



STAR WARS  
MILESTONE III  
REVENGE OF THE LISTS

# Group 100 11-791 Project

*Milestone 3:*  
***REVENGE OF THE LISTS***

**Joseph Chang, Nikolas Wolfe,  
Di Xu, Prajwal Yadapadithaya**

# Project Progress

- Decided on Lists for Q/A
- Improved Document, Concept and Triples retrieval. (numbers coming soon!)
- More caching!
- BioASQ pipeline completion.
- Testing various strategies to improve answer retrieval.
- Deployment. (Yet to be done)

# Improved Results

Query Type	MAP	GMAP ( $\epsilon=0.01$ )
Documents*	0.1822 (+44.03%)	0.033 (+43.47%)
Wipe out all stoppers; use unigram model with window size 2 to scan the query. Add “AND” between two terms where both of the have a probability $< 1e-4$ .		
Concepts	0.4056 (+106.8%)	0.2698 (+371.6%)
The “Year” hack. (Well, not really a hack)		
Triples	0.0119 (From 0.00)	0.0134 (+34%)
All the retrieved triples from the webservice API were incomplete. Hence partial evaluation of Triples were considered.		



# Time For Error Analysis!

Because....



# Error Analysis - Concepts

## Concepts:

- Gold standard data is wrong (year 2012)
- Query Normalization
  - Punctuation, Spacing, Casing, Lemmatization
  - Difficult to evaluate when you can't get anything back from the WebAPI
- Pseudo-Relevance Feedback
  - Take the results that are initially returned from a given query and to use information about whether or not those results are relevant to perform a new query
  - Recursively iterate query with PRF
  - Difficult to evaluate when you can't get anything back from the WebAPI

# Error Analysis - Concepts

- Pseudo-Relevance Feedback

Example (Initial Query):

*Is Rheumatoid Arthritis more common in men or women?*

Using Top-10 ranked Concept results, stemming, stop-word removal, lower-casing, and word frequencies we obtain:

*chain woman medium rheumatoid acid arthritis fatty syndrome synthase  
thioesterase serpin acyl juvenile man felty*

Iterating again in this way we get, oh wait, nothing... **Web API fail.**

# Error Analysis

## Pseudo-Relevance Feedback Example

Trial	Iterations	Avg. Disease Ontology Concepts	MAP
Baseline	0	40.0	0.0725
PSR 1	1	66.3	0.027
PSR 2	2	75.2	0.019
PSR 3	3	53.9	0.018



# Error Analysis

## Discussion

- In principle, would be good to try some form of PSR to improve results in other areas
- Experiment improving Concepts baseline (without Uniprot / Joachem results) shows
  - PRF can destroy results
  - More results increases recall
  - Narrow results increases precision

# Error Analysis - Snippets

Similarity Metric	F1-Score*
Cosine	0.03675
Jaccard	0.03387
Dice	0.03387
Average	0.03387

\*excluding cases where the gold standard does not have a solution

# List Q&A Results

## Algorithm:

1. Identify **pivot terms** in the query, using **unigram heuristics**
2. Locate pivot terms in abstract, extract text in a **window** of fixed size.
3. Run **BioNER** on the extracted text.
4. Limit to **unigram entities** for the moment.
5. Wipeout recognized entities with unigram heuristics.

	Precision	Recall	F1
Exact Match	0.0053	0.1875	0.0101
Token Match	0.0699	0.25	0.0631

# List Q&A Results

**Question:** Which acetylcholinesterase inhibitors are used for treatment of myasthenia gravis?

**pivot terms:** acetylcholinesterase, inhibitors, myasthenia, gravis

## **Our Ranked Answer:**

MuSK 3538944.000000

AChR 82944.000000

Myasthenia 55296.000000

cholinesterase 41472.000000

**pyridostigmine** 1024.000000

thymectomy 432.000000

Alzheimer 384.000000....

edrophonium 128.000000

**neostigmine** 64.000000

ColQ 48.000000

## **Gold Standard:**

**neostigmine**

**pyridostigmine**

# List Q&A Results

- **Advantage:**

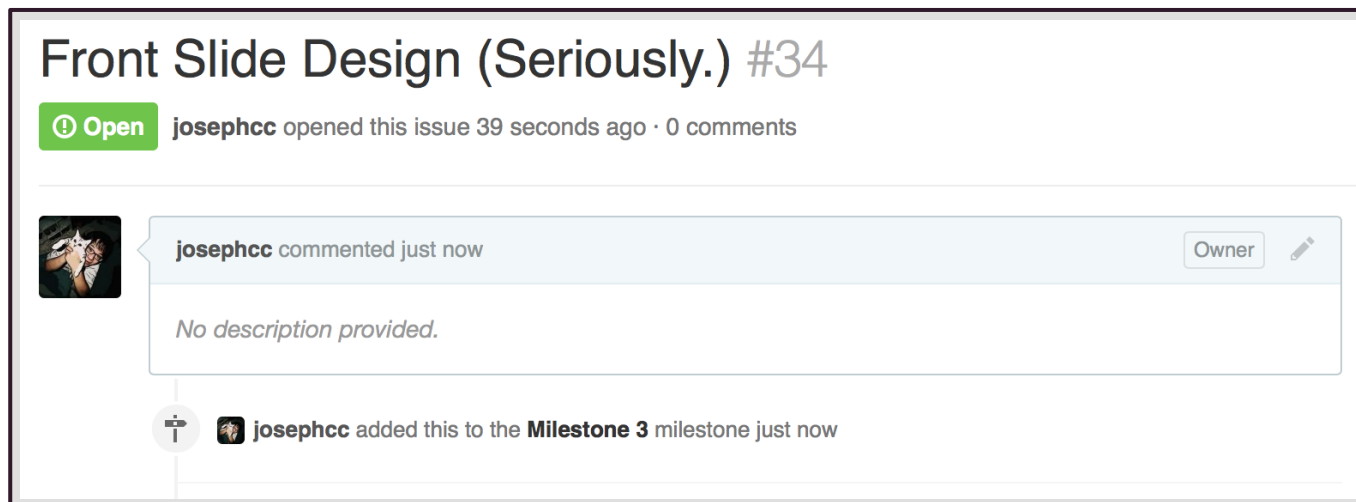
- This method is robust and its assumption holds.
  - smaller window size -> higher P, lower R
  - shorter list -> higher P, lower R
- Robust to overfitting
- Efficient

- **To be improved:**

- Bigram LM for bigram List entities.
- Need NLU module to understand what the question is looking for.
- Need ontology corpus for concept matching

# Logistics - More issues on Github!

- Front Slide Design (Seriously)
  - Maximize win
  - Minimize subtlety





# Logistics

- **MOARRR HACKATHONS!!!**

- 3 hackathons last week (7-8 hours each)
- Managed to work distantly.
  - Slack, emails, documentations, modular design
  - Robust framework
- Reworking various strategies. (Mentioned earlier)
- Pair Programming.
- Had fun!



- **DIVISION OF LABOR**

- Document reworking - Di and Prajwal
- Updating triple evaluation - Prajwal
- Concepts reworking and Cache enhancements - Nik
- List answers retrieval - Di
- Snippets, evaluations, reader, type system - Joseph

# Team 100 FTW.

## Questions?

*Stand on the shoulders of giants, Wednesday teams!*  
*Fork our code! We dare you...*

 [11791-04 / project-team04](#)  
 [michaelzhhan1990 / project-team04](#)

# Okay?





BTW

STAR WARS  
MILESTONE III  
REVENGE OF THE LISTS