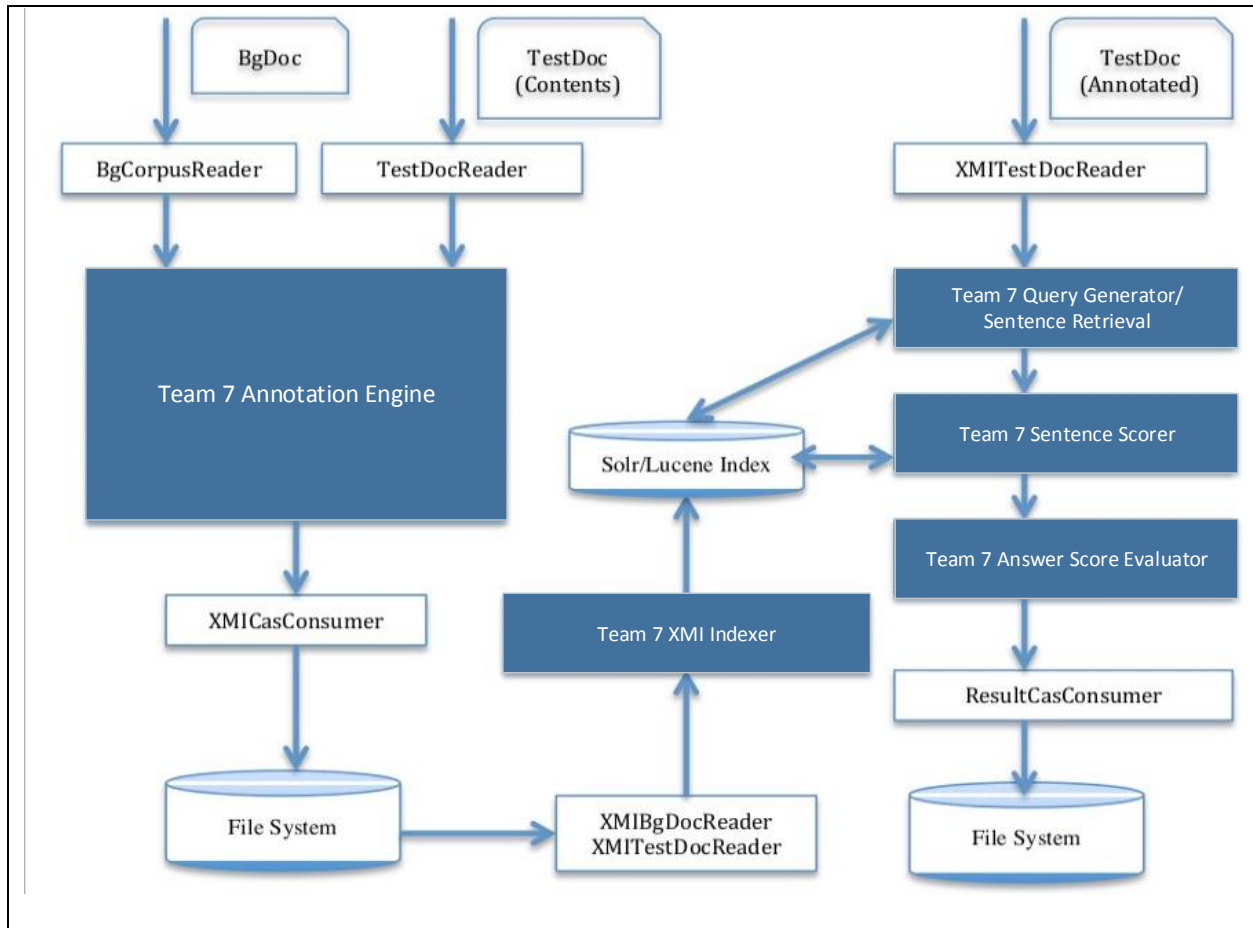
1 Initial Pipeline

The initial pipeline will work similarly to that of the baseline code (qa4mre-alzheimers project). This may later be adjusted, based on internal discussions throughout development iterations.



By adopting some existing components (such as document readers in the qa4mre project, we can focus on the components that most dictate the performance of the system. We have identified the following as key areas for development:

- Annotator Chain
- XMI Indexer
- Candidate Sentence Searching Strategist
- Answer Choice Scoring Strategist
- Answer Choice Selection Strategist



Our pipeline will attempt to improve these five components by implementing our own methods for annotating, indexing, searching, and scoring.

2 Initial Type-System Design

Our type-system will leverage the initial type system utilized in q4mre-base:

- Answer
- CandidateAnswer
- CandidateSentence
- Dependency
- NER
- NounPhrase
- Question
- QuestionAnswerSet
- Sentence
- Synonym

- TestDocument
- Token

Additionally, our initial type-expansion list includes:

- AnswerScore
- LineGroup
- NamedEntity
- QuestionType(s)
 - Factoid
 - Causal
 - Method
 - Purpose
 - True/False

3 Annotators

In addition to the Annotators that currently exist in the q4mre-base and q4mre-alzheimers projects, we will implement annotators including:

- Line Groupers
- Question Type Annotators
- Named Entity Annotators
- Additional Answer Scorers.

4 Initial Division of Work

As we are not fully aware of how much time each task will take from the start, we will attempt to re-distribute work in each iteration.

4.1 Mini-Iteration 1: Investigation Phase

Our first iteration will be an investigation phase. In this phase we attempt to get the baseline code running, and investigate different components to see what comprises the current benchmark.

Component	Member(s)	Description
Annotation Engine	Laleh, Jeff	Determine what annotation strategies are currently being utilized in the existing QA4MRE pipeline.
XMI Indexer	Zhiyul	Determine how XMI objects are being indexed into Lucene/Solr
Candidate Sentence Searching Strategist	Zhiyul, Jeff	Determine how candidate sentence queries are being formed based on questions
Answer Choice Scoring Strategist	Zhengxi	Determine how candidate answers are being scored.

Answer Choice Selection Strategist	Adam	Determine how the answer selection process works.
---	------	---

4.2 Areas of Focus

Throughout the development cycle, we will all share the work when it comes to actual code development. However, there may be many cases in which specific people may be identified as references for a given field. This identification is done based on the background (area of study) of each team member.

Field	Member(s)
Software Engineering, Presentation	Adam (MISM)
Machine Learning	Zhengxi (VLIS), Zhiyul (VLIS)
Natural Language Processing	Laleh (MLT), Jeff (MIIS)
Solr/Lucene	Zhiyul, Jeff