ARQMath

Answer Retrieval for Questions on Math

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

CLEF 2020 Thessaloniki

Please join us! Feedback also appreciated!
Email: rxzvcs@rit.edu

Richard Zanibbi
Rochester Institute of Technology
rxzvcs@rit.edu

Douglas W. Oard
University of Maryland
oard@umd.edu

Anurag Agarwal
Rochester Institute of Technology
axasma@rit.edu

#ARQMath

— Goals —

Advance techniques for math-aware search and semantic analysis of mathematical notation and texts

— Collection —

Math Stack Exchange community QA forum (~**1.1 million questions**)

— Formula Representations —

Appearance: the arrangement of symbols on writing lines.
- LaTeX
- Presentation MathML

Semantic: operators, arguments, order of operations
- Content MathML

Key Dates

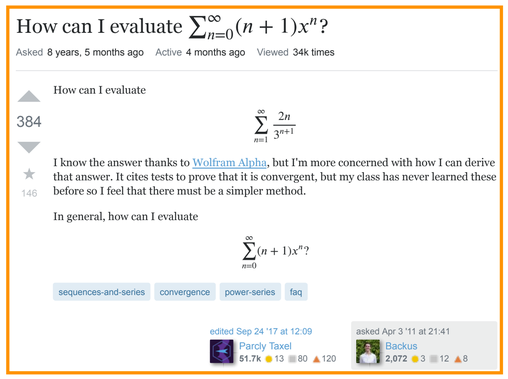
2019
- Nov 1 Release data, sample queries
- Nov 5 Registration opens

2020
- Jan 15 Test queries released
- May 1 Submissions close
- Aug 14 Final lab report
- Sep 22-25 ARQMath@CLEF 2020

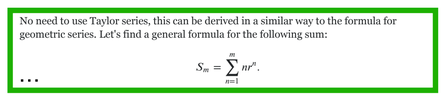
Task 1: Finding Answers to Math Questions

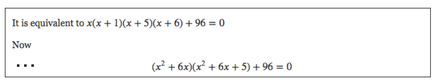
Given a **posted question as a query**, search all **answer posts** and return **relevant answers**.

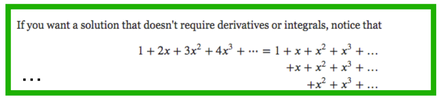
Query



Search Results

1


2


3


Topics and Runs

- **Goal:** Create 100 topics using questions containing text and at least one formula
- Queries may be processed using either text, math, or both text and math
- Manual and automatic runs will be collected

Evaluation

- Top-k hits (e.g., top-20) + additional manual runs by organizers will be pooled
  - Most topics assessed once, some doubly-assessed to check agreement
  - Assessors will include volunteers from teams, along with hired assessors
- During assessment, we propose **organizing topics into three sets**: (1) all topics, (2) topics where the text seems to characterize the topic on its own, and (3) topics where the formula(s) seem to characterize the topic on their own.
- We propose using **extended inferred Average Precision (xinfAP)** over the three topic sets as the primary measure to support comparison with future systems that do not contribute to the judgement pool

Task 2: Formula Search

Given a **formula query from a question**, search **formulas in question and answer posts** and return **relevant formulas.**

Query

$\sum_{n=0}^{\infty}(n+1)x^n$

Search Results

1 $\sum_{n=0}^{\infty}(n+1)x^n$

2 $\sum_{n=0}^{\infty}(n+1)x^n$

3 $\int_0^1 \frac{\ln(x+1)}{x^2+1}\mathrm{d}x$

Topics and Runs

- **Goal:** Create 50 topics using formulas extracted from topics in Task 1
  - **Formulas provided in LaTeX, Presentation MathML, and Content MathML**
- Annotators create **human-readable narratives** to define *relevant* formulas, e.g., subexpressions, alternative notation, simplification, applications in specific fields, etc.
- Manual and automatic runs will be collected

Evaluation

- Top-k formulas from participants + additional manual runs by organizers will be pooled
  - Assessors can use formula hits + pools from Task 1 to identify similar formulas
  - Most topics assessed once some doubly-assessed to check agreement
  - Assessors include volunteers from teams along with hired assessors
- We propose using **extended inferred Average Precision (xinfAP)** over formula topics as the primary measure for comparing systems