LING 573
Deliverable 2

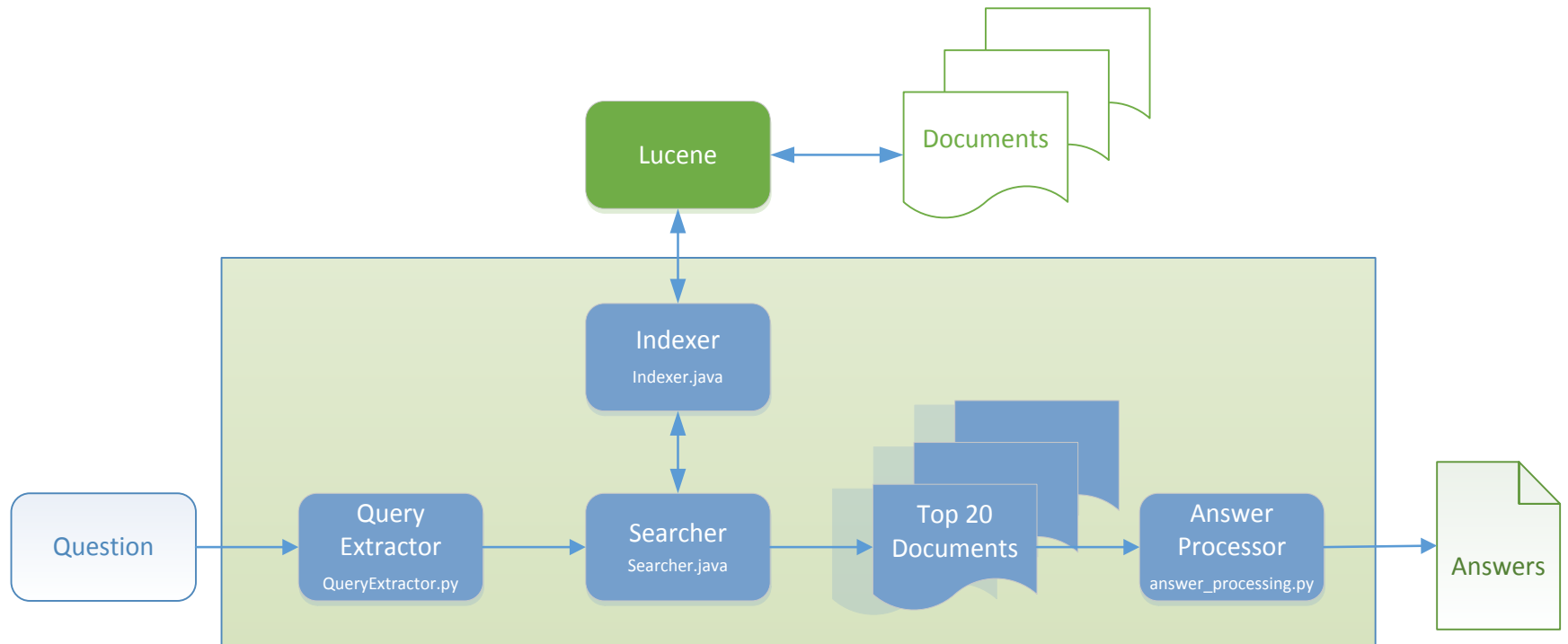RYAN ALDRICH | MATTHEW BENAK | NORAH HOGOBOOM

System Overview

Approach

Query Formulation (primary developer Matt)
- Collect all of the question files and pass them to the query-generation module

Document Retrieval (primary dev Ryan)
- Start up the document-retrieval engine (Lucene), run each query, and pass the resulting top 20 documents to the answer-processor

Answer Processing (primary dev Norah)
- Run the answer processor to find strings matching query, extract the passage from the document, and print the output in the requested format

Results

The only way to go is up!

0.0	Strict
0.21	Lenient

The answer processor searches(via RegEx) for *any* of the terms in the query, and returning the 250 characters that appear after the first hit in the document.

D2 Assessment

Challenges:

- Software versions available on cluster
- Integrating Python and Java
- Changing software and formats at last minute
- Forming as a virtual group
- Very different availability schedules

Successes:

- The system works end-to-end
- Software options stable
- Team coming together
- Team schedule and strengths understood

Next Steps

Query Formulation

◦ Improve queries and add types

Document Retrieval

Improvements in indexing and retrieval

Answer Processing

Implement and evaluate a number of different approaches.