Team 5:

Fellowship of the Larman

Maya Tydykov, Chao-Hung Chen, Niloy Gupta, Hakim Sidahmed

Pipeline

List Question → Collection Reader

Expanded Question → Question Expander

Document Passage Extractor → Concept Annotator, Document Annotator, Snippet Annotator

Candidate Answer → Candidate Answer Annotator

Answer Evidence Retrieval → Re-Run Annotators

Answer → CAS Consumer
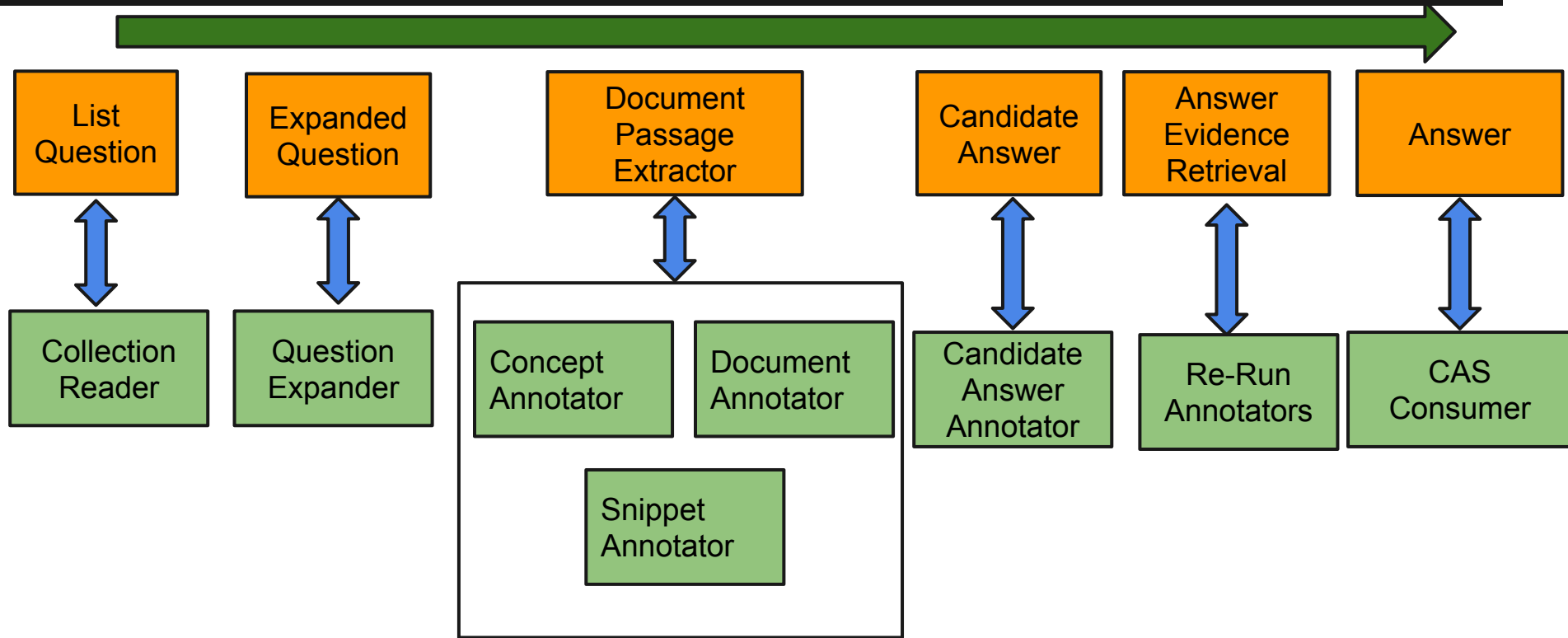
Performance trend

First iteration

	MAP	GMAP
Concept	0.017241	0.013243
Document	0.030855	0.027856
Triples	0.000000 *(issues with gold standard and web service)	0.000000

Performance trend

Second iteration

	MAP	GMAP
Concept	0.017241	0.013243
Document	0.030855	0.027856
Triples	0.000000 *(issues with gold standard and web service)	0.000000

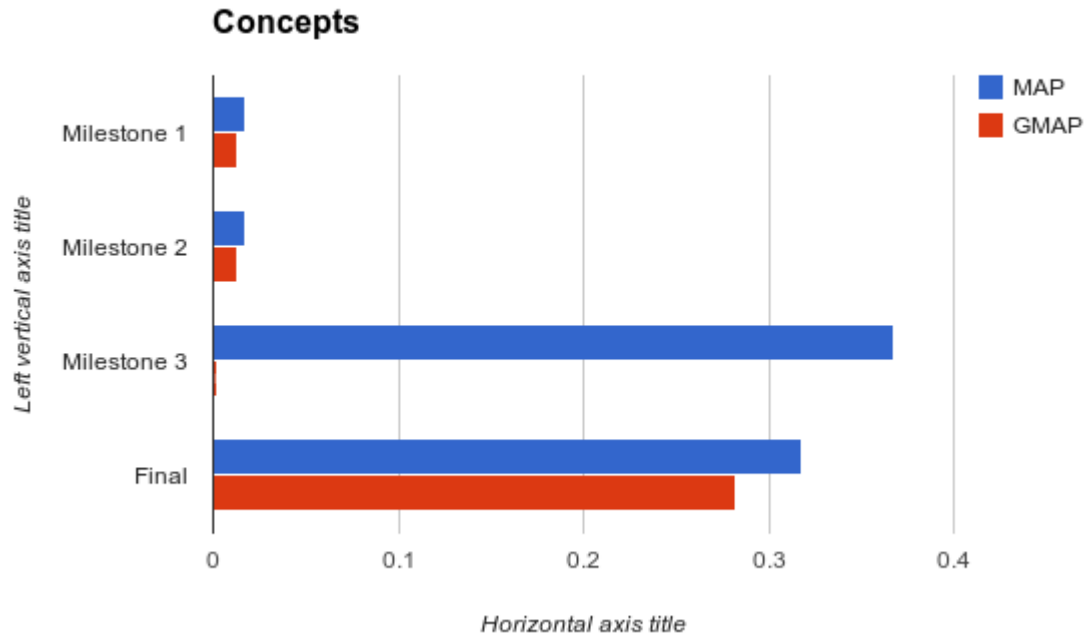# Performance trend

Third iteration

	MAP	GMAP
Concept	0.367308	0.001340
Document	0.125000	0
Snippets	0.015993	0

Performance trend

Final results

	MAP	GMAP
Concept	0.318056	0.282326
Document	0.067085	0.000003
Snippets	0.003205	0
Exact Answers	0.003503	0.00000

Performance trend

Performance trend

Documents

Legend: MAP (blue), GMAP (red)

Vertical axis (Left vertical axis title): Milestone 1, Milestone 2, Milestone 3, Final

Horizontal axis title: 0, 0.04, 0.08, 0.12, 0.16
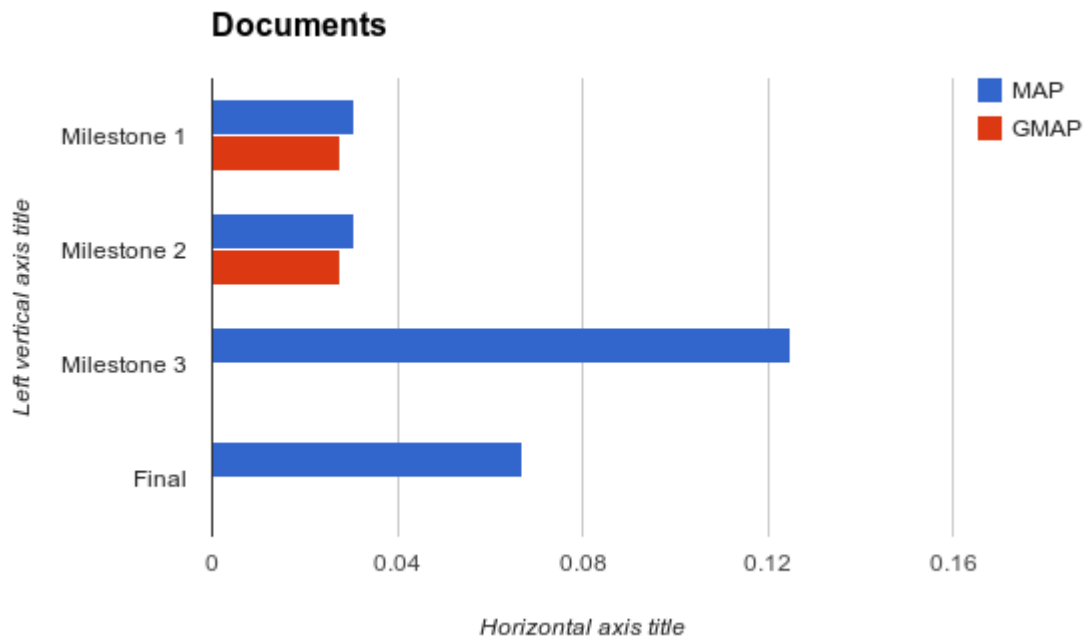
Error Analysis

A Journey down the pipeline

Taming PubMed

- Poor precision and recall for documents
- PubMed is sensitive to the query
- Small variations in query return different results
- Included concept terms, UMLS synonyms
- Binary Operators and MeSH

Taming PubMed

Q1: $(t_1$ or $<s_n>) + (t_2$ or $<s_n>) + (c_1$ or c_2) => D $\{d_1, d_2, d_3, d_4, \ldots d_n\}$

Q2: $(t_1$ or $<s_n>) + (c_1$ or c_2) => D $\{d_1, d_2, d_3, d_4, \ldots d_n\}$

Q3: $(t_1$ or $<s_n>) + (t_2$ or $<s_n>$) => D $\{d_1, d_2, d_3, d_4, \ldots d_n\}$

Q4: $(t_2$ or $<s_n>) + (c_1$ or c_2) => D $\{d_1, d_2, d_3, d_4, \ldots d_n\}$

Q5: $(t_2$ or $<s_n>$) => D $\{d_1, d_2, d_3, d_4, \ldots d_n\}$

(* if t_2 is a biological term)

Output $D_k = \{d_i\}$

if d_i has been returned more than K times.

Where K is the threshold set to 4.

Limit number of runs by thresholding size of output D

Needle in the Haystack

- Passage Term Matching
- Skip-Bigram
- Windowing

Needle in the Haystack

- Window Size: 20 words
- Window Increment: 6 words
- Window that has at least one biological term is analysed
- Ranking: Number of bigram matches (question term + synonyms)

What's in a name (d Entity)?

- Candidate Answer is a biological term
- How do you find if the token a biological term?
- AbNER : Low true positive
- Lingpipe: Only trained on genes. No training data for biological terms.
- UMLS and PubMed Ontologies: Many false positives

An Educated Guess

- Extract nouns from stemmed snippets
- Record occurrence frequency
- Hypothesis: Answers should not have very low frequencies
- Ideal: Answers should occur around the mean if snippet recall is good.

Where is the answer?

Definitely not here!

Frequency
of occurrence in snippets
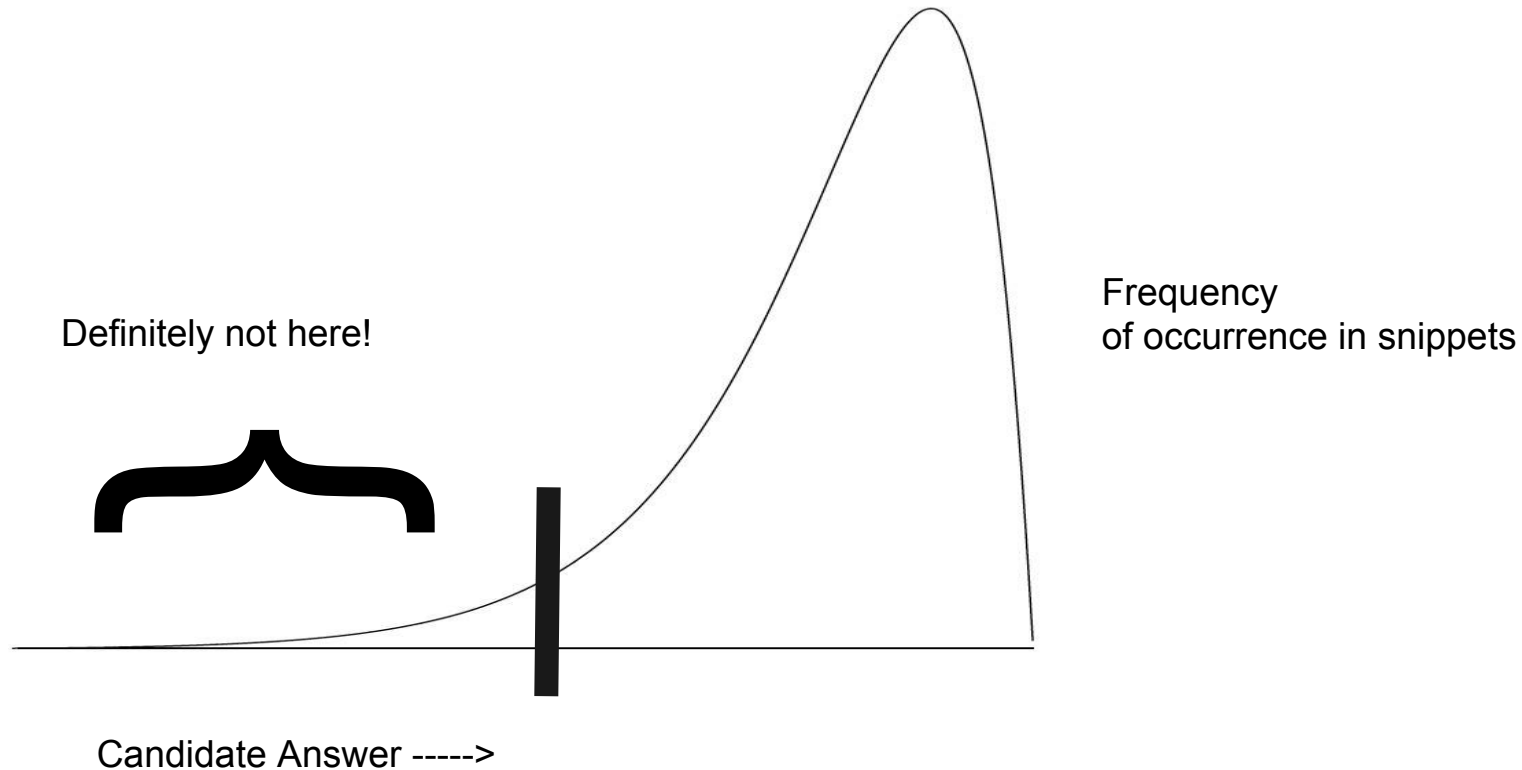
Candidate Answer ----->

Supporting Evidence Retrieval

- Combine the candidate answer with the search terms.
- Hypothesis: Combining candidate answer with the question will give correct documents and snippets. Reinforce answer ranking

But did it work?

- Works only if the candidate answer is a biological term.
- Due to ineffective Biological Term extractor SER did not improve results

What next?

- Train a better Bio Named Entity Recognizer.
- Textual Alignment, Logical Form Answer Candidate Scorer, Semantics
- Asynchronous processing?
- Submit recommendations to PubMed team

Bibliography

- Biological Question Answering: A Survey
- Textual evidence gathering and analysis
- Question analysis: How Watson reads a clue

Breakdown of Work

Niloy	Hakim	Chao	Maya
- Brute Force Query - Bio Term Extractor- AbNER, Lingpipe - SER - Pipeline bug fixes - Performance tuning	- Collection Reader - Cache - Document Annotation - Document Error Analysis - Global code Review	- Snippet Annotators with both sentence based and window based - Answer Annotators - Error analysis for snippet annotators	- Query Expansion - Evaluation - SER - Biological term filtering - Exact answer thresholding

Questions for us?

^The pipeline can't answer your questions