

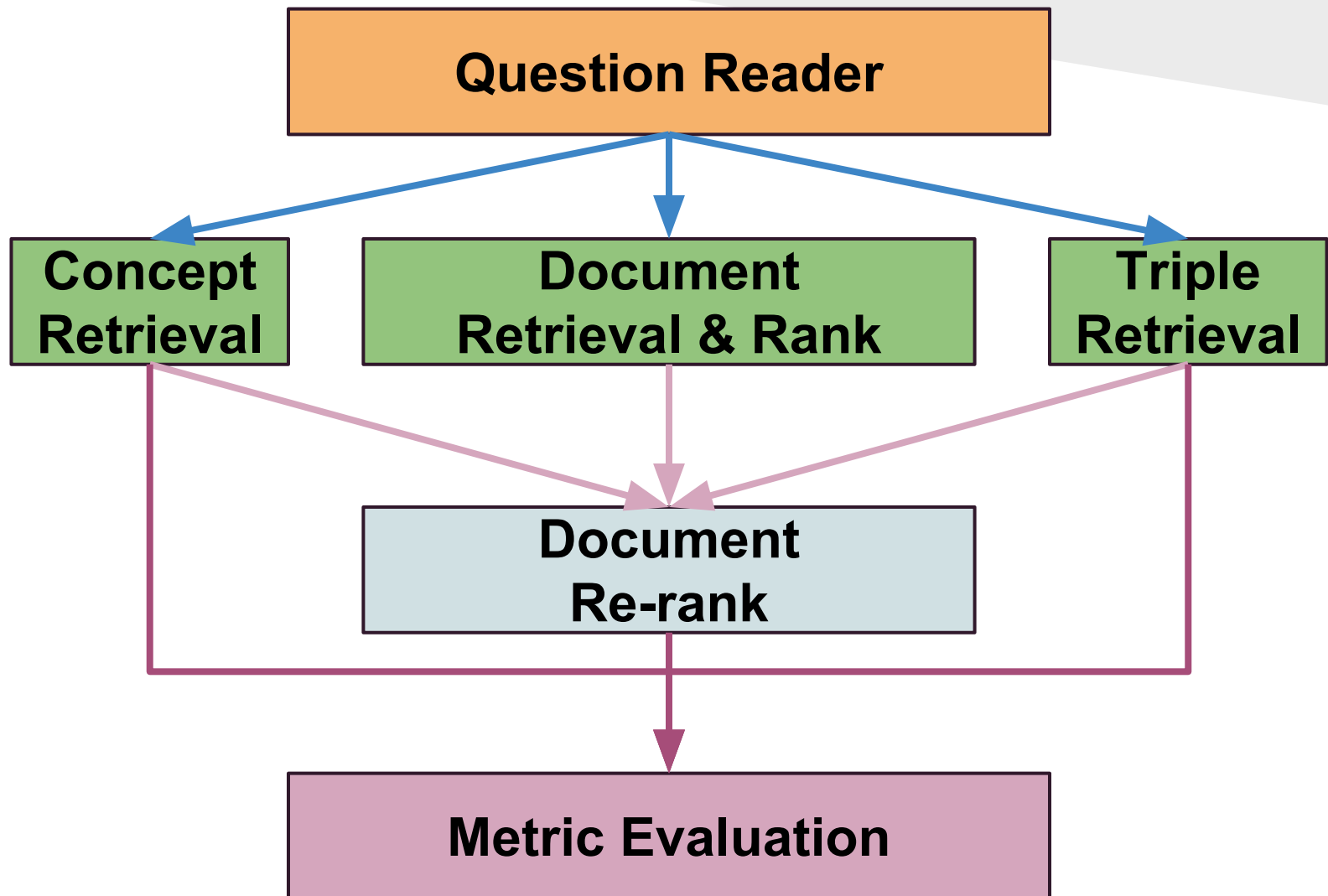
Group 100

11-791 Project

Milestone 1: Mostly Head Scratching

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

Architecture



Division of Tasks

- **Di**
 - Document Retrieval and ranker
 - MAP / GMAP Metric
- **Nik**
 - Concept Retrieval
 - API Service Helper
- **Prajwal**
 - Triples Retrieval
- **Joseph**
 - Collection Reader
 - CAS Consumer / Metric integration

Project Progress

- **API Calls**

- GoPubMed **Singleton**
- API accessed through **Proxy**
- Caching for efficiency?

- **Concept Retrieval:**

- Concept Analysis Engine
- API calls → Sort & Rank Results → Build ConceptSeachResult(s) → Add to Indexes
- Sorted Results → Concept MAP/GMAP

Project Progress

- **Document Retrieval:**

- Retrieve document information from PubMed
- Krovetz Stemmer
- Stopper Removal
- Rank document by their relevance to the query
- Okapi, Dirichlet, Indri, XQL (Zhai, et al. 2012)
 - *PubMed articles are different from TREC corpuses*
 - *Insufficient background knowledge*

- **Triples Retrieval:**

- Retrieve triples using the LinkedLifeDataService API
- Use the scores to rank the triples
- TripleSearchResult type to rank triples.


Team Meetings


- We had **four** meetings for Milestone 1
 - Meeting 0 [11/3] - Meeting Team Members
 - Meeting 1 [11/5] - Basic Idea Discussion
 - Meeting 2 [11/7] - Discussing Division of Labour and Implementation Details
 - Meeting 3 [11/9] - Implementation and Consolidating of Individual Work


Logistics


- Issue tracking on **GitHub**


☐ 3 Open ✓ 8 Closed


☐  **findUniprotEntitiesPaged() function fails** wontfix
#10 opened 13 hours ago by nikwolfe7


☐  **Output to file**
#9 opened 19 hours ago by josephcc 🚧 Milestone 2


☐  **Ranker (consumer or an AE)**
#8 opened 19 hours ago by josephcc 🚧 Milestone 2


☐  3 Open ✓ 8 Closed


☐  **Evaluation Metric Helper Class**
#11 opened 12 hours ago by josephcc 🚧 Milestone 1


☐  **Retrieve Tripples**
#7 opened 19 hours ago by josephcc 🚧 Milestone 1


☐  **Retrieve Documents**
#6 opened 19 hours ago by josephcc 🚧 Milestone 1

☐  **Retrive Concepts**
#5 opened 19 hours ago by josephcc 🚧 Milestone 1

☐  **Evaluation (Consumer)**
#4 opened 19 hours ago by josephcc 🚧 Milestone 1

☐  **Basic Consumer**
#3 opened 19 hours ago by josephcc 🚧 Milestone 1

☐  **Collection Reader (Annotations)**
#2 opened 3 days ago by josephcc 🚧 Milestone 1

☐  **Collection Reader (question body)**
#1 opened 3 days ago by josephcc 🚧 Milestone 1

Logistics

● Online discussion on Slack

The screenshot shows a Slack interface for a workspace named 'project-team04'. The left sidebar contains a list of channels, with '# general' selected. Below the channels are sections for 'DIRECT MESSAGES' (listing slackbot, dix, nwolfe, prajwal_y) and 'PRIVATE GROUPS' (with a link to 'New private group...'). The main area displays the '#general' channel conversation. The chat history shows several automated messages from a bot, each starting with '[11791-04/project-team04] Issue created by josephcc' followed by a topic like '#7 Retrieve Tripples', '#8 Ranker (consumer or an AE)', and '#9 Output to file'. The current conversation starts with a message from 'josephcc' at 4:21 PM saying 'yep' and 'basically all github activities comments / push / merges'. This is followed by a message from 'nwolfe' at 4:24 PM saying 'cool...' and 'oh so you've added all the issues to finish before the presentation, i got it'. Then 'josephcc' at 4:25 PM says 'i didn't assign all of them to MS1' and 'not sure why its posting everything twice' (with a star icon). 'dix' at 4:47 PM says 'joined #general'. 'dix' at 4:50 PM says 'This so called Slack seems to be the product of a stat-up company..Not sure how far they can go be for getting swallowed by bigger companies....'. Finally, 'nwolfe' at 4:51 PM says 'well that'd be nice, right? make cool little tool and then google buys you for a billion dollars.... never have to work for nobody ever again'. The bottom of the screen shows the input area with a search icon and the text 'x|'. The user 'josephcc' is shown as 'online' in the bottom left corner.

project-team04 ▾

CHANNELS

- # general
- +2 more...

DIRECT MESSAGES

- slackbot
- dix
- nwolfe
- prajwal_y

PRIVATE GROUPS

New private group...

#general ▾

[11791-04/project-team04] Issue created by josephcc

#7 Retrieve Tripples

[11791-04/project-team04] Issue created by josephcc

#7 Retrieve Tripples

[11791-04/project-team04] Issue created by josephcc

#8 Ranker (consumer or an AE)

[11791-04/project-team04] Issue created by josephcc

#8 Ranker (consumer or an AE)

[11791-04/project-team04] Issue created by josephcc

#9 Output to file

[11791-04/project-team04] Issue created by josephcc

#9 Output to file

josephcc 4:21 PM

yep

basically all github activities

comments / push / merges

nwolfe 4:24 PM

cool...

oh so you've added all the issues to finish before the presentation, i got it

josephcc 4:25 PM

i didn't assign all of them to MS1

4:25 ★ not sure why its posting everything twice

bad robot

dix 4:47 PM

joined #general

dix 4:50 PM

This so called Slack seems to be the product of a stat-up company..Not sure how far they can go be for getting swallowed by bigger companies....

nwolfe 4:51 PM

well that'd be nice, right? make cool little tool and then google buys you for a billion dollars.... never have to work for nobody ever again

josephcc online

x|

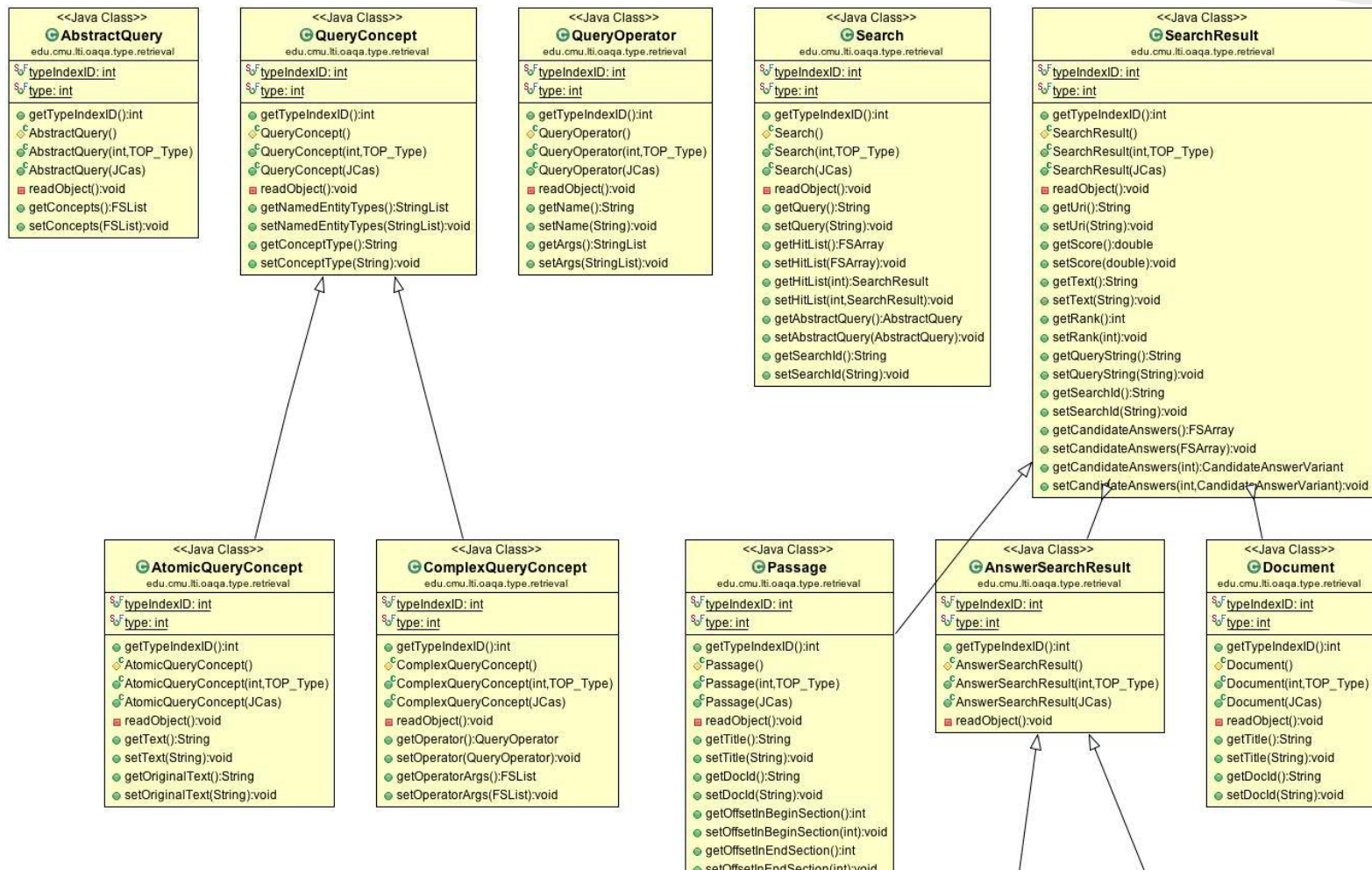
Results

Query Type	MAP	GMAP ($\epsilon=0.01$)
Document*	0.1038	0.0200
Concept	0.1961	0.0572
Triples	0.000	0.0100

* MAP for document retrieval has been improved from baseline (0.09)

Type System UML

* Available in Team04's repo





Thanks!

Questions?

We know we're awesome. .. You shall see... Eventually...