



# ARQMath

Answer Retrieval for Questions on **Math**

<https://www.cs.rit.edu/~dprl/ARQMath>



#ARQMath

**Richard Zanibbi**

**Rochester Institute of  
Technology, USA**

**rxzvcs@rit.edu**

**Douglas W. Oard**

**University of Maryland  
USA**

**oard@umd.edu**

**Anurag Agarwal**

**Rochester Institute of  
Technology, USA**

**axasma@rit.edu**

**Behrooz Mansouri**

**Rochester Institute of  
Technology, USA**

**bm3302@rit.edu**

# Goals

ARQMath aims to advance techniques for math-aware search, and semantic analysis of mathematical notation and texts

## Collection

**Math Stack Exchange** (MSE) is a widely-used community question answering forum containing over **1 million questions**

- Internet Archive provides free & public MSE snapshots
- **Collection:** Questions and answers from 2010-2018
- **Topics:** Questions from 2019

Formulas in appearance (**LaTeX, Presentation MathML**) and ‘semantic’ operation encodings (**Content MathML**)

# ARQMath Tasks

1. Finding answers to math questions
2. Formula search

**Note:** Task 2 queries are from Task 1 questions

# Task 1: Finding answers to math questions

Given a posted question as a query, search answer posts, and return relevant answers

Query	Search Results
<p>How can I evaluate <math>\sum_{n=0}^{\infty} (n + 1)x^n</math>?</p> <p>Asked 8 years, 5 months ago Active 4 months ago Viewed 34k times</p> <p>How can I evaluate</p> <p>384 <math>\sum_{n=1}^{\infty} \frac{2n}{3^{n+1}}</math></p> <p>I know the answer thanks to <a href="#">Wolfram Alpha</a>, but I'm more concerned with how I can derive that answer. It cites tests to prove that it is convergent, but my class has never learned these before so I feel that there must be a simpler method.</p> <p>146 In general, how can I evaluate</p> <p><math>\sum_{n=0}^{\infty} (n + 1)x^n</math></p> <p>sequences-and-series convergence power-series faq</p> <p>edited Sep 24 '17 at 12:09 Parcly Taxel 51.7k 13 80 120</p> <p>asked Apr 3 '11 at 21:41 Backus 2,072 3 12 8</p>	<p>1 No need to use Taylor series, this can be derived in a similar way to the formula for geometric series. Let's find a general formula for the following sum:</p> <p>...</p> $S_m = \sum_{n=1}^m nr^n.$ <p>2 It is equivalent to <math>x(x + 1)(x + 5)(x + 6) + 96 = 0</math></p> <p>Now</p> <p>...</p> $(x^2 + 6x)(x^2 + 6x + 5) + 96 = 0$ <p>3 If you want a solution that doesn't require derivatives or integrals, notice that</p> $1 + 2x + 3x^2 + 4x^3 + \dots = 1 + x + x^2 + x^3 + \dots$ $+ x + x^2 + x^3 + \dots$ $+ x^2 + x^3 + \dots$ <p>...</p> <p>▪</p> <p>▪</p> <p>▪</p>

# Task 2: Formula search

Given a formula in a question, search questions and answers, and return relevant formulas with their posts (**context**)

Query	Search Results
<p>How can I evaluate <math>\sum_{n=0}^{\infty} (n+1)x^n</math>?</p> <p>Asked 8 years, 5 months ago Active 4 months ago Viewed 34k times</p> <p>How can I evaluate</p> <p>384 <math>\sum_{n=1}^{\infty} \frac{2n}{3^{n+1}}</math></p> <p>I know the answer thanks to <a href="#">Wolfram Alpha</a>, but I'm more concerned with how I can derive that answer. It cites tests to prove that it is convergent, but my class has never learned these before so I feel that there must be a simpler method.</p> <p>146</p> <p>In general, how can I evaluate</p> <p><math>\sum_{n=0}^{\infty} (n+1)x^n</math>?</p> <p>sequences-and-series convergence power-series faq</p> <p>edited Sep 24 '17 at 12:09  Parcly Taxel 51.7k ● 13 ■ 80 ▲ 120</p> <p>asked Apr 3 '11 at 21:41  Backus 2,072 ● 3 ■ 12 ▲ 8</p>	<p>1 <math>\sum_{n=0}^{\infty} (n+1)x^n</math></p> <p>2 <math>\sum_{n=0}^{\infty} (n+1)x^n</math></p> <p>3 <math>\int_0^1 \frac{\ln(x+1)}{x^2+1} dx</math></p>

# Submitted Runs

Manual and Automatic

## Task 1

5 Teams

18 Runs

+5 Baselines

## Task 2

4 Teams

11 Runs

+1 Baseline

## Total:

6 Teams

29 Team runs

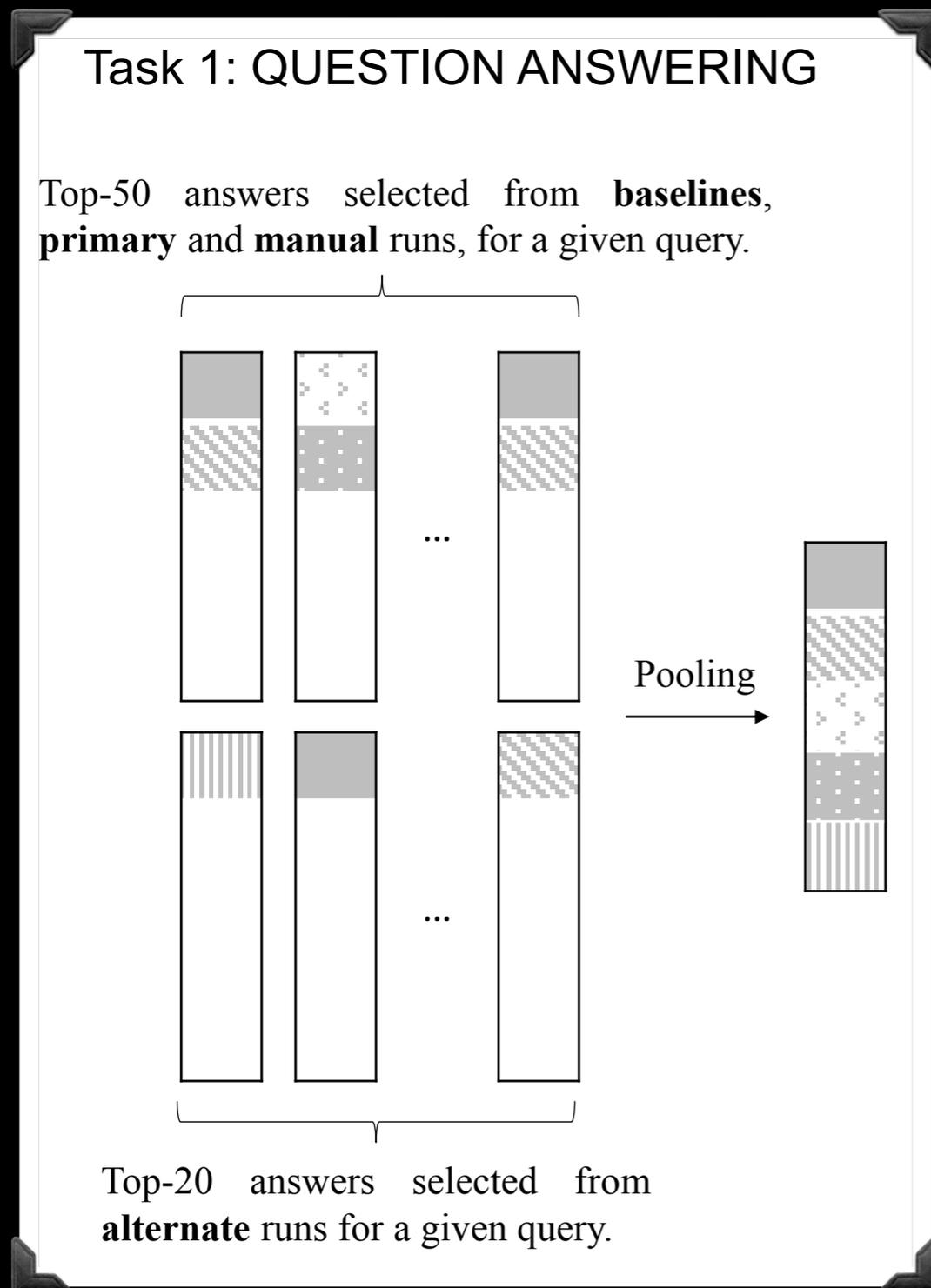
35 Total runs

	Automatic Runs		Manual Runs	
	Primary	Alternate	Primary	Alternate
TASK 1: QUESTION ANSWERING				
<i>Baselines</i>	4			1
DPRL	1	3		
MathDowers	1	3		1
MIRMU	3	2		
PSU	1	2		
ZBMath			1	
TASK 2: FORMULA RETRIEVAL				
<i>Baseline</i>	1			
DPRL	1	3		
MIRMU	2	3		
NLP-NIST	1			
ZBMath			1	

Teams were from **Canada** (MathDowers), the **Czech Republic** (MIRMU), **Germany** (ZBMath), **India** (NLP-NIST), and **USA** (DPRL, PSU)

# Evaluation:

## Answer Retrieval (77 topics)



**Evaluation pool:** set of unique answers in top-k results from runs

**Pool Depths (k)**

50 Primary, manual, **baseline**

20 Alternate runs

**Pooled Hits (answers)**

> 39,000 hits ( Avg: 508.5 / topic )

**Average Time to Assess a Hit**

63.1 seconds

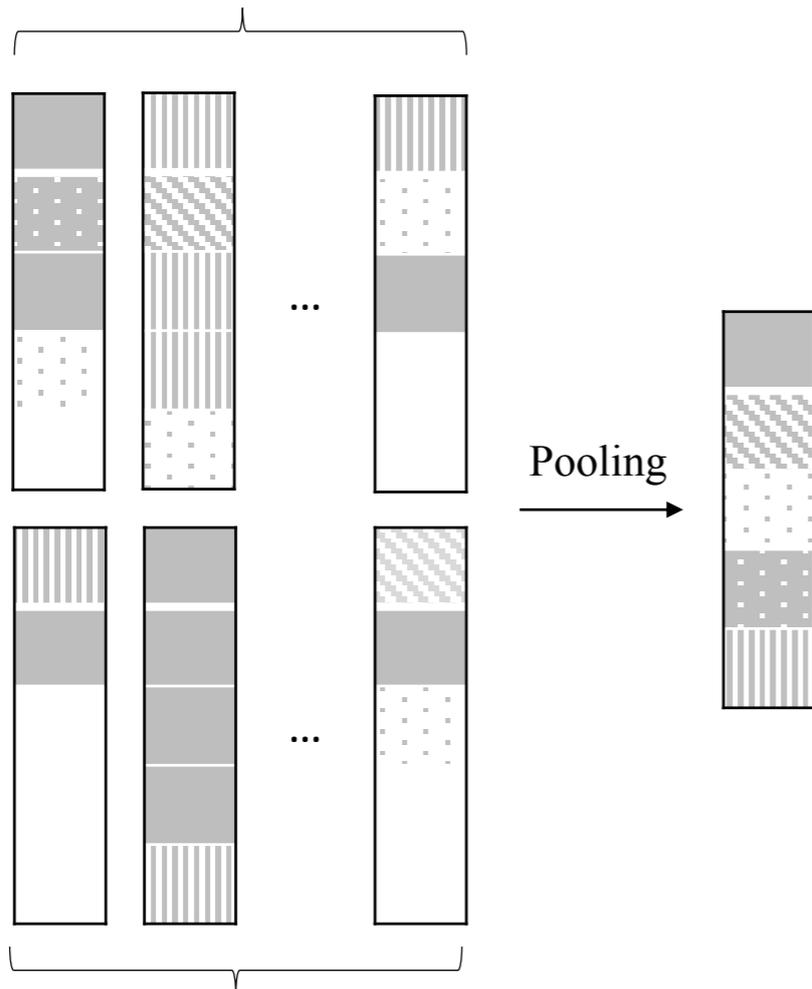
- 4-level relevance (Not, Low, Med, High)

# Evaluation:

## Formula Search (45 topics)

### Task 2: FORMULA RETRIEVAL

Top-25 visually distinct formulae selected from **baseline** and each **primary** run, for a given formula query.



Top-10 visually distinct formulae selected from each **alternate** run for a given formula query.

**Evaluation pool:** visually distinct formula set, differing by symbol positions on writing lines where available, LaTeX otherwise

Up to 5 posts per distinct formula selected  
MAX relevance score used for each formula

**Pool Depths for Distinct Formulas (k)**

25 Primary, baseline

10 Alternate runs

**Pooled Visually Distinct Formulas**

> 5,600 ( Avg: 125 distinct formulae / topic )

- Only 1.6% of formulas in > 5 posts

**Avg. Formula Eval. Time (1-5 posts apiece)**

38.1 seconds - 4-level relevance (N,L,M,H)

**ARQMath**

# Answer Retrieval Results (77 topics)

**Rank Metric:** avg. nDCG', prime for evaluated hits only (Sakai & Kando, 2008). Uses graded relevance.

**Binarization:** avg. MAP', avg. Precision@10 with Medium + High ratings considered 'relevant'

**Linked MSE Post Baseline:** semi-oracle, access to MSE duplicate question links. All answers from duplicate questions ranked by votes

**MathDowers:** BM25+ ranking over Symbol Layout Tree (SLT) features and keywords in a single framework, Tangent-L (Fraser et al., 2018)

RUN	DATA	RUN TYPE		EVALUATION MEASURES		
		P	M	nDCG'	MAP'	P@10
<b>Baselines</b>						
<i>Linked MSE posts</i>	n/a	(✓)		<b>(0.279)</b>	<b>(0.194)</b>	<b>(0.384)</b>
<i>Approach0*</i>	Both		✓	0.250	0.099	0.062
<i>TF-IDF + Tangent-S</i>	Both	(✓)		0.248	0.047	0.073
<i>TF-IDF</i>	Text	(✓)		0.204	0.049	0.073
<i>Tangent-S</i>	Math	(✓)		0.158	0.033	0.051
<b>MathDowers</b>						
alpha05noReRank	Both			<b>0.345</b>	<b>0.139</b>	<b>0.161</b>
alpha02	Both			0.301	0.069	0.075
alpha05translated	Both		✓	0.298	0.074	0.079
alpha05	Both	✓		0.278	0.063	0.073
alpha10	Both			0.267	0.063	0.079
<b>PSU</b>						
PSU1	Both			0.263	0.082	0.116
PSU2	Both	✓		0.228	0.054	0.055
PSU3	Both			0.211	0.046	0.026
<b>MIRMU</b>						
Ensemble	Both			0.238	0.064	0.135
SCM	Both	✓		0.224	0.066	0.110
MIaS	Both	✓		0.155	0.039	0.052
Formula2Vec	Both			0.050	0.007	0.020
CompuBERT	Both	✓		0.009	0.000	0.001
<b>DPRL</b>						
DPRL4	Both			0.060	0.015	0.020
DPRL2	Both			0.054	0.015	0.029
DPRL1	Both	✓		0.051	0.015	0.026
DPRL3	Both			0.036	0.007	0.016
<b>zbMATH</b>						
zbMATH	Both	✓	✓	0.042	0.022	0.027

# Formula Search Results (45 topics)

**Rank Metric:** avg. nDCG'

**Tangent-S baseline:** SLT and Operator Tree (OPT) feature + structure matching + score weights (Davila & Zanibbi, 2017)

**TangentCFTED:** TangentCFT (Mansouri et al., 2019) FastText SLT and OPT tuple embeddings + tree edit-distance reranking

RUN	DATA	P	EVALUATION MEASURES		
			nDCG'	MAP'	P@10
<b>Baseline</b>					
<i>Tangent-S</i>	Math	(✓)	( 0.506 )	( 0.288 )	( 0.478 )
<b>DPRL</b>					
TangentCFTED	Math	✓	0.420	0.258	0.502
TangentCFT	Math		0.392	0.219	0.396
TangentCFT+	Both		0.135	0.047	0.207
<b>MIRMU</b>					
SCM	Math		0.119	0.056	0.058
Formula2Vec	Math	✓	0.108	0.047	0.076
Ensemble	Math		0.100	0.033	0.051
Formula2Vec	Math		0.077	0.028	0.044
SCM	Math	✓	0.059	0.018	0.049
<b>NLP_NITS</b>					
formulaembedding	Math	✓	0.026	0.005	0.042

# Closing Notes

Training models directly from MSE votes / selections was not beneficial for a number of teams

‘Pure’ embedding models did not obtain the strongest results. Surprisingly, best performing systems did not use embeddings

**Task 1** is the first CQA task for math-aware search; **Task 2** is the first context-aware formula retrieval task

For Task 2, +27 topics after evaluation, **74 Task 2 topics** now available in addition to the **77 topics for Task 1**

Collection data, tools, and assessments available online.

# ARQMath Assessors



Kiera Gross



Gabriella Wolf



Justin Haverlick



Josh Anglum



Ken Shultes



Riley Kieffer

Wiley Dole

Minyao Li

Assessors are senior & recently graduated **undergraduate math students** from RIT

# ARQMath Assessors



Kiera Gross



Gabriella Wolf



Justin Haverlick



Josh Anglum



Ken Shultes



Riley Kieffer

Wiley Dole

**Important Note:**  
Justin, Josh and Minyao will participate in panels on assessment during ARQMath sessions Friday



Minyao Li

# ARQM**Math**



Please join our sessions on Friday!

Also, please consider participating  
next year at CLEF 2021!

<https://www.cs.rit.edu/~dprl/ARQMath>



#ARQMath

Send Email to: [rxzvcs@rit.edu](mailto:rxzvcs@rit.edu)

Our thanks to the National Science Foundation (USA)

