**Team 11: Initial Report**

**Nov. 2013**

# Team Information

**Team number:** 11

**Team leader:** Di Wang (diw1)

**Team member:** Di Wang (diw1)

Xiaohua Yan (xiaohuay)

Serena Jeblee (sjeblee)

Yuanchi Ning (yuanchin)

# Initial Pipeline / Workflow Design

// Di & Serena? Maybe Di can describe the whole architecture of our project and Serena can talk about the detail of query re-construction in the pipeline here.

# Initial Type-System Design

//Feel free to correct or add something here

Besides the seven given UIMA annotators:

* TextSegmenter
* SentenceExtractor
* NoiseFilter
* NamedEntityRecognizer
* PhraseExtractor (Maybe eliminated based on the future experiments)
* SynonymExpander (Maybe eliminated based on the future experiments)
* DependencyExtractor

Two more annotators will be added into the type system in the initial phase of the project:

* **N-Sentence Annotator:** Similarly to the N-Gram Annotator implemented in the previous homework, the N-Sentence Annotator extracts groups of consecutive sentences in the documents in order to evaluate the overall overlap between the query and the potentially relevant documents. The window size (N) may vary depending on the specific context and the type of given query.
* **SemanticRole Annotator:** Semantic role analysis is an important part in Natural Language Processing, which captures the relationship between a predicate and syntactic constituents. SemanticRole Annotator parses the sentences of the query and the documents to do semantic labeling, and generates semantic arguments to facilitate the subsequent analysis of the task. (SemanticRole Annotator may not be included in the initial implementation of the project, but will definitely be utilized later)

# Baseline Methods to Implement

//Xiaohua & Di? Maybe an initial description of scoring methods could be given by Xiaohua and Di can add some additional methods which is necessary within the whole workflow.

# Initial Division of Work

Di Wang: Engineering and framework setup

Xiaohua Yan: Implementation of baseline scoring method

Serena Jeblee: Implementation of query re-construction

Yuanchi Ning: Implementation of N-Sentence Annotator