Team 01 -- *The One!* Milestone 3

11-791 Group Project

Team Members

- Jeremy Doornbos
- Anurag Kumar
- Zexi Mao
- Fei Xia

Vision

- Ultimate Goal:
  - Build a system for biomedical question answering, specifically for yes/no question!

- Milestone 3:
  - Improve performance of concept and document retrieval
  - Improve performance of snippet extraction
  - Exact answer generation and evaluation

What's New

- Better query building
- Multiple Sources -- Concept Retrieval
- Better snippet extraction
- Exact answer generation
- Answer evaluation

Pipeline
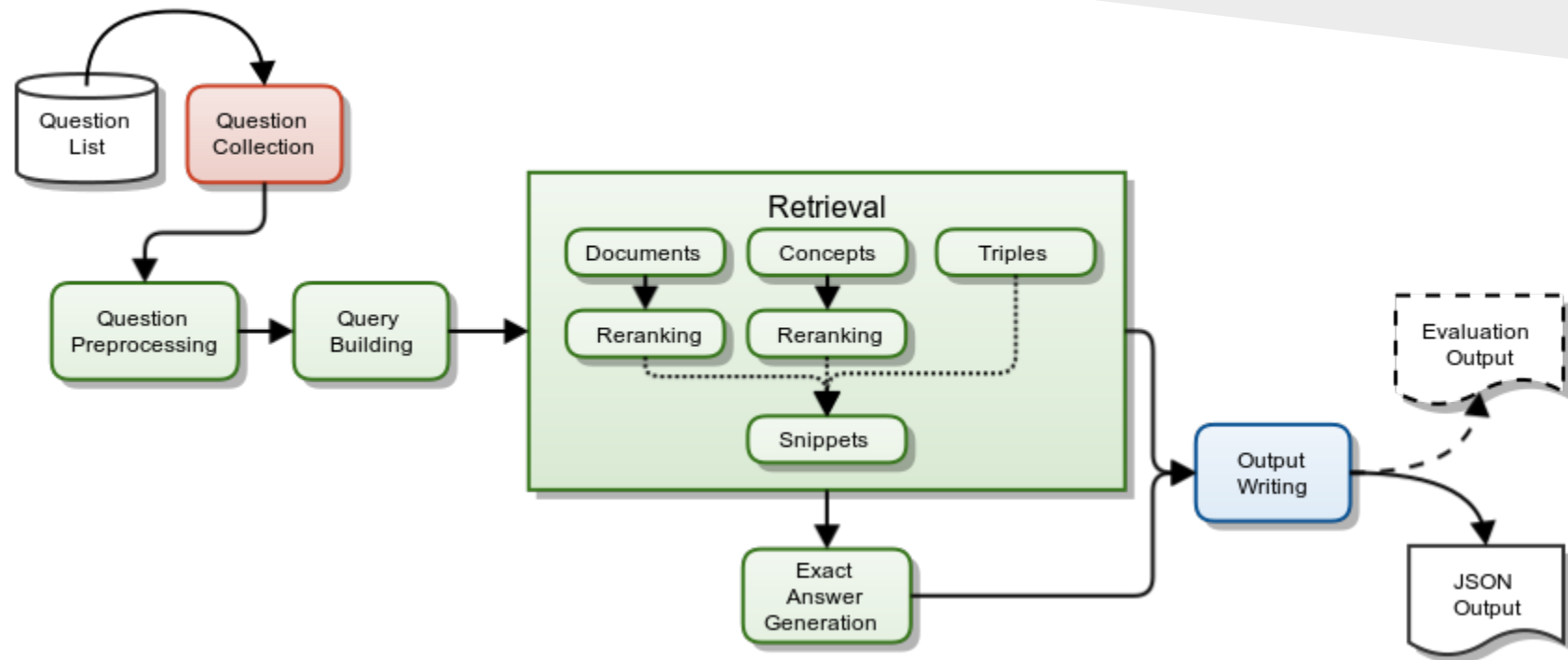
Query Building

- Complex query of form (a AND b) OR (c AND d)
  - Too many results retrieved
- For yes/no question specifically, add root of the dependency parse tree to query
  - e.g. "relate", "approve"
- Classify entity type (e.g. protein, gene) for concept retrieval source selection
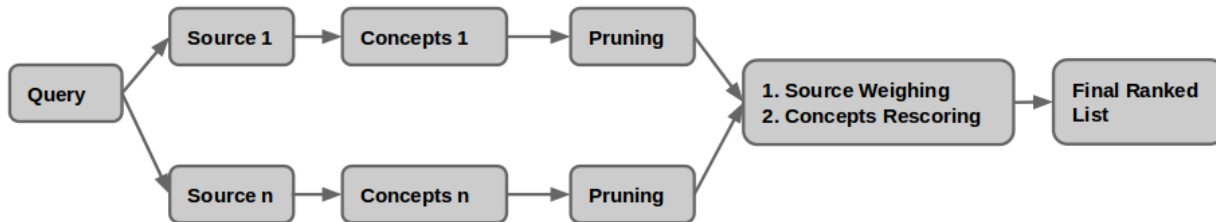
Query Building

Results for Document Retrieval:

Query Formulation	Precision	Recall	F-measure	MAP	GMAP
AND operator	**0.214**	**0.240**	**0.188**	**0.278**	**0.017**
(AND) OR (AND)	0.110	0.139	0.115	0.139	0.005
AND with sentence root	0.115	0.161	0.100	0.163	0.008

Concept Retrieval

- Single Source vs. Multiple Sources
- Need Multiple Sources
- Different Sources in different contexts
- Two step process
  - Get relevant things from all source
  - Rerank
    - Confidence on Source
    - Based on Query

Query → Source 1 → Concepts 1 → Pruning

Query → Source n → Concepts n → Pruning

→ 1. Source Weighing 2. Concepts Rescoring → Final Ranked List

Concept Retrieval

Effect of pruning when using multiple sources

Concept Perf.	Precision	Recall	F-Measure
Mesh Only	0.213	0.269	0.208
All-No pruning	0.096	0.439	0.129
All - Pruning	**0.142**	**0.405**	**0.178**

Available test set heavily dependent on MeSH only

Concept Retrieval

Effect of re-ranking strategies

Concept Performance	MAP	GMAP
Mesh Only	0.479	0.060
All - No pruning	0.344	0.106
All - Pruning	**0.435**	**0.119**
All - Prun. + Source Confidence (SW)	**0.445**	**0.130**
All - Prun. + Query Based(QS)	**0.426**	**0.142**
All-Prun.+QS+SW	0.435	0.120

More tuning of parameter in alpha might have shown improvements.
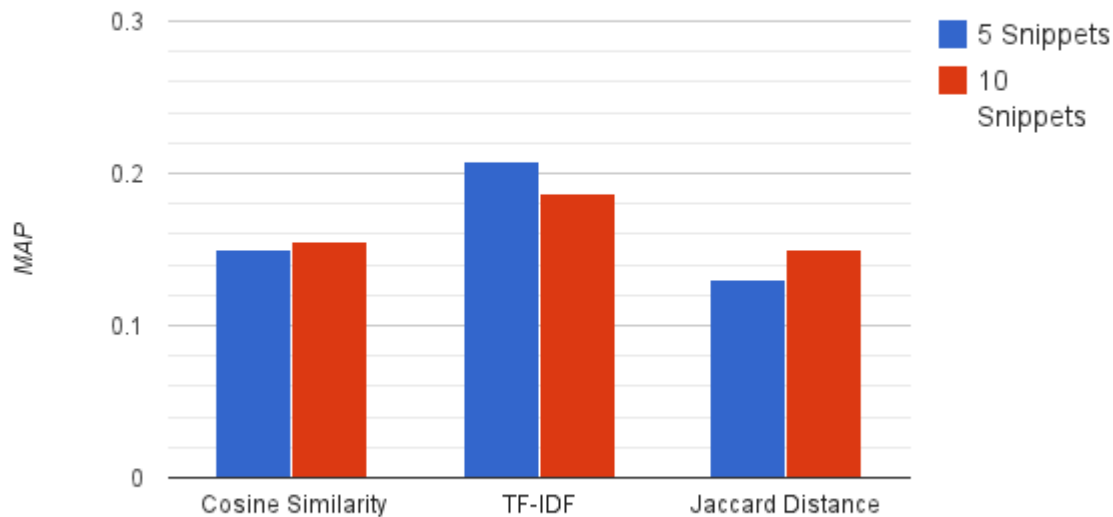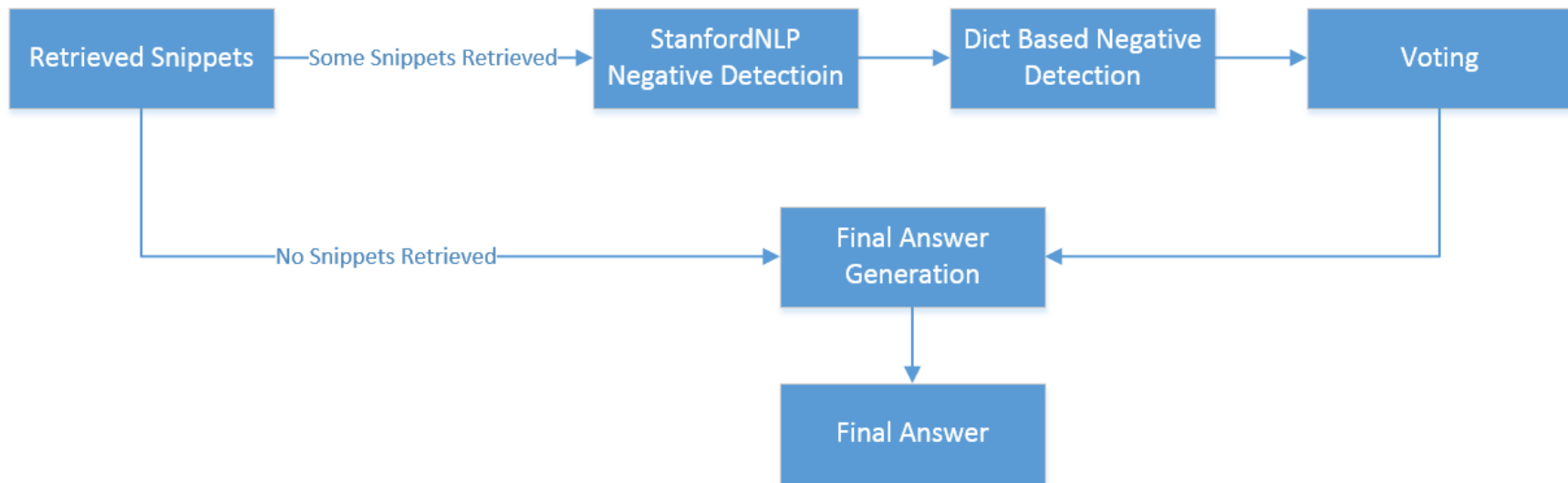
Document Retrieval
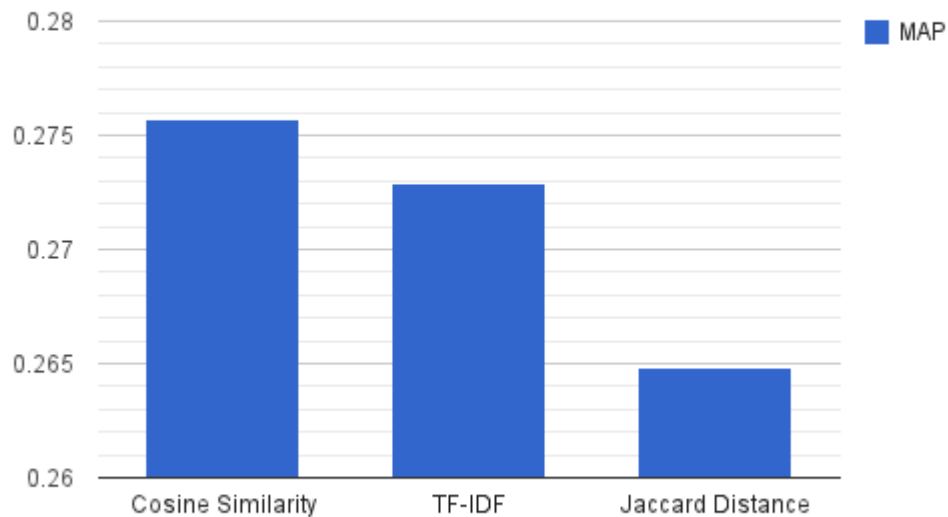
Snippet Extraction
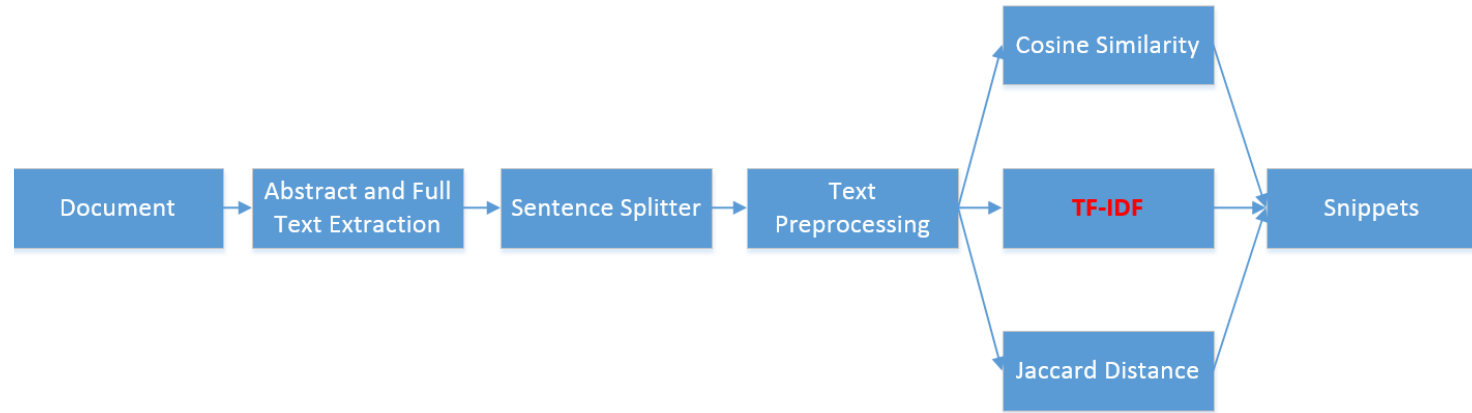
Snippet Extraction

Exact Answer Generation

Exact Answer Generation

Question Type: Yes/No

Number of questions: 8 (6 yes, 2 no)

Baseline (always Yes) Accuracy: 75%

Our Accuracy: 100%

??

- How did we achieve these results?

- Are we just awesome, the chosen 1?

Exact Answer Generation

Affirmative Examples

Question: "Is Mammaprint approved by the United States Food and Drug Administration?"

Snippet (TP): "The MammaPrint assay has the advantages of a 510(k) clearance by the US Food and Drug Administration"

Answer: Yes

Question: "Is thrombophilia related to increased risk of miscarriage?"

Gold Snippet: "Thrombotic risk during pregnancy and the puerperium is higher in asymptomatic women with than without thrombophilia"

Snippet (FP): "OBJECTIVES: Recently, it has been found that women who have thrombophilia have increased risk of fetal loss."

Answer: Yes

Exact Answer Generation

Negative Example

Question: "Are there any desmins present in plants?"

Snippets: None retrieved

Gold Snippets:

- "Characterization of a zebrafish (Danio rerio) desmin cDNA: an early molecular marker of myogenesis."
- "Mutations in desmin have been associated with a subset of human myopathies."
- "Desmin is a muscle-specific protein…"

Answer: No

Exact Answer Generation

However…

Question: "Is depression associated with poor prognosis of brain tumor patients?"

Snippets:

- "Three patients with temporal depressions were operated on by this method with satisfactory results."
- "BACKGROUND: Fatigue, cognitive deficits, and depression are frequently reported but often undertreated symptoms that can profoundly affect daily life in patients with primary brain tumors (PBTs).
- "While depression and delirium are frequent in cancer patients, hypomania during chemotherapy is rare."

Answer: Yes ??

Team Management

- Multiple long work sessions
  - Includes discussions and actual work
- Also worked separately
- Work
  - Jeremy: Evaluation
  - Anurag: Concept Retrieval, Reranking
  - Zexi: Query Processing
  - Fei: Snippet Retrieval and Reranking, Exact Answer Generation

Reflections

- Work
  - Group work sessions helped in resolving issues
- Problems
  - Maven release - biggest issue (spent hours)
  - Dataset too small to get comprehensive analysis
- Potential Future
  - Answer generation could be more robust
  - Incorporating retrieved concepts to query again