11693-02 final presentation

Biomedical question answering system
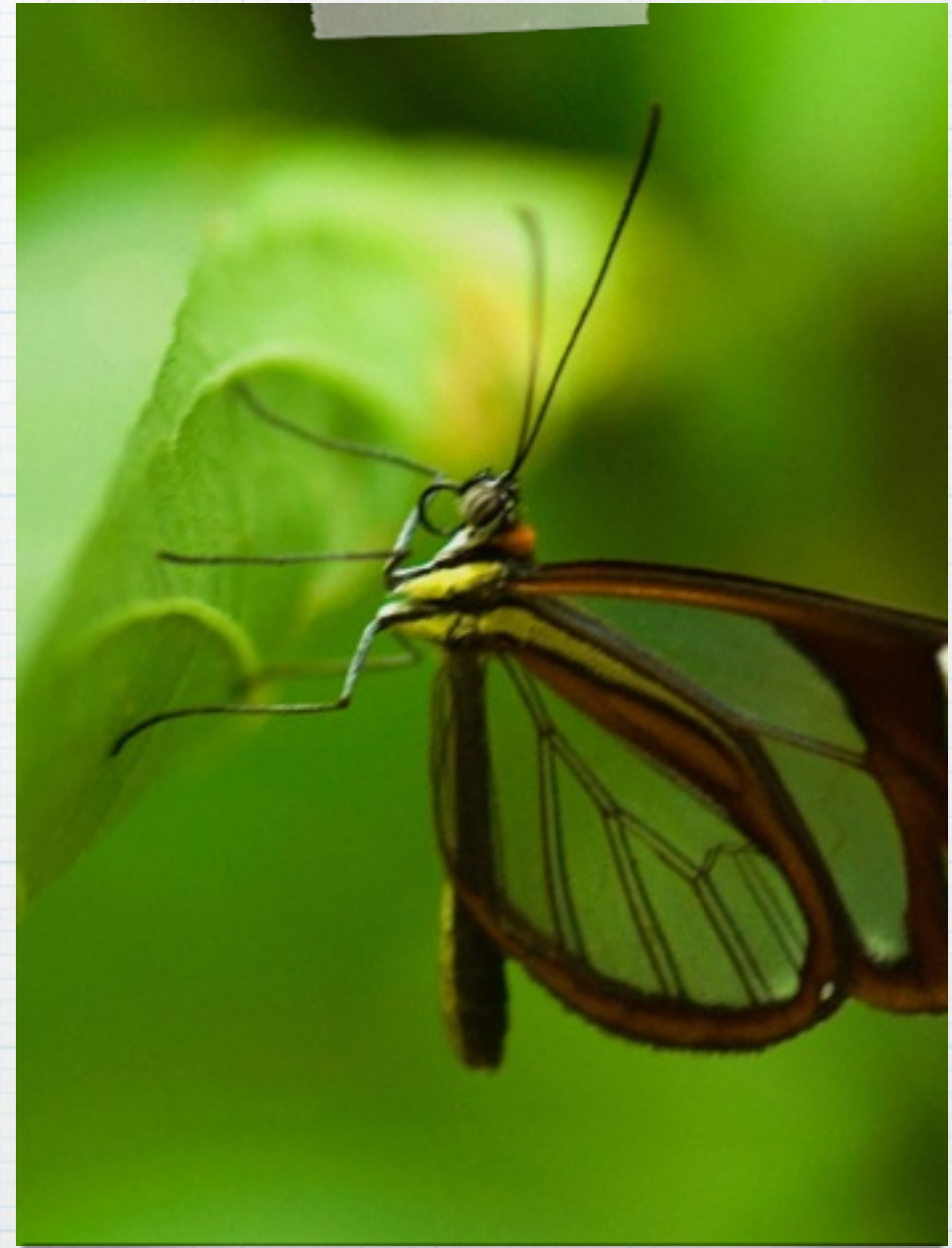
Our Best team members

* Xi LIU

* Lei XIAO

* Yifu WANG

* Yan HE

* Xiaoxu LU

Vision

* Answer YES/NO questions

* In Biomedical domain

* With High precision

* Example: Are there any DNMT3 proteins present in plants?

Significance

* BioQA system could expedite biomedical search by queries, answering the given question with an exact answer instead of returning a list of relevant documents.

* Evidenced-Based retrieval and Natural Language Processing enables high precision of the system.

Milestones

Milestone0: pipeline built

Milestone1: document, concept, triple retrieval

Milestone2: snippet retrieval and evaluation

Milestone3: answer extraction and selection
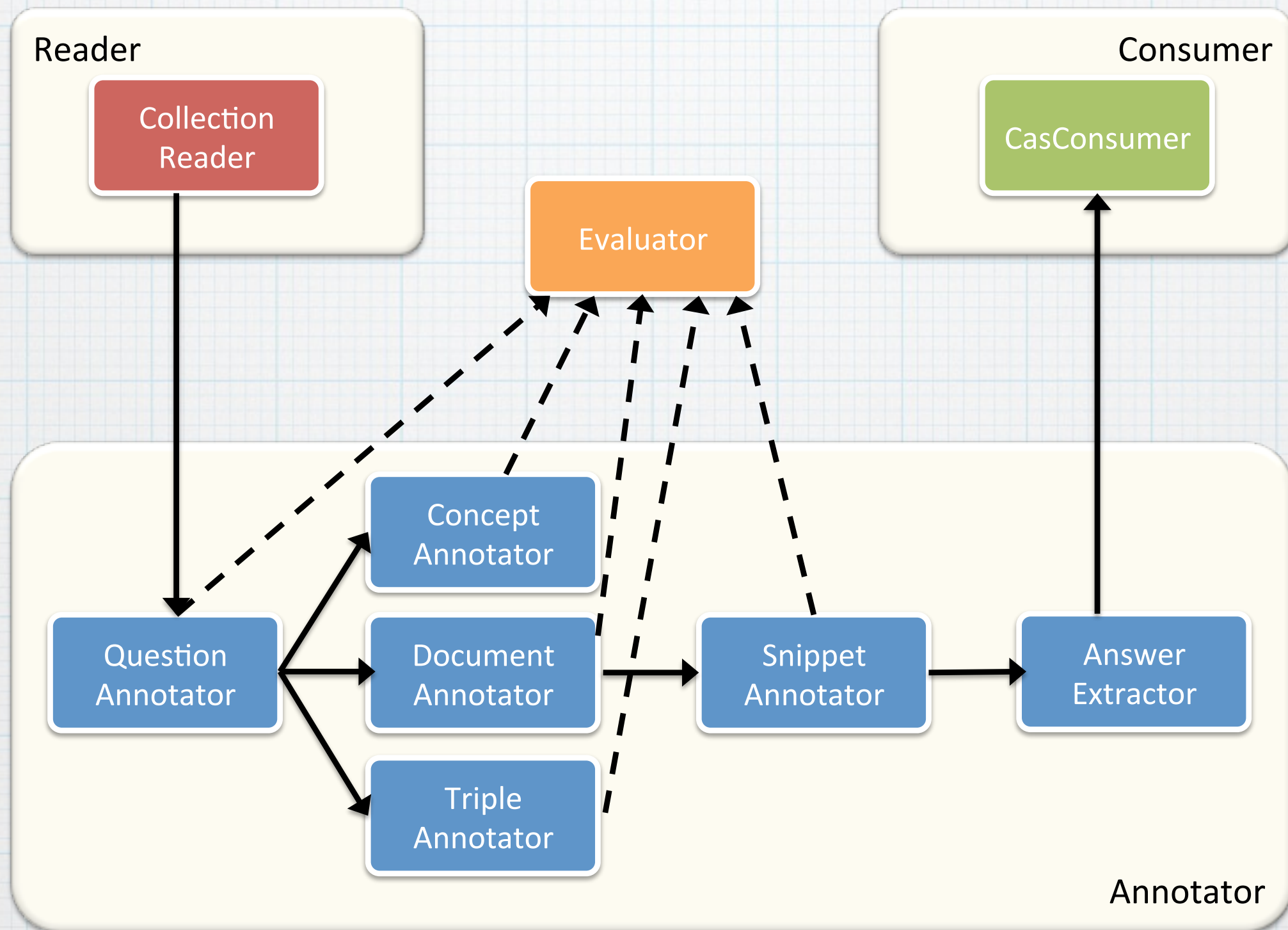
Milestones & Timelines

Milestone0: pipeline built Oct.24–27

Milestone1: document, concept, triple retrieval Oct.28–Nov.10

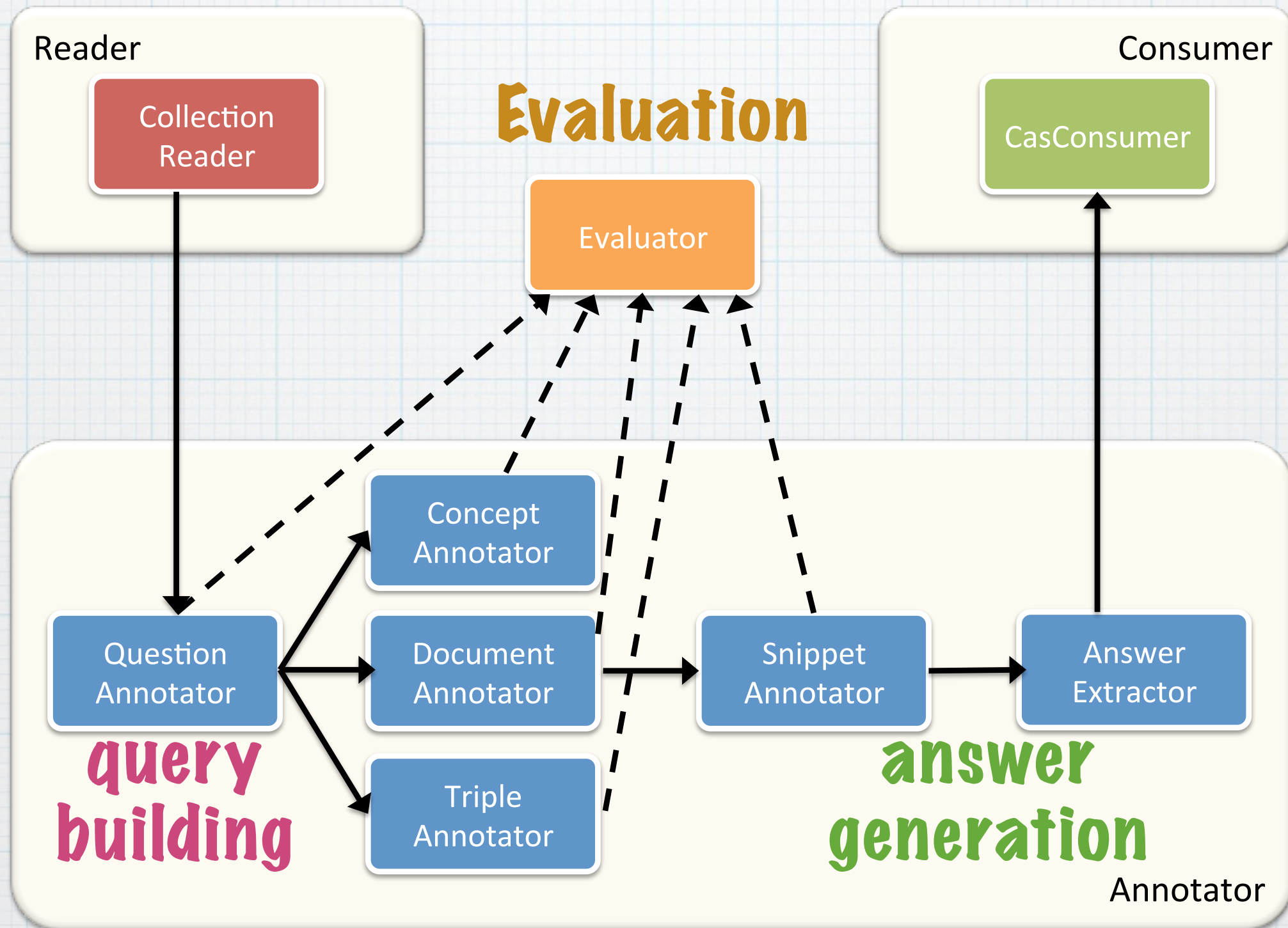Milestone2: snippet retrieval and evaluation Nov.11–Nov.20

Milestone3: answer extraction and selection Nov.21–Dec.01

Carnegie Mellon
SCHOOL OF COMPUTER SCIENCE

Pipeline Design

Query Building

Are there any DNMT3 proteins present in plants?

* **Stanford NLP Lemmatizer**

be there any dnmt3 protein present in plants?

* **OPEN NLP Tokenizer**

[be, there, any, dnmt3, protein, present, in, plants, ?]

* **stop word removal**

dnmt3 protein present in plants

* **Lingpipe Name Entity Recognizer**

dnmt3 protein

* **Query operator**

dnmt3 AND protein AND present AND in AND plants

dnmt3 protein[mesh] present in plants

Carnegie Mellon
SCHOOL OF COMPUTER SCIENCE

Answer Extraction

* ## Cosine similarity (Snippet Retrieval)

Compare question with each sentence in retrieved document.

Higher cosine similarity sentences -> snippet.

* ## Position Scoring (Sentence scoring)

Sentence position in a paragraph indicates likelihood for summarizing key points. First and last sentence -> high; middle part -> low.

* ## AFINN-Sentiment Scoring (Word scoring)

Tf-idf for word importance and sentiment dictionary for positive or negative level.

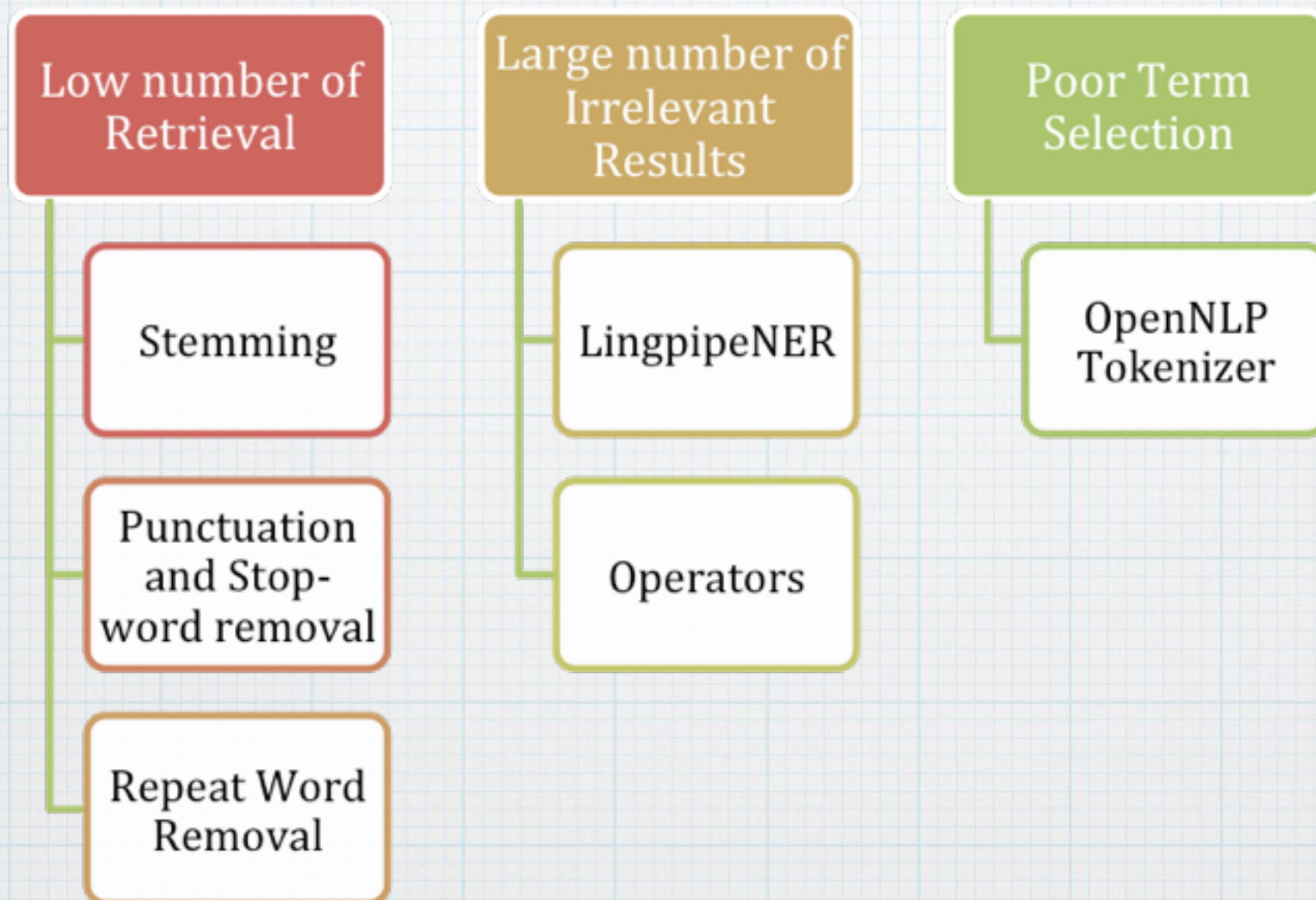* ## Voting system (YES/NO)

Each sentence gets one chance to vote for Yes/No, majority rules applied.

Experiment

## A. Improve Retrieval Precision by Building Complex Queries

Low number of Retrieval
- Stemming
- Punctuation and Stop-word removal
- Repeat Word Removal

Large number of Irrelevant Results
- LingpipeNER
- Operators

Poor Term Selection
- OpenNLP Tokenizer

Experiment

B. Trade-off Retrieval Precision with Snippet Generation

Document Retrieval <—> Snippet <—> Answer Candidate Pool

Doc Retrieval Precision

Snippet Candidate Pool
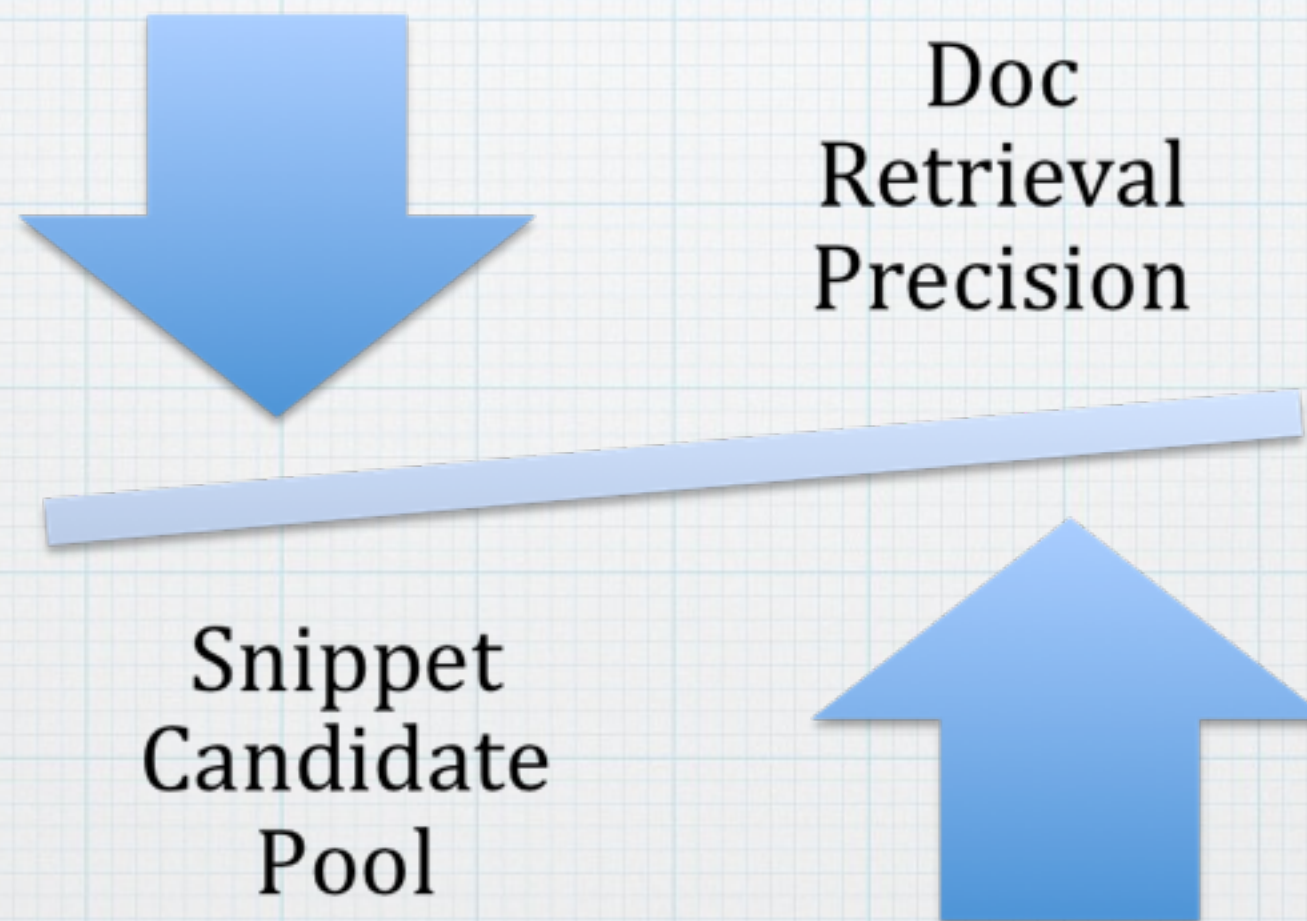
Experiment

C. Select Best Answer Selection Strategy

Sentence Similarity, Sentiment Scoring

Voting Strategy

Position Weighting
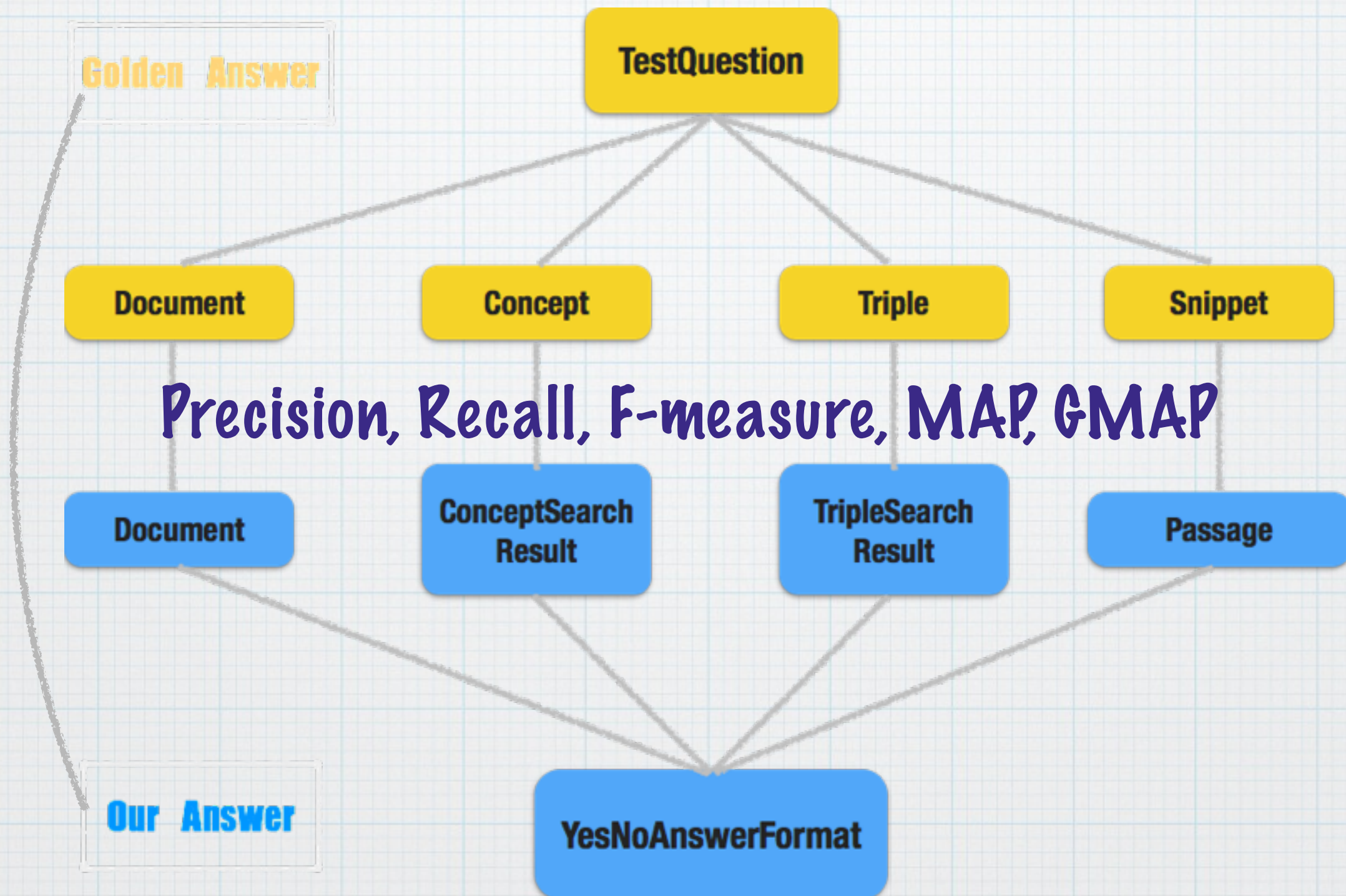
Performance

```
+---------------------------------------
|        Performance Report      |
+--------------------------------------+
| Document
| Precision: 0.1032
| Recall: 0.0723
| F-Score: 0.0842
| MAP: 0.2375
| GMAP: 0.7093
+--------------------------------------+
| Concept
| Precision: 0.0079
| Recall: 0.0024
| F-Score: 0.0063
| MAP: 0.0074
| GMAP: 0.023
+--------------------------------------+
| Triple
| Precision: 0.0000
| Recall: 0.0000
| F-Score: 0.0000
| MAP: 0.0000
| GMAP: 0.0000
+--------------------------------------+
| Snippet
| Precision: 0.0017
| Recall: 0.0012
| F-Score: 0.0015
| MAP: 0.0021
| GMAP: 0.0307
+--------------------------------------+
| ExactAnswer
| accuracy: 0.625 right number : 5 total number:8
+--------------------------------------+
```

Error Analysis

* Query ??

* Document ??

* Snippet ??

Future Work

* Sentiment dictionary enhancement

* Query expansion and integration with concept and triple

* List/Factoid/Summary questions

Q&A

Thanks!!