

Dowsing for Math Answers with Tangent-L

Yin Ki NG¹, Dallas J. Fraser², Besat Kassaie¹, George Labahn¹,
Mirette S. Marzouk¹, Frank Wm. Tompa¹, and Kevin Wang¹

¹ David R. Cheriton School of Computer Science, University of Waterloo,
Waterloo, ON, Canada, N2L 3G1

² DigitalEd, 630 Weber Street North, Suite 100, Waterloo, ON, Canada, N2V 2N2

Abstract. We present our application of the math-aware search engine Tangent-L to the ARQMath Community Question Answering (CQA) task. Our approach performs well, placing in the top three positions out of all 23 submissions, including the baseline runs.

Tangent-L, built on the text search platform Lucene, handles math formulae by first converting a formula's Presentation MathML representation into a Symbol Layout Tree, followed by extraction of math tuples from the tree that serve as search terms. It applies BM25⁺ ranking to all math tuples and natural language terms in a document during searching.

For the CQA task, we index all question-answer pairs in the Math Stack Exchange corpus. At query time, we first convert a topic question into a bag of formulae and keywords that serves as a formal query. We then execute the queries using Tangent-L to find the best matches. Finally, we re-rank the matches by a regression model that was trained on metadata attributes from the corpus.

Our primary run produces an nDCG' value of 0.278 and MAP' value of 0.063, where these are two common measures of quality for ranked retrieval. However, our best performance, an nDCG' value of 0.345 and MAP' value of 0.139, is achieved by an alternate run without re-ranking. Follow-up experiments help to explain which aspects of our approach lead to our success.

Keywords: Community Question Answering (CQA) · Mathematical Information Retrieval (MathIR) · Symbol Layout Tree · Lucene · Mathematics Stack Exchange.

1 Research Objectives

Mathematical Information Retrieval (MathIR) focuses on using mathematical formulae and terminology to search and retrieve documents that include mathematical content [11, 22]. MathIR is important because content expressed in formal mathematics and formulae is often crucial and non-negligible in STEM papers [27]. In the recent decade, the MathIR research community has been

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CLEF 2020, 22-25 September 2020, Thessaloniki, Greece.

growing and developing ever-improved math-aware search systems (e.g. [7, 10, 13, 14, 23–28, 30, 32]). Most of these efforts have been encouraged through a series of MathIR evaluation workshops at NTCIR-10 [2], NTCIR-11 [3], and NTCIR-12 [29]. The workshops have provided corpora derived from arXiv and Wikipedia for traditional ad-hoc retrieval tasks and formula search tasks, and the data and tasks have since served as benchmarks for the research community.

The ARQMath Lab (ARQMath) [16, 31] is the first Community Question Answering (CQA) task with questions involving math data, using collections from the Math Stack Exchange (MSE),³ a math question answering site. The training corpus contains approximately 1.1 million mathematical questions and 1.4 million answers, covering MSE threads from the year 2010 to 2018. Like the NTCIR workshops that precede it, it is centered around an evaluation exercise that aims to advance math-aware search and the semantic analysis of mathematical notation and texts. The main task of ARQMath is the answer retrieval task, in which participating systems need to find answers to a set of mathematical questions among previously posted answers on MSE. A secondary task considers matching relevant formulae drawn from the mathematical questions from the same collection. Participating teams were asked to submit up to five runs for either or both tasks, and selected results received relevance assessments from human evaluators.

Related tasks to the ARQMath Lab task have been held previously: a recent math question answering task was held as part of SemEval-2019 [12], following CQA challenge series held at SemEval-2015 [18], SemEval-2016 [19], and SemEval-2017 [17]. The math question answering task at SemEval-2019 considered a math question set that was derived from Math SAT practice exams. This task was different from the ARQMath CQA task, since the data does not involve question-answer threads from a community forum, and the task targeted identification of one uniquely correct answer by multiple-choice selection or by numerical computation, instead of retrieving all relevant answers from a corpus. On the other hand, the earlier CQA challenge series at SemEval involved question-comment threads from the Qatar Living Forum, which is a data collection that is similar to the MSE collection. This CQA challenge series, however, differs from the ARQMath CQA task in that the questions are not necessarily mathematical, and the task objective is answer-ranking instead of answer retrieval from a corpus. Besides SemEval tasks, related tasks under the question-answering context were also held previously at TREC, CLEF, and NTCIR ([1], [20]), but the data involved was not drawn from the mathematical domain and the data did not follow a community-forum structure in general.

Our team of MathDowsers⁴ participated in the ARQMath CQA task, with an approach based on the Tangent-L system, a math-aware search engine proposed by Fraser et al. [9]. Tangent-L is a traditional math-aware query system developed after NTCIR-12,

³ <https://math.stackexchange.com>

⁴ Waterloo researchers dowsing for math (Fig.1)

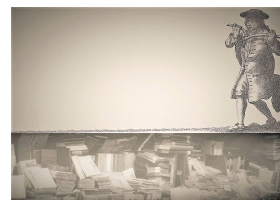


Fig. 1. Dowsing for math.

using the data provided for all three NTCIR math search workshops, and appeared to be competitive with the systems participating in those workshops. We wished to determine whether it could, in fact, perform well against other traditional math-aware query systems and whether a traditional math-aware query system could compete with modern machine-learning approaches that might be adopted by other workshop participants in a question-answering task.

Our experiments included five submitted runs that were designed to address the following research objectives:

- RQ1** What is an effective way to convert each mathematical question (expressed in mathematical natural language) into a formal query consisting of keywords and formulae?
- RQ2** Should keywords or math formulae be assigned heavier weights in a query?
- RQ3** What is the effect of a re-ranking algorithm that makes use of metadata?

We present an overview of Tangent-L in Section 2. In Section 3, we describe our approach to CQA and provide details on how we retrieve and rank answer matches for a mathematical question from the MSE corpus with the use of Tangent-L. The submitted runs and the results are discussed in Section 4. In Section 5, we present conclusions and propose future work.

2 Overview of Tangent-L

The Tangent-L system is the core component of our submission. It is a traditional query system built on the popular Lucene [4] text search platform, adding methods adapted from the Tangent-3 math search system [30] so that it is capable of matching documents based on queries with keywords and formulae.

Tangent-L handles formulae by converting them into a bag of “math terms” to be matched in the same way that natural language terms are handled by Lucene. More specifically, Tangent-L takes as input a formula in Presentation MathML [5] format⁵ and converts it into a *symbol layout tree* (SLT) [30] where nodes represent the math symbols and edges represent spatial relationships between these symbols (Figure 2). Thereafter, this tree-like representation is traversed to extract a set of features, or “math tuples,” of four types to capture local characteristics of a math formula as depicted in Figure 3. In preparation for search, the math tuples replace the formula itself in the document and are then considered by Tangent-L as if each were a term in the text to be matched.

After formula conversion, Tangent-L applies BM25⁺ ranking [15] to all the terms in a document. Specifically, given a collection of documents D containing $|D|$ documents and a query q consisting of a set of query terms, the score for a document $d \in D$ is given by

$$\text{BM25}^+(q, d) = \sum_{w \in q} \left(\frac{(k+1)tf_{w,d}}{k \left(1.0 - b + b \frac{|d|}{d} \right) + tf_{w,d}} + \delta \right) \log \left(\frac{|D| + 1}{|D_w|} \right) \quad (1)$$

⁵ A formula in L^AT_EX representation can be converted into MathML by using LaTeXXML (<https://dlmf.nist.gov/LaTeXXML/>).

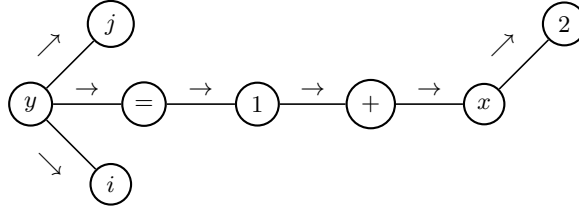


Fig. 2. Symbol Layout Tree for $y_i^j = 1 + x^2$.

| Feature Type | Definition and Extracted Features |
|---------------------|--|
| Symbol pairs | For each edge, start and end symbols and edge label (y, j, \nearrow) $(y, =, \rightarrow)$ (y, i, \searrow) $(=, 1, \rightarrow)$ $(1, +, \rightarrow)$ $(+, x, \rightarrow)$ $(x, 2, \nearrow)$ |
| Terminal symbols | List of symbols with no outedges (j, Δ) (i, Δ) $(2, \Delta)$ |
| Compound symbols | List of outedge labels for nodes with more than one outedge $(y, \nearrow \searrow \rightarrow)$ |
| Augmented locations | For each feature of the first three types, that feature together with the path to the feature's (first) symbol $(y, j, \nearrow, \emptyset)$ $(y, =, \rightarrow, \emptyset)$ $(y, i, \searrow, \emptyset)$ $(=, 1, \rightarrow, \rightarrow)$ $(1, +, \rightarrow, \rightarrow \rightarrow)$ $(+, x, \rightarrow, \rightarrow \rightarrow \rightarrow)$ $(x, 2, \nearrow, \rightarrow \rightarrow \rightarrow \rightarrow)$ (j, Δ, \nearrow) (i, Δ, \searrow) $(2, \Delta, \rightarrow \rightarrow \rightarrow \rightarrow \nearrow)$ $(y, \nearrow \searrow \rightarrow, \emptyset)$ |

Fig. 3. Extracted features (“math tuples”) to represent the formula in Figure 2.

where k , b , and δ are constants (following common practice, chosen to be 1.2, 0.75, and 1, respectively); $tf_{w,d}$ is the number of occurrences of term w in document d ; $|d|$ is the total number of terms in document d ; $\bar{d} = \sum_{d \in D} \frac{|d|}{|D|}$; and $|D_w|$ is the number of documents in D containing term w . This formula is easily applied to a bag of query terms: if a term is repeated in the query, the corresponding score for that term is simply accumulated multiple times. To allow for math tuples to be given a weight that differs from natural language terms, we assign weights to query terms as follows:

$$BM25_w^+(q_t \cup q_m, d) = BM25^+(q_t, d) + \alpha \cdot BM25^+(q_m, d) \quad (2)$$

where q_t is the set of keywords in a query, q_m is the set of math feature tuples in that query, and α is a parameter to adjust the relative weight applied to math tuples.

In the NTCIR-12 arXiv Main task benchmark, where queries are composed of formulae and keywords, the Tangent-L system gives a comparable performance to other MathIR systems [9]. We are interested in determining whether and how Tangent-L could be adapted to address the ARQLab CQA task.

3 Methodology

After indexing the corpus, we adopt a three-phase architecture to return answer matches for a mathematical question:

1. *Conversion*: Transform the input (a mathematical question posed on MSE) into a well-formulated query consisting of a “bag” of formulae and keywords.
2. *Searching*: Use Tangent-L to execute the formal query to find the best matches against the indexed corpus (MSE question-answer pairs).
3. *Re-ranking*: Re-order the best matches by considering additional metadata (such as votes and tags) associated with question-answer pairs.

3.1 Conversion: Extracting Formulae and Keywords from Questions

For the CQA task, participants are given 98 real-world mathematical questions selected from MSE posts in 2019.⁶ Each mathematical question is a *topic* that contains: (1) the *topic-ID*, (2) the *title* for this topic, (3) the *question* (body text) for this topic, and (4) the *tags* for this topic. The title and body text are free text fields describing a question in mathematical natural language (as opposed to formal logic, for example), and the tags indicate the question’s academic areas.

We adopt the following automated mechanism to extract a list of topic formulae and keywords:

Topic formulae. Formulae within the topic’s title and body text are extracted. All formulae within the title are selected as topic formulae. Formulae within the body text are selected only if they are not single variables (e.g., n or i) nor isolated numbers (e.g., 1 or 25).

Topic keywords. Keyword selection is summarized in Algorithm 1. Each of the topic’s tags is selected as one of the topic keywords. For the topic’s title and body text, we first tokenize the text (Algorithm 2) to obtain a list of potential word tokens. A token is then selected as a keyword if it contains a hyphen (such as “Euler-Totient” or “Cesáro-Stolz”) (Algorithm 3) or it is not a stopword (Algorithm 4) and its *stem* appears on a pre-constructed list of *mathematical stems*.

The mathematical stem list is created in a pre-processing step by automatically extracting terms from two sources: (1) tags from the indexed MSE dataset and (2) titles from English Wikipedia articles comprising the corpus for the NTCIR-12 MathIR Wikipedia task[29]. For the former source, each tag present in the ARQMath MSE corpus is tokenized, stemmed, and added to the stem list. For example, two stems (“totient” and “function”) are added to the list for the tag “totient-function.” For the latter, we first collect the HTML filenames of all articles used in the NTCIR-12 MathIR Wikipedia corpus, where each filename reflects the corresponding Wikipedia article’s title. The filenames are then transformed by removing file extensions and replacing underscores and parentheses

⁶ In addition, three extra questions are provided as samples with annotated answers.

with spaces. Each hyphenated term is expanded to include all components as well as the hyphenated term itself. The resulting cleaned text strings are tokenized, stemmed, and punctuation removed, and the resulting stems are added to the stem list. Thus, for example, four stems (“exponenti,” “logarithm,” “distribut,” and “exponential-logarithm”) are added for the filename Exponential-logarithmic_distribution.html. The resulting mathematical stem list comprises about 1200 stems from MSE tags and about 21,000 stems from Wikipedia article titles.

```

for tags in topic do
    Split input by ',' to obtain a list of tags;
    foreach tag do
        add tag to keyword list;
        if hyphen in tag then
            further split by '-' to obtain a list of parts;
            foreach part do
                add part to keyword list;
for title, body text in topic do
    foreach token in Tokenize(input) do
        if !IsStopword(token) && (IsPreserved(token) ||
            OnMathList(Stem(token))) then
            append to keyword list;

```

Algorithm 1: ExtractKeywords(topic)

```

Split the input by space into substrings;
foreach substring do
    if hyphen in substring then
        add substring to the token list;
Replace '-' with space throughout input;
foreach token in Treebank-Tokenization(input) do
    add token to the token list;

```

Algorithm 2: Tokenize(input)

```

return ('-' in token) || ('-' in token)

```

Algorithm 3: IsPreserved(token)

```

return (token in stopwords set provided by the NLTK library) || (token
contains only a single character) || (token is a numeric string)

```

Algorithm 4: IsStopword(token)

Table 1. Average, minimum and maximum number of topic formulae and keywords per topic. Example topics are included in the parentheses.

| | <i>Average</i> | <i>Min</i> | <i>Max</i> |
|----------|----------------|------------|------------|
| Formulae | 8 | 1 (A.12†) | 35 (A.81) |
| Keywords | 38 | 5 (A.13) | 155 (A.73) |

(† among others)

When selecting keywords with this procedure, we use the Treebank tokenizer, the Porter stemmer, and a list of English stop-words provided by the Python NLTK library.⁷ Using this approach, on average 8 topic formulae and 38 topic keywords were extracted for the CQA task (Table 1). The complete list of topic formulae and keywords can be found in the Appendix.

3.2 Searching: Answer Retrieval from the Indexed Corpus

We use the Tangent-L system to retrieve answers that match the extracted keywords and formulae.

Indexing. We build the indexed corpus with question-answer pairs. Each indexed unit includes an MSE answer along with the content of its associated question.⁸ For each question-answer pair, we extract the following content from the corpus XML files:

- From the answer:** the body text and the number of votes (from *Posts.xml*);
- From the associated question:** the title, body text, and tags of the question (from *Posts.xml*) plus comments associated with each question (from *Comments.xml*). Additionally, we also include the *titles* of all related and duplicated posts for this question (from *PostLinks.xml*), having first converted all one-way links between posts into two-way links.

All formulae within the text are replaced by their Presentation MathML form using the formula mapping in the TSV files provided with the corpus data. An HTML file containing the final extracted content is then assembled as an indexing unit (Figure 4.). The indexable version of the MSE corpus includes a total of 1,445,488 documents, indexed by Tangent-L in preparation for search.

Searching. Searching the corpus is a straightforward application of Tangent-L for the converted topics. The list of keywords and formulae (in MathML) are passed to the search engine. Tangent-L then converts the formulae to math tuples and uses corresponding postings lists in the index to compute BM25⁺ scores (Equation 2), weighted by a value α that depends on the experimental setup for the run.

⁷ <https://www.nltk.org/>

⁸ As many of us were admonished in school: “You should always include the question as part of your answer.”

```

<body>
  <div class="row" id="question-header">
    <h1> The cow in the field problem (intersecting circular areas) </h1>
  <div class="question">
    <div class="question" data-questionid="#QID#" id="question">
      <div class="post-text">
        <p>What length of rope should be used to tie a cow to an <strong>exterior fence
        post</strong> of a <em>circular</em> field so that the cow ...
      </div>
    </div>
    <div class="post-taglist">
      <span class="post-tag"> geometry </span> <hr>
    </div>
  </div>
  <div class="row" id="answers">
    <div class="answer" data-answerid="62"> <div class="post-text"> <p>So, the area
    of the field is <span class="math-container" id=795 fid=795><?xml version="1.0"
    encoding="UTF-8"?><math xmlns="http://www.w3.org/1998/Math/MathML" alttext=
    "\pi r^2" display="block"> <mrow> <mi>\pi</mi> <mo>~</mo> <msup> <mi>r</mi>
    <mn>2</mn> </msup> </mrow></math></span> and you want the cow to be able to
    graze an area equal to half of that.</p> <p>All you need to do is set up the equation ...
    </div>
  </div>
  <div class="row" id="question-comments">
    <table>
      <tr><td> I'm guessing that most fence posts are at the edge of a field, which
      makes this a far more interesting problem. </td></tr><tr><td> ....
      </tr>
    </table>
  </div>
  <div class="row" id="duplicated">
    <table>
      <tr><td> A confusing word problem related to geometry (circles) </td></tr><tr>
      <td> Is it possible to express the area of the intersection of 2 circles as a
      closed-form expression? </td></tr>
    </table>
  </div>
  <div class="row" id="related">
    <table>
      <tr><td> Find the area where dog can roam </td></tr><tr><td> A goat tied to a
      corner of a rectangle </td></tr>
    </table>
  </div>
</body>

```

Fig. 4. A sample indexed unit.

3.3 Using Answers' Metadata for Re-ranking

The CQA task is made more challenging by the potential to use semantic information relating candidate answers to topic questions. For instance, an answer should be ranked higher in position if it answers the topic question correctly. However, a correct answer does not necessarily have the highest score based merely on matching keywords and formulae by similarity.

As such, we are motivated to explore a re-ranking algorithm applied to the results returned from the search engine. To this end, we wish to use information from the complete MSE corpus to build a model that reflects how valuable a potential answer might be. We hypothesize, with the law of simplicity in mind, that a linear function of the following four variables might serve well:

1. **similarity:** The similarity of keywords and formulae between the target answer (including its associated question) and the topic query is clearly an important component. This is captured by the search score returned from the Tangent-L system.
2. **tags:** The number of overlapping tags between the question associated with the answer and the topic query reflects how well the answer matches the query's academic area(s).
3. **votes:** The fraction of votes received by the answer when posted with its associated question reflects the community's belief in the answer's value.
4. **reputation:** The reputation of the author who wrote an answer implies the trustworthiness of that answer. Intuitively, a good answer comes from an author with a good reputation. This can be computed from the user reputation score, the number of user up-votes, and the number of user down-votes.

The remaining problem is to determine what coefficient values to use when linearly combining these inputs. For this we need a training set that includes relevance assessments, which are not available as part of the corpus.

Mock Relevance Assessments. As a substitute for assessed relevance, we can build a training set of queries from questions already in the corpus: we hypothesize that relevant answers include those that were actually provided for those questions as well as those provided for *duplicate* and *related* questions.

We use tags, related posts, and duplicate posts of a question, and the number of votes for each answer to calculate mock relevance assessments for answers to the training topics, based on the following two observations:

1. Considering the question associated with a target answer, the more "on-topic" that question is to the query, the more relevant the target answer is likely to be for that query. A corpus question and a query are related if they have overlapping tags, but are more related if one is marked as a related post of the other, and still more related if it is marked as a duplicate post of the other (or is, in fact, the same question).

2. If two potential answers are associated with the same question, the one with more votes should be preferred.

With these assumptions, an assessment value can be computed as follows:

– **Integral assessment value.**

1. An answer gets an assessment value of **2**, if the question associated with that answer is a duplicate post to the query question, or if the answer comes from the same question.
2. An answer gets an assessment value of **1**, if the question associated with that answer is a related post to the query question;
3. Otherwise, the answer gets an integral assessment value of **0**

– **Fractional assessment value.**

1. Assuming that all votes for answers are non-negative, an answer gets a normalized fractional assessment value **between 0 and 1**, calculated as the fraction of its votes to the sum of votes of all answers for the associated question, if the associated question is the same or a duplicate or related post to the query question or if the associated question contains any overlapping tags with the query question.
2. Otherwise, the answer gets a fractional assessment value of **0**.

The final mock relevance assessment is the sum of the integral and fractional assessment values and, when all answers have non-negative votes, ranges from 0 to 3 (matching the assessment range expected in the CQA task):

| <i>Score</i> | <i>Meaning</i> |
|--------------|-----------------------------------|
| 0 | irrelevant |
| 0 .. 1 | tags overlap |
| 1 .. 2 | posts are related |
| 2 .. 3 | posts are identical or duplicates |

In fact, some answers in the corpus have negative votes (down-votes). In this case, the fractional assessment value is adjusted as follows: for a thread containing any answers with negative votes, the fraction’s denominator is the absolute value of the total negative votes plus the sum of votes within a thread. If an answer has positive votes, the fraction’s numerator is the absolute value of the total negative votes plus the number of votes for the answer, otherwise, the fraction’s numerator is the (negative) value of the votes. For example, suppose a thread contains three answers, and their votes are -2, 2, 6, respectively. The fractional assessment values for these answers are -0.2, 0.4, 0.8, respectively. Thus the mock relevance assessment penalizes answers with negative votes.

Linear Regression Formula. To determine the coefficients for the linear regression model, we first generate 1300 training topics from the MSE collection. We use the Tangent-L system with $\alpha = 1.0$ (in Equation 2) to retrieve the top 10,000 answers for each valid topic, resulting in a total of 12,970,000 answers.

Table 2. Example items of generated training data for the linear regression model, in decreasing order of the search score returned from the Tangent-L system.

| <i>Regressand</i> | | <i>Regressors</i> | | | | | |
|-------------------|------------|-------------------|-------|----------|--------------|-----------|-------------|
| mock relevance | similarity | tags | vote* | votes-d† | u-reputation | u-upvotes | u-downvotes |
| 3 | 2617.5 | 3 | 2 | 2 | 45496 | 26331 | 66 |
| 3 | 1226.3 | 1 | 5 | 5 | 305891 | 20695 | 647 |
| 1 | 1197.6 | 2 | 3 | 3 | 6227 | 130 | 2 |
| 1 | 1012.6 | 1 | 0 | 0 | 884 | 26 | 9 |
| 1 | 932.8 | 2 | 5 | 5 | 4268 | 244 | 27 |
| 1 | 930.0 | 1 | 2 | 2 | 1743 | 128 | 2 |
| 0.5 | 911.4 | 2 | 1 | 2 | 77437 | 2507 | 17 |
| 1 | 895.3 | 2 | 1 | 1 | 52161 | 1796 | 252 |
| 1 | 864.5 | 1 | 1 | 1 | 53311 | 3710 | 504 |
| 1 | 852.5 | 2 | 4 | 4 | 48172 | 9298 | 1270 |
| 0 | 811.7 | 0 | 1 | 1 | 294098 | 5057 | 218 |

* the original answer vote

† the denominator of the fractional assessment value

Table 3. Parameters of the linear regression model, trained on un-normalized data.

| <i>Parameter</i> | <i>Coefficient Value</i> |
|----------------------|--------------------------|
| (intercept) | 0.021 |
| similarity | $1.68 \times 1e-6$ |
| tags | 0.398 |
| vote | $1.59 \times 1e-4$ |
| vote-d (denominator) | $-6.17 \times 1e-5$ |
| user reputation | $9 \times 1e-8$ |
| user upvotes | $1.5 \times 1e-7$ |
| user downvotes | $1.20 \times 1e-6$ |

Next, we generate the mock relevance assessments for these answers and associate these assessments with the values for similarity, tags, votes, and reputation for those answers, as discussed above. These 12,970,000 tuples then serve as training data for the linear regression model. Example items of the training data are shown in Table 2, and the trained coefficients are shown in Table 3.

We validate the trained linear regression model by first applying the model to predict a relevance score for the top 1000 retrieved answers of a separate set of another 100 topics, and then re-ranking the answers according to the predicted mock relevance score. Finally, we adopt the normalized Discounted Cumulative Gain (nDCG), which measures the gain of a document based on its position in the result list and its graded relevance. We compare the nDCG value of the results for those 100 topics according to the mock relevance assessment before

and after re-ranking. This simple model gives a slight improvement in nDCG after the re-ranking (from 0.3192 to 0.3318).

4 Experiments

4.1 Description of Runs

The MathDowers submitted a primary run and four alternate runs, each run returning at most 1000 answer matches for all 98 topics. The primary run is designed as a combination of our hypotheses for the best configuration for all research objectives. The alternate runs are designed to test these hypotheses: the setup for each of them is the same as the primary run, except for a single aspect that is associated with a testing hypothesis, as described here and summarized in Table 4.

RQ1 *Topic Conversion*

In the primary run, we use the auto-extracted topic keywords and formulae described in Section 3.1 and listed in the Appendix as the query input to the search engine. To compare the effectiveness of our extraction algorithm with human understanding of the questions, an alternate run takes as input lists of up to five topic keywords and up to two formulae per topic, all manually chosen by the lab organizers after reading each question and made available to all participants. This alternate run, which includes “translated” in its label, is therefore a manual run, in contrast to all other runs that are “automatic.”

RQ2 *Formula Weight*

We test the weights of math formulae in a query by tuning Tangent-L’s α parameter (Equation 2). In the primary run, $\alpha = 0.5$, which means math terms are given half the weight of keywords in a query. The reasoning behind this choice is that each formula generates many terms, one for each feature extracted, and previous experiments showed some benefit in reducing the weight accordingly [9]. Two alternate runs operate with $\alpha = 0.2$ (even less weight on formulae) and $\alpha = 1.0$ (Tangent-L’s default setting), respectively.

RQ3 *Re-ranking*

In the primary run, we re-rank the results from the search engine, using the model described in Section 3.3. To evaluate the effectiveness of the model, an alternate run has no re-ranking.

Table 4. A summary of the setup for MathDowser submissions. **P** indicates the primary run, **A** indicates an alternate run, and **M** indicates a manual run.

| | <i>Formula Weight</i> | <i>Topic Conversion</i> | <i>Re-ranking</i> |
|---------------------------|-----------------------|-------------------------|-------------------|
| alpha05 (P) | $\alpha = 0.5$ | auto-extracted | yes |
| alpha02 (A) | $\alpha = 0.2$ | auto-extracted | yes |
| alpha10 (A) | $\alpha = 1.0$ | auto-extracted | yes |
| alpha05-translated (A, M) | $\alpha = 0.5$ | manual | yes |
| alpha05-noReRank (A) | $\alpha = 0.5$ | auto-extracted | no |

4.2 Machine configuration

Indexing. Indexing is done on a Ubuntu 16.04.6 LTS server, with two Intel Xeon E5-2699 V4 Processors (22 cores 44 threads, 2.20 GHz for each), 1024GB RAM and 8TB disk space (on an USB3 external hard disk). The size of the document corpus is 24.3GB.

Tangent-L requires 5.0GB of storage on the hard drive and approximately 6 hours to index all documents with parallel processing.

Searching and Re-ranking. Training and testing the model for re-ranking is done on a Linux Mint 19.1 machine, with an Intel Core i5-8250U Processor (4 cores 8 threads, up to 3.40 GHz), 24GB RAM and 512GB disk space.⁹

Model training using the Python scikit-learn library¹⁰ takes less than 30 seconds, and re-ranking for all 98 topics requires around 3 seconds per run.

Searching is executed on this same Mint machine, and retrieval time statistics for Tangent-L are reported in Table 5.

Table 5. Retrieval times per topic, in seconds, using single-threaded processing.

| | <i>Avg. (sec)</i> | <i>Top-2 Min (sec)</i> | <i>Top-2 Max (sec)</i> |
|--------------------|-------------------|-----------------------------|---------------------------|
| alpha05 † | 13.3 | 0.669 (A.67) / 0.775 (A.94) | 63.4 (A.76) / 48.7 (A.28) |
| alpha02 | 13.3 | 0.661 (A.67) / 0.850 (A.94) | 59.1 (A.76) / 49.5 (A.28) |
| alpha10 | 13.1 | 0.616 (A.67) / 0.784 (A.83) | 54.0 (A.76) / 48.7 (A.28) |
| alpha05-translated | 5.3 | 0.247 (A.99) / 0.291 (A.94) | 32.8 (A.11) / 25.0 (A.67) |
| alpha05-noReRank | 13.3 | 0.669 (A.67) / 0.775 (A.94) | 63.4 (A.76) / 48.7 (A.28) |

† The run *alpha05* does not, in fact, take any *additional* retrieval time, since it merely re-ranks the retrieved items from the *alpha05-noReRank* run.

4.3 Results and Observations

The primary measure for the task is the Normalized Discounted Cumulative Gain (nDCG) with unjudged documents removed (thus nDCG'). Following the organizers' practice [31], we also measure Mean Average Precision (MAP) with unjudged documents removed (MAP') and Precision at top-10 matches (P@10). Additionally, we calculate the bpref measure, which other researchers have found useful, and the count for *Unjudged Answers* within the top-*k* retrieved answers for each topic.

⁹ A NVIDIA GeForce MX150 graphics card with 2GB on-card RAM is available on the machine, but it was not used for the experiments.

¹⁰ <https://scikit-learn.org>

The lab organizers obtained relevance assessments for 77 of the 98 topics. With respect to the primary measure, MathDowers placed in the top three positions out of all 23 submissions including the five baseline runs [31]. The retrieval performance for the MathDowers’ submissions is summarized in Table 6, along with the performance of baseline systems provided by the lab organizers. One of the five baselines is an unrealizable model and the other four are traditional text or math-aware query systems adapted for the task. They are described by the lab organizers as follows [31]:

Linked MSE posts: an unrealizable model “built from duplicate post links from 2019 in the MSE collection (which were not available to participants). This baseline returns all answer posts from 2018 or earlier that were in threads from 2019 or earlier that MSE moderators had marked as duplicating the question post in a topic. The posts are sorted in descending order by their vote scores.”

Approach0: “the ECIR 2020 version of the Approach0 text + math search engine [32], using queries manually created by the third and fourth authors. This baseline was not available in time to contribute to the judgement pools and thus was scored post hoc.” Formulae are represented by *operator trees* (OPT) where internal nodes are operators and leaves are operands, and formulae are matched efficiently by finding structurally identical sub-trees with dynamic pruning.

TF-IDF + Tangent-S: “a linear combination of TF-IDF and Tangent-S results [see below]. To create this combination, first the relevance scores from both systems were normalized between 0 and 1 using min-max normalization, and then the two normalized scores were combined using an unweighted average.”

TF-IDF: “a TF-IDF (term frequency–inverse document frequency) model from the Terrier platform [21]... Formulae are represented using their $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ string... The TF-IDF baseline used default parameters in Terrier.”

Tangent-S: like Tangent-L, an extension from the Tangent-3 math search system [30]: “a formula search engine using SLT and OPT formula representations [6]. One formula was selected from each Task 1 question title if possible; if there was no formula in the title, then one formula was instead chosen from the question’s body. If there were multiple formulae in the selected field, the formula with the largest number of nodes in its SLT representation was chosen... (Tangent-S) retrieves formulae independently for each representation, and then linearly combines SLT and OPT scoring vectors for retrieved formulae [6]. For ARQMath, we used the average weight vector from cross validation results obtained on the NTCIR-12 formula retrieval task.”

From the top half of Table 6, the first observation is that the primary run, with our presumed best configuration, performs better than only one of our alternate runs. The second observation is that lowering the weight placed on math terms ($\alpha = 0.2$) improves the performance, and using the default weight ($\alpha = 1.0$) hurts the performance. Thirdly, the alternate run using manually extracted

Table 6. CQA (main task) results, averaged over 77 topics. For post-experiment runs, top-2 values in each column are highlighted. Parentheses indicate a baseline relying on information not available to participants.

| | <i>Evaluation Measures</i> | | | | <i>Unjudged Answers</i> | | |
|---------------------------|----------------------------|----------------|----------------|----------------|-------------------------|---------------|---------------|
| | <i>nDCG'</i> | <i>MAP'</i> † | <i>P@10</i> † | <i>bpref</i> † | <i>top-10</i> | <i>top-20</i> | <i>top-50</i> |
| Baseline | | | | | | | |
| <i>Linked MSE posts</i> | (0.279) | (0.194) | (0.384) | (0.214) | 15 | 33 | 45 |
| <i>Approach0</i> | 0.247 | 0.099 | 0.062 | 0.115 | 587 | 1215 | 3202 |
| <i>TF-IDF + Tangent-S</i> | 0.248 | 0.047 | 0.073 | 0.044 | 0 | 1 | 6 |
| <i>TF-IDF</i> | 0.204 | 0.049 | 0.073 | 0.043 | 1 | 2 | 4 |
| <i>Tangent-S</i> | 0.158 | 0.033 | 0.051 | 0.033 | 1 | 7 | 20 |
| Submissions | | | | | | | |
| alpha05 (P) | 0.278 | 0.063 | 0.073 | 0.041 | 0 | 1 | 3 |
| alpha02 (A) | 0.301 | 0.069 | 0.075 | 0.044 | 0 | 1 | 530 |
| alpha10 (A) | 0.267 | 0.063 | 0.079 | 0.042 | 0 | 2 | 236 |
| alpha05-translated (A) | 0.298 | 0.074 | 0.079 | 0.050 | 0 | 1 | 3 |
| alpha05-noReRank (A) | 0.345 | 0.139 | 0.161 | 0.126 | 2 | 3 | 1796 |
| Post-experiment | | | | | | | |
| alpha10-noReRank | 0.327 | 0.134 | 0.158 | 0.123 | 4 | 79 | 1865 |
| alpha10-trans-noR | 0.357 | 0.149 | 0.164 | 0.137 | 211 | 557 | 2253 |
| alpha05-trans-noR | 0.365 | 0.152 | 0.162 | 0.140 | 207 | 545 | 2227 |
| alpha02-noReRank | 0.368 | 0.146 | 0.178 | 0.134 | 35 | 157 | 1784 |
| alpha02-trans-noR | 0.372 | 0.153 | 0.168 | 0.138 | 193 | 535 | 2205 |
| alpha01-noReRank | 0.388 | 0.153 | 0.182 | 0.142 | 105 | 362 | 1930 |
| alpha01-trans-noR | 0.387 | 0.158 | 0.169 | 0.142 | 198 | 560 | 2203 |
| alpha001-noReRank | 0.247 | 0.091 | 0.048 | 0.100 | 616 | 1271 | 3296 |
| alpha001-trans-noR | 0.348 | 0.157 | 0.155 | 0.155 | 371 | 884 | 2734 |

† using H+M binarization

formulae and keywords performed better than the primary run. Finally, the alternate run without re-ranking achieves the best performance in all evaluation measures. This best submission also performs well with respect to all other evaluation measures, with the best (but unrealizable) baseline system, *Linked MSE posts*, being the only submission to perform better with respect to those other measures.

In retrospect, it appears as if all aspects of our primary run leave room for improvement. In order to explore these observations more closely, we execute several additional runs designed after the conclusion of the formal experiment (*post-experiment*). As detailed in Table 7, these runs examine the performance of our system without re-ranking for additional values of α and with automatic or manual choice of keywords and formulae. We can now summarize our insights from all of Table 6 with respect to our research objectives.

The effect of re-ranking. Comparing the result from the submissions and the post-experiment runs, we see that our re-ranking design was detrimental to the performance (**RQ3**). A consistent drop of at least 0.06 in nDCG' can be observed for runs after re-ranking (e.g *alpha02* vs. *alpha02-noReRank*). A similar deterioration can be observed in other evaluation measures as well.

The effect of α . Considering runs without re-ranking, we confirm that the performance gradually improves with the decrease of weight for math formulae in a query. When we decrease α from 1.0 to 0.1, a gain of 0.06 in nDCG' is achieved for runs with auto-extracted topic conversion (*alpha10-noReRank* vs. *alpha01-noReRank*). Similarly, a gain of 0.03 in nDCG' is achieved for runs with manual topic conversion (*alpha10-trans-noR* vs. *alpha01-trans-noR*). Gains are also observed using other evaluation measures.

Unfortunately, when α is set to a much smaller value, namely 0.01, the overall performance is questionable since the result of different evaluation measures is contradictory. We observe that for the run with auto-extracted topic conversion

Table 7. A summary of the setup of post-experimental runs.

| | <i>Formula Weights</i> | <i>Topic Conversion</i> | <i>Re-ranking</i> |
|--------------------|------------------------|-------------------------|-------------------|
| alpha10-noReRank | $\alpha = 1.0$ | auto-extracted | no |
| alpha10-trans-noR | $\alpha = 1.0$ | manual | no |
| alpha05-trans-noR | $\alpha = 0.5$ | manual | no |
| alpha02-noReRank | $\alpha = 0.2$ | auto-extracted | no |
| alpha02-trans-noR | $\alpha = 0.2$ | manual | no |
| alpha01-noReRank | $\alpha = 0.1$ | auto-extracted | no |
| alpha01-trans-noR | $\alpha = 0.1$ | manual | no |
| alpha001-noReRank | $\alpha = 0.01$ | auto-extracted | no |
| alpha001-trans-noR | $\alpha = 0.01$ | manual | no |

(*alpha001-noReRank*), all measures are at their lowest compared to runs with other choices of α . However, for the run with manual topic conversion (*alpha001-trans-noR*), nDCG' and P@10 are still at their lowest, but MAP' is the second-best, and bpref is at its highest among all other choices of α .

Table 8 shows that with $\alpha = 0.01$ unjudged answer coverage at top- k for *alpha001-noReRank* is extraordinary high (at least 80%). Similarly, among the manual runs, *alpha001-trans-noR* also has by far the highest unjudged answer coverage for various values of k across all choices of α . The large percentage of unjudged answers implies that the evaluation of these two runs might not actually be informative.

We conclude that keywords should be assigned heavier weights in a query (**RQ2**). Good performance is achieved when math terms are given one-tenth of the weight of keywords in a query ($\alpha = 0.1$). We are unable to determine the effect when the weight of math formulae is set even smaller.

The effect of topic conversion. Table 6 shows that with high α values (0.5 and 1.0), runs with manual topic conversion consistently perform better than runs with auto-extracted topic conversion in all evaluation measures. However, with lower α values (0.1 and 0.2), runs with auto-extracted topic conversion performs better for P@10. When $\alpha = 0.1$, which we have just concluded to be the best setting for α , the performance of the two runs with respect to nDCG' and bpref measures becomes essentially indistinguishable. We conclude that the proposed method to convert a topic into a query composed of keywords and formulae is competitive with human ability to select search terms (**RQ1**). We discuss this observation further in Section 5.

Table 8. *Unjudged answers* coverage for runs without re-ranking for 77 topics.

| <i>Topic Conversion</i> | <i>Top-10(%)</i> | <i>Top-20(%)</i> | <i>Top-50(%)</i> |
|-------------------------|------------------|------------------|------------------|
| Auto-extracted | | | |
| alpha10-noReRank | 0.5 | 5.1 | 48.4 |
| alpha05-noReRank | 0.3 | 0.2 | 46.6 |
| alpha02-noReRank | 4.5 | 10.2 | 46.3 |
| alpha01-noReRank | 13.6 | 23.5 | 50.1 |
| alpha001-noReRank | 80.0 | 82.5 | 85.6 |
| Manual | | | |
| alpha10-trans-noR | 27.4 | 36.2 | 58.5 |
| alpha05-trans-noR | 26.9 | 35.4 | 57.8 |
| alpha02-trans-noR | 25.1 | 34.7 | 57.3 |
| alpha01-trans-noR | 25.7 | 36.4 | 57.2 |
| alpha001-trans-noR | 48.3 | 57.4 | 71.0 |

Table 9. Number of test topics in each category.

| <i>Dependency</i> | <i>Count</i> | <i>Topic Type</i> | <i>Count</i> | <i>Difficulty</i> | <i>Count</i> |
|-------------------|--------------|-------------------|--------------|-------------------|--------------|
| Text | 13 | Computation | 26 | Easy | 32 |
| Formula | 32 | Concept | 10 | Medium | 21 |
| Both | 32 | Proof | 41 | Hard | 24 |

Table 10. Category performance of *alpha05-noReRank* and *Linked MSE posts* in all evaluation measures, with the better performance measure for each category highlighted in bold.

| | <i>alpha05-noReRank</i> | | | | <i>Linked MSE posts</i> | | | |
|-------------------|-------------------------|-------------|-------------|--------------|-------------------------|----------------|----------------|----------------|
| | <i>nDCG'</i> | <i>MAP'</i> | <i>P@10</i> | <i>bpref</i> | <i>nDCG'</i> | <i>MAP'</i> | <i>P@10</i> | <i>bpref</i> |
| Overall | 0.345 | 0.139 | 0.161 | 0.126 | (0.279) | (0.194) | (0.384) | (0.214) |
| Dependency | | | | | | | | |
| Text | 0.273 | 0.084 | 0.115 | 0.085 | (0.286) | (0.145) | (0.418) | (0.193) |
| Formula | 0.446 | 0.192 | 0.203 | 0.164 | (0.260) | (0.173) | (0.323) | (0.194) |
| Both | 0.272 | 0.109 | 0.138 | 0.105 | (0.295) | (0.198) | (0.360) | (0.203) |
| Topic Type | | | | | | | | |
| Computation | 0.325 | 0.116 | 0.146 | 0.106 | (0.211) | (0.128) | (0.273) | (0.139) |
| Concept | 0.189 | 0.025 | 0.040 | 0.023 | (0.277) | (0.183) | (0.360) | (0.176) |
| Proof | 0.395 | 0.182 | 0.200 | 0.164 | (0.323) | (0.210) | (0.405) | (0.240) |
| Difficulty | | | | | | | | |
| Easy | 0.361 | 0.147 | 0.178 | 0.132 | (0.293) | (0.201) | (0.392) | (0.206) |
| Medium | 0.306 | 0.114 | 0.129 | 0.098 | (0.293) | (0.175) | (0.342) | (0.195) |
| Hard | 0.357 | 0.151 | 0.167 | 0.143 | (0.248) | (0.152) | (0.315) | (0.189) |

4.4 Strength and Weakness

To gain a deeper insight into the behaviour of our best-performing submission run (*alpha05-noReRank*), we examine its performance within each topic category, as determined by the lab organizers over three aspects. *Dependency* shows to what extent a topic depends on the text, formula, or both. *Topic Type* gives a broad categorization of whether a topic asks about a computation, a concept, or a proof. *Difficulty* approximates how hard a topic is to answer, using three levels of difficulty: easy, medium, and hard. The breakdown counts of the 77 test topics by category is shown in Table 9.

Table 10 tabulates the performance by category for our best-performing submission run vs. the best-performing (but unrealizable) baseline. It shows that our system has a weaker performance than the ideal with respect to MAP', P@10, and bpref measures in all categories. However, with respect to the nDCG' measure, its better overall performance results from its performance in some particular categories.

Table 11. Number of test topics of different *Topic Types* and *Difficulties* for each *Dependency*.

| | <i>Dependency</i> | | |
|-------------------|-------------------|----------------|-------------|
| | <i>Text</i> | <i>Formula</i> | <i>Both</i> |
| <i>Topic Type</i> | | | |
| Computation | 3 | 13 | 10 |
| Concept | 4 | 0 | 6 |
| Proof | 6 | 19 | 16 |
| <i>Difficulty</i> | | | |
| Easy | 6 | 16 | 10 |
| Medium | 4 | 9 | 8 |
| Hard | 3 | 7 | 14 |

In terms of dependency, our system has strong performance for topics that rely heavily on formulae. We attribute this strength to Tangent-L, the underlying MathIR system that is competitive with other state-of-the-art systems for formula retrieval [9]. We further conclude that, in spite of our earlier conclusion that performance improves for lower values of α , setting $\alpha = 0$ —assigning no weight to math terms in a query—would be a poor design decision for our system.

In terms of topic type, we observe that our system is strong at *Computation*-type and *Proof*-type topics, but is particularly weak at *Concept*-type topics when compared with the other two types of topics. With further inspection, we see that none of the *Concept*-type topics have a *Formula*-dependency as shown in Table 11, which might be the reason why our system does not perform as well in that category.

Our system excels at all three levels of difficulty: *Easy*, *Medium*, and *Hard*. We attribute this performance to the even distribution of a significant number of topics relying on formulae (with either a *Formula*-dependency or *Both*-dependency) among the three levels, as observed in Table 11.

Similar conclusions can also be made when comparing the nDCG' of our best-performing submission with that of other baselines, as shown in Table 12.

Finally, we observe that our system might be tuned to boost performance to accommodate its weakness if the type of category were known in advance. For instance, in Table 13 we observe that the overall nDCG' value can be slightly improved if our trained re-ranking model is applied on *Concept*-type topics only. This boosting effect is only essentially distinguishable for the setting of $\alpha = 1.0$ and *not effective* for our recommended setting of $\alpha = 0.1$.¹¹ This leads us to speculate that some advantage might be gained by building a system that adapts to the type of query being posed, instead of seeking a one-size-fits-all solution.

¹¹ This effect might be partially attributable to the fact that the re-ranking model is trained over answers retrieved by Tangent-L under the setting $\alpha = 1.0$, as described in Section 3.3.

Table 12. Category performance of *alpha05-noReRank* and all baselines in nDCG', with the best performance for each category highlighted in bold.

| | <i>alpha05</i> <i>-noReRank</i> | <i>Linked</i> <i>MSE posts</i> | <i>Approach0</i> | <i>TF-IDF</i> <i>+Tangent-S</i> | <i>TF-IDF</i> | <i>Tangent-S</i> |
|-------------------|------------------------------------|-----------------------------------|------------------|------------------------------------|---------------|------------------|
| Overall | 0.345 | (0.279) | 0.247 | 0.248 | 0.204 | 0.158 |
| Dependency | | | | | | |
| Text | 0.273 | (0.286) | 0.240 | 0.319 | 0.291 | 0.163 |
| Formula | 0.446 | (0.260) | 0.251 | 0.227 | 0.136 | 0.214 |
| Both | 0.272 | (0.295) | 0.246 | 0.240 | 0.235 | 0.101 |
| Topic Type | | | | | | |
| Computation | 0.325 | (0.211) | 0.201 | 0.260 | 0.208 | 0.157 |
| Concept | 0.189 | (0.277) | 0.233 | 0.334 | 0.373 | 0.095 |
| Proof | 0.395 | (0.323) | 0.280 | 0.219 | 0.159 | 0.175 |
| Difficulty | | | | | | |
| Easy | 0.361 | (0.293) | 0.269 | 0.264 | 0.188 | 0.193 |
| Medium | 0.306 | (0.293) | 0.179 | 0.189 | 0.174 | 0.102 |
| Hard | 0.357 | (0.248) | 0.277 | 0.278 | 0.250 | 0.161 |

Table 13. Performance of synthetic runs in nDCG' with *Concept*-type topics being re-ranked, as compared with runs with and without a full re-ranking.

| | <i>Full</i> <i>re-ranking</i> | <i>No</i> <i>re-ranking</i> | <i>Concept-only</i> <i>re-ranking</i> |
|----------------------------|----------------------------------|--------------------------------|--|
| alpha10 ($\alpha = 1.0$) | 0.267 | 0.327 | (0.330) |
| alpha05 ($\alpha = 0.5$) | 0.278 | 0.345 | (0.346) |
| alpha02 ($\alpha = 0.2$) | 0.301 | 0.368 | (0.369) |
| alpha01 ($\alpha = 0.1$) | 0.312 | 0.388 | (0.386) |

5 Conclusions and Future Work

We conclude that a traditional math-aware query system remains a viable option for addressing a CQA task specialized in the mathematical domain. In particular, Tangent-L is a math-aware search engine well-suited to retrieve answers to many computational and proof-like questions in the presence of math formulae. We hypothesize that part of the success for all five MathDowers runs results from the search engine having indexed question-answer pairs instead of answers only, thereby providing a context for evaluating the suitability of each answer in serving as an answer to newly posed topic questions.

Nevertheless, several of our initial experimental design decisions turn out to be somewhat disappointing. In the remainder of this section, we share our thoughts regarding room for improvement.

Improvements for re-ranking. The first improvement is with respect to the design and training of the re-ranking model described in Section 3.3. Several avenues are open:

1. We realize in retrospect that we should normalize the scores returned by Tangent-L within each topic: as for many search engines, the scores for one query are not comparable to the scores for another query. However, we have since found that even with proper normalization, the linear regression model does not produce an overall valuable re-ranking.
2. The mock relevance scores used for training the model might not be indicative of assessed relevance, therefore giving a poor model for re-ranking. For future work, now that there are actual assessments from the ARQMath Lab, we could use those for training and avoid the need for mock scores. This approach could be tested by conducting some cross-validation studies.
3. Perhaps using linear regression is not appropriate for this application. An alternate approach is to first transform (mock or actual) assessments into a set of discrete scores, and then treat these scores as categories to be predicted by a classification model, such as a Support Vector Machine. Other approaches to model CQA-type features should also be investigated, including those that have been shown to be successful in the SemEval CQA Challenge series [17–19].

Improvements for topic conversion. Another area for improvement is with respect to the design of the auto extraction algorithm described in Section 3.1. Although the use of our automatic extraction algorithm performs comparably to manual topic conversion, we have not considered constraining the number of keywords and formulae in the algorithm. With a widely fluctuating formula-to-keyword ratio across topics, the performance of our extraction algorithm might be hindered by using any fixed α value for all queries. Preliminary investigations have been unable to establish a correlation between the best value for α and the ratio of the number of keyword terms to the number of terms extracted from

math formulae [8], but this new benchmark might provide additional insights into how to choose a value for α “on the fly” that depends on the number of keyword terms and math terms.

Alternatively, constraining the maximum number of keywords and formulae and the sizes of formulae extracted from a topic description might be a better approach, since it would also constrain the number of keyword and math terms. However this poses a new question on determining the best maximum limit. We believe that research related to Automatic Term Extraction (ATE) in technical domains, or in mathematical domains, might provide valuable insights into our problem.

Acknowledgements

We gratefully acknowledge financial support from the Waterloo-Huawei Joint Innovation Lab and from NSERC, the Natural Science and Engineering Research Council of Canada in support of our project entitled “Searching Documents with Text and Mathematical Content Using a Pen Based Interface.” Prof. Gordon Cormack kindly let us use one of his research machines for indexing the corpus. We also thank the ARQMath Lab organizers (including, notably, Behrooz Mansouri) for the time and effort expended on developing the idea for the Lab and submitting the proposal to CLEF, as well as preparing the corpus, the questions, the manual translation of the topic questions into formulae and keywords, and the relevance assessments. Andrew Kane and anonymous reviewers from other participating lab teams made several valuable suggestions for improving our presentation after reading a draft version of this paper.

References

1. Abacha, A.B., Agichtein, E., Pinter, Y., Demner-Fushman, D.: Overview of the medical question answering task at TREC 2017 LiveQA. In: Voorhees, E.M., Ellis, A. (eds.) Proceedings of the 26th Text REtrieval Conference, TREC 2017. NIST Special Publication, vol. 500-324 (2017)
2. Aizawa, A., Kohlhase, M., Ounis, I.: NTCIR-10 Math Pilot task overview. In: Proceedings of the 10th NTCIR Conference on Evaluation of Information Access Technologies (NTCIR-10). pp. 654–661 (2013)
3. Aizawa, A., Kohlhase, M., Ounis, I., Schubotz, M.: NTCIR-11 Math-2 task overview. In: Proceedings of the 11th NTCIR Conference on Evaluation of Information Access Technologies (NTCIR-11). pp. 88–98 (2014)
4. Białeccki, A., Muir, R., Ingersoll, G.: Apache Lucene 4. In: SIGIR 2012 Workshop on Open Source Information Retrieval. pp. 17–24 (2012)
5. Carlisle, D., Ion, P., Miner, R.: Mathematical markup language (MathML) version 3.0 2nd edition. W3C recommendation (Apr 2014), <http://www.w3.org/TR/2014/REC-MathML3-20140410/>
6. Davila, K., Zanibbi, R.: Layout and Semantics: Combining Representations for Mathematical Formula Search. ACM Reference (2017). <https://doi.org/10.1145/3077136.3080748>

7. Davila, K., Zanibbi, R., Kane, A., Tompa, F.W.: Tangent-3 at the NTCIR-12 MathIR task. In: Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies (NTCIR-12). pp. 338–345 (2016)
8. Fraser, D.J.: Math Information Retrieval using a Text Search Engine. Master’s thesis, Cheriton School of Computer Science, University of Waterloo (2018)
9. Fraser, D.J., Kane, A., Tompa, F.W.: Choosing math features for BM25 ranking with Tangent-L. In: Proceedings of the 18th ACM Symposium on Document Engineering (DocEng 2018). pp. 17:1–17:10 (2018)
10. Gao, L., Yuan, K., Wang, Y., Jiang, Z., Tang, Z.: The math retrieval system of ICST for NTCIR-12 mathir task. In: Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies (NTCIR-12). pp. 318–322 (2016)
11. Guidi, F., Sacerdoti Coen, C.: A survey on retrieval of mathematical knowledge. *Mathematics in Computer Science* **10**(4), 409–427 (2016)
12. Hopkins, M., Le Bras, R., Petrescu-Prahova, C., Stanovsky, G., Hajishirzi, H., Koncel-Kedziorski, R.: SemEval-2019 task 10: Math question answering. In: Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval-2019). pp. 893–899 (June 2019)
13. Kristianto, G.Y., Topić, G., Aizawa, A.: Combining effectively math expressions and textual keywords in Math IR. In: Proceedings of the 3rd International Workshop on Digitization and E-Inclusion in Mathematics and Science 2016 (DEIMS2016). pp. 25–32 (2016)
14. Kristianto, G.Y., Topic, G., Aizawa, A.: MCAT math retrieval system for NTCIR-12 mathir task. In: Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies (NTCIR-12). pp. 323–330 (2016)
15. Lv, Y., Zhai, C.: Lower-bounding term frequency normalization. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM’11). pp. 7–16 (2011)
16. Mansouri, B., Agarwal, A., Oard, D.W., Zanibbi, R.: Finding old answers to new math questions: The ARQMath lab at CLEF 2020. In: Advances in Information Retrieval, Proceedings of the 42nd European Conference on IR Research (ECIR 2005). *Lecture Notes in Computer Science*, vol. 12036, pp. 564–571. Springer (2020)
17. Nakov, P., Hoogeveen, D., Mårquez, L., Moschitti, A., Mubarak, H., Baldwin, T., Verspoor, K.: SemEval-2017 task 3: Community question answering. In: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017). pp. 27–48 (Dec 2018)
18. Nakov, P., Mårquez, L., Magdy, W., Moschitti, A., Glass, J., Randeree, B.: SemEval-2015 task 3: Answer selection in community question answering. In: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015). pp. 269–281 (2015)
19. Nakov, P., Mårquez, L., Moschitti, A., Magdy, W., Mubarak, H., Freihat, A.A., Glass, J., Randeree, B.: SemEval-2016 task 3: Community question answering. In: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). pp. 525–545
20. Olvera-Lobo, M.D., Gutiérrez-Artacho, J.: Question answering track evaluation in TREC, CLEF and NTCIR. In: Advances in Intelligent Systems and Computing, vol. 353, pp. 13–22 (2015)
21. Ounis, I., Amati, G., Plachouras, V., He, B., Macdonald, C., Johnson, D.: Terrier information retrieval platform. In: Advances in Information Retrieval, Proceedings of the 27th European Conference on IR Research (ECIR 2005). *Lecture Notes in Computer Science*, vol. 3408, pp. 517–519. Springer (2005)

22. Pineau, D.C.: Math-aware search engines: Physics applications and overview. CoRR abs/1609.03457 (2016)
23. Růžička, M., Sojka, P., Líška, M.: Math indexer and searcher under the hood: Fine-tuning query expansion and unification strategies. In: Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies (NTCIR-12). pp. 331–337 (2016)
24. Schubotz, M., Meuschke, N., Leich, M., Gipp, B.: Exploring the one-brain barrier: A manual contribution to the NTCIR-12 mathir task. In: Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies (NTCIR-12). pp. 309–317 (2016)
25. Schubotz, M., Youssef, A., Markl, V., Cohl, H.S.: Challenges of mathematical information retrieval in the NTCIR-12 math Wikipedia task. In: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2015). pp. 951–954 (2015)
26. Sojka, P., Líšaka, M.: The art of mathematics retrieval. In: Proceedings of the 11th ACM Symposium on Document Engineering (DocEng 2011). pp. 57–60. ACM Press, New York, New York, USA (2011)
27. Sojka, P., Novotný, V., Ayetiran, F., Lupták, D., Štefánik, M.: Quo Vadis, math information retrieval. Tech. rep. (2019)
28. Thanda, A., Agarwal, A., Singla, K., Prakash, A., Gupta, A.: A document retrieval system for math queries. In: Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies (NTCIR-12). pp. 346–353 (2016)
29. Zanibbi, R., Aizawa, A., Kohlhase, M., Ounis, I., Topić, G., Davila, K.: NTCIR-12 MathIR task overview. In: Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies (NTCIR-12). pp. 299–308 (2016)
30. Zanibbi, R., Davila, K., Kane, A., Tompa, F.W.: Multi-stage math formula search: Using appearance-based similarity metrics at scale. In: Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2016). pp. 145–154 (2016)
31. Zanibbi, R., Oard, D., Agarwal, A., Mansouri, B.: Overview of ARQMath 2020: CLEF Lab on answer retrieval for questions on math. In: Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum (2020)
32. Zhong, W., Rohatgi, S., Wu, J., Giles, C.L., Zanibbi, R.: Accelerating substructure similarity search for formula retrieval. pp. 714–727 (2020)

Appendix

Below is the list of formulas and keywords converted from the given topics using the method described in Section 3.1. Among the 98 topics submitted for each run, 77 were (partially) assessed for the CQA task. The topics with no assessments are A.2, A.6, A.22, A.25, A.34, A.46, A.57, A.64, A.70, A.71, A.73, A.76, A.81, A.82, A.84, A.91, A.92, A.94, A.95, A.97 and A.100, marked with a dagger (†) below.

Topic List of formulas and keywords

- A.1 $[c, f(x) = \frac{x^2+x+c}{x^2+2x+c}, [-1, -\frac{1}{3}], f(x) = \frac{x^2+x+c}{x^2+2x+c}, f(x), [-1, -\frac{1}{3}],$ Finding, value, range, rational, function, does, contain, calculate, range, rational, function, reverse, across, this, find, value, range, does, contain, functions]

- A.2† $[f'(x) = f(x + 1), \frac{df}{dx} = f(x + 1), \text{Solving, differential, equations, form, solve, differential, equations, form, ordinary, differential, equations, ordinary-differential-equations}]$
- A.3 $[\sqrt{5}, 10^{-10}, \sqrt{5}, 10^{-10}, \sqrt{5}, f(x), [a, b], f(a), f(b), \text{Approximation, correct, resolve, problem, Find, approximation, correct, using, bisection, has, placed, function, sure, go, function, Mathematica, given, calculations, function, interval, opposite, signs, tolerance, number, iterations, numerical, methods, algorithms, bisection, numerical-methods}]$
- A.4 $[\sum_{k=0}^n \binom{n}{k} k, n2^{n-1}, \text{compute, this, combinatoric, sum, sum, know, result, know, does, one, even, sum, like, this, has, binomial, coefficients, combinatorics, number, theory, summation, proof, explanation, number-theory, proof-explanation}]$
- A.5 $[P((2)|(1)) = \frac{P((2) \cap (1))}{P((1))} = \frac{P((2))P((1))}{P((1))} = P((2)), \text{family, has, two, Given, one, boy, probability, boys, family, has, two, Given, one, boy, probability, boys, was, this, question, using, conditional, probability, event, first, boy, event, second, probability, second, boy, given, first, boys, formula, since, second, boy, does, depend, first, detailed, solution, correct, probability, proof, verification, conditional, probability, proof-verification, conditional-probability}]$
- A.6† $[5^{133} \pmod{8}, 5^n \pmod{8} = 5, 5^{133} \pmod{8} = 5, \text{calculate, mod, number, big, exponent, find, noticed, lead, say, know, prove, prove, this, case, find, solution, algebra, precalculus, arithmetic, algebra-precalculus}]$
- A.7 $[\frac{1110-1}{100}, 1110 - 1, 1110 - 1 = x \pmod{100}, 112 \equiv 21 \pmod{100}, (112)2 \equiv (21)2 \pmod{100}, 114 \equiv 441 \pmod{100}, 114 \equiv 41 \pmod{100}, (114)2 \equiv (41)2 \pmod{100}, 118 \equiv 1681 \pmod{100}, 118 \equiv 81 \pmod{100}, 118 \times 112 \equiv (81 \times 21) \pmod{100}, 1110 \equiv 1701 \pmod{100} \implies 1110 \equiv 1 \pmod{100}, 1110-1 \equiv (1-1) \pmod{100} \implies 1110 - 1 \equiv 0 \pmod{100}, x = 0, 1110 - 1, \text{Finding, remainder, using, modulus, divided, solve, term, mod, tried, mod, mod, mod, mod, mod, mod, mod, mod, mod, value, divisible, this, approach, long, time, competitive, exam, math, contest, without, using, process, determine, remainder, above, problem, very, helpful, advance, elementary, number, theory, modular, arithmetic, divisibility, alternative, proof, elementary-number-theory, modular-arithmetic, alternative-proof}]$
- A.8 $[\lim_{n \rightarrow \infty} \sqrt[n]{\frac{(27)^n (n!)^3}{(3n)!}}, \lim_{n \rightarrow \infty} \sqrt[n]{\frac{(27)^n (n!)^3}{(3n)!}}, \text{finding, value, Finding, value, try, solve, help, limits}]$
- A.9 $[\sum_{n=0}^N nx^n, \sum_{i=0}^n i^2 = \frac{(n^2+n)(2n+1)}{6}, \text{this, series, need, write, series, form, does, involve, summation, notation, example, Does, anyone, idea, this, multiple, ways, including, using, generating, functions, sequences, series, sequences-and-series}]$
- A.10 $[\int_0^\infty \frac{\sin x}{x^a}, \int_0^\infty \frac{\sin x}{x^a}, \text{Find, values, improper, integral, sin, converges, Find, values, improper, integral, sin, converges, expand, using, series, expansion, improper, integrals, improper-integrals}]$

- A.11 $[R^3, u = (a, b, c), v = (d, e, f), R^2, R^3, u = (a, b), v = (d, e), R^2, R^3, \Phi(u, v) = (f(u), g(v)), R^2, R^3, R^2, \Phi(u, v) = (2u \cos v, u \sin v),$ cross, product, dimensions, math, book, using, states, cross, product, two, vectors, defined, direction, resultant, determined, curling, fingers, vector, pointing, direction, cross, product, cross, product, defined, Is, degenerate, case, cross, product, like, this, type, determinant, instance, parameterization, needed, calculate, examples, book, calculating, determinate, cos, sin, multivariable, calculus, vectors, multivariable-calculus]
- A.12 $[(1 + i\sqrt{3})^{1/2},$ Finding, roots, complex, number, was, solving, practice, problems, this, question, It, find, roots, sketch, linear, algebra, complex, numbers, polar, coordinates, linear-algebra, complex-numbers, polar-coordinates]
- A.13 $[\int_a^b f(x)dx + \int_{f(a)}^{f(b)} f^{-1}(x)dx ? , \int_a^b f(x)dx + \int_{f(a)}^{f(b)} f^{-1}(x)dx ? , bf(b) - af(a),$ expression, expression, answer, answer, calculus]
- A.14 $[y = xy' + \frac{1}{2}(y')^2, \frac{1}{2}y'(2x + y') = y, x^2 + y = t,$ Help, solving, first, order, differential, equation, first-order, first, order, differential, equation, this, find, way, solve, use, derivate, idea, first-order, ordinary, differential, equations, ordinary-differential-equations]
- A.15 $[\sum_{i=1}^n ix^{i-1}, 1 + 2x + 3x^2 + 4x^3 + 5x^4 + \dots + nx^{n-1} + \dots, x \neq 1, |x| < 1, S_n, S_2 = 1 + x + x^2 + x^3 + x^4 + \dots, \frac{d(S_2)}{dx} = S_1, |x| < 1, S_2, \frac{1-x^n}{1-x} = \frac{1}{1-x}, S_1 = \frac{d(S_2)}{dx} = \frac{d(\frac{1-x}{1-x})}{dx} = \frac{1}{(1-x)^2},$ Derive, sum, series, need, find, partial, sums, finally, sum, tried, series, source, series, sum, geometric, progression, this, answer, sequences, series, convergence, summation, power, series, sequences-and-series, power-series]
- A.16 $[\int_0^1 \frac{\ln(1+x)\ln(1-x)}{1+x} dx, \int_0^1 \frac{\ln(1+x)\ln(1-x)}{1+x} dx, I(a, b) = \int_0^1 \frac{\ln(1-ax)\ln(1+bx)}{1+x} dx, \frac{d^2 I(a,b)}{da db},$ Finding, Calculate, My, try, Let, compute, happy, see, ideas, order, kill, this, integral, integration, sequences, series, definite, integrals, closed, form, sequences-and-series, definite-integrals, closed-form]
- A.17 $[\int_{x=0}^{\infty} \frac{\sin(x)}{x}, \frac{e^{iz}}{z}, \int_{x=0}^{\infty} \frac{\sin(x)}{x}, f(z) = \frac{e^{iz}}{z}, \Gamma = \gamma_1 + \gamma_R + \gamma_2 + \gamma_\epsilon, \gamma_1(t) = t, t \in [i\epsilon, iR], \gamma_R(t) = Re^{it}, t \in [-\frac{\pi}{2}, \frac{\pi}{2}], \gamma_2(t) = t, t \in [-iR, -i\epsilon], \gamma_\epsilon(t) = \epsilon e^{it}, t \in [-\frac{\pi}{2}, \frac{\pi}{2}], \frac{\sin(x)}{x}, \int_{\gamma_\epsilon} f = -i\pi, \epsilon \rightarrow 0, \int_{\gamma_R} f = 0, R \rightarrow \infty,$ Calculate, sin, function, calculate, sin, function, using, closed, path, use, sin, even, function, has, anti, derivative, integral, closed, path, managed, show, show, Help, complex, analysis, improper, integrals, complex-analysis, improper-integrals]

- A.18 $\left[\lim_{n \rightarrow \infty} \frac{[(n+1)(n+2)\cdots(n+n)]^{1/n}}{n}, \lim_{n \rightarrow \infty} \frac{[(n+1)(n+2)\cdots(n+n)]^{1/n}}{n} \right]$,
 $\log 2 + n \log \frac{n}{n+1}, \frac{2}{e}, \frac{4}{e}, 2/e$, Evaluate, Evaluate, using, Stolz, know,
many, question, like, this, solve, using, Stolz, method, log, applied, Stolz,
log, log, answer, answer, help, Edit, On, log, log, log, log, log, log, log,
log, log, log, Cesàro-Stolz, Cesàro-Stolz, Cesàro-Stolz,, sequences, series,
sequences-and-series]
- A.19 $[p^4 - 1, p^4 - 1, 7^4 - 1, 2^4 * 3 * 5 * 2, 11^4 - 1, 2^4 * 3 * 5 * 61, 2^4 * 3 * 5,$
 $p^4 - 1, (p^2 + 1)(p - 1)(p + 1), 2^4, 16n + x$, Greatest, common, factor, was,
find, greatest, common, factor, primes, First, value, has, divisors, has,
divisors, has, prove, divisible, even, integers, know, prove, divisibility,
since, check, numbers, prove, divisibility, assigning, divisibility, greatest,
common, divisor, greatest-common-divisor]
- A.20 $[n \in \mathbb{N} \setminus \{41\}, \phi(n) = 40, n \in \mathbb{N}, \phi(n) = 40, \phi, n = 41, n's$, Calculate,
looking, Euler, Totient, Function, found, one, namely, calculate, Euler-
Totient, totient, function, totient-function]
- A.21 $[9^{9^{\cdots 9}}, 9^{9^{\cdots 9}}, 9^{9^{\cdots 9}} \equiv x \pmod{100}, 0 \leq x \leq 100, 9^{9^{\cdots 9}}$ (nine9s) =
 $9^a, 9^a \pmod{100}, a \pmod{\phi(100)}, \phi(100) = 40, a = b \pmod{40},$
 $9^{9^{\cdots 9}}$ (eight9s) = $9^b, 9^b \pmod{40}, b \pmod{\phi(40)}, \phi(40) = 16, b =$
 $c \pmod{16}, 9^{9^{\cdots 9}}$ (seven9s) = $9^c, 9^c \pmod{16}, c \pmod{\phi(16)}, \phi(16) = 8,$
 $c \pmod{8}, 9 = 1 \pmod{8}, c = 1 \pmod{8}$, Finding, last, two, digits, nine,
continuing, learning, modular, arithmetic, confused, this, question, Find,
last, two, digits, nine, phi, function, used, this, problem, far, this,
In, order, know, need, know, In, order, know, need, know, In, order,
know, need, know, need, find, like, made, along, way, lot, back, order,
value, last, two, anyone, help, this, number, theory, modular, arithmetic,
number-theory, modular-arithmetic]
- A.22† $[d, d + 1, d + 2 \dots = N, N, d, d + 1, d + 2 \dots, N = 30, Ans = 3, d_1 =$
 $4; d_2 = 6; d_3 = 8, d_1 : 4 + 5 + 6 + 7 + 8 = 30, (\sum(d + n) - \sum(d - 1)),$
Find, number, satisfies, challenge, cat, trip, walk, way, sum, given, many,
ways, Example, Edit, way, see, many, subsets, way, sequences, series,
number, theory, sequences-and-series, number-theory]
- A.23 $[2^7 \cdot 3^8 \cdot 5^2 \cdot 7^{11}, 2^3 \cdot 3^4 \cdot 5, \cdot x = 2^3 \cdot 3^4 \cdot 5x$, find, product, two, integers,
greatest, common, divisor, least, common, multiple, this, question, help,
solve, tried, assuming, Gcd, product, factors, prime, numbers, great-
est, common, divisor, least, common, multiple, prime-numbers, greatest-
common-divisor, least-common-multiple]
- A.24 $[\sqrt{2i - 1}?, \sqrt{2i - 1}?, 2i - 1 = (a + bi)^2, a^2 + 2abi - b^2, a^2 - b^2 = -1, 2ab =$
 $2, a^2 = b^{-2}, b^{-2} - b^2 = -1, -b^4 + 1 = -1, b^4 = 2, b = \sqrt[4]{2}, \sqrt{2i - 1}$, Is,
this, only, way, evaluate, work, solve, way, algebra, precalculus, algebra-
precalculus]

- A.25† [$P(x^2 + 1) = (P(x))^2 + 1$, $P(0)$, $P(x^2 + 1) = (P(x))^2 + 1$, $P(x)$, $P(0) = 0$, $P(1) = 1$, $P(2) = 2$, $P(5) = 5$, $P(26) = 26$, $P(677) = 677$, $P(x) = x$, $y = P(x)$, $y = x$, $P(0) = 2$, $P(x) = (x^2 + 1)^2 + 1$, $P(0) = 3$, $\lim_{x \rightarrow \infty} \log_x P(x)$, $P(x)$, $P(0)$, $P(x)$, polynomial, polynomial, Let, points, log, does, converge, So, values, makes, polynomial, polynomials]
- A.26 [$\int_0^\infty \frac{\sin x}{x} dx$, $\int_0^x \frac{\sin x}{x} dx$, solve, indefinite, integral, using, Taylor, series, trying, show, integral, convergent, sin, My, first, taylor, series, sin, next, step, real, analysis, calculus, integration, taylor, expansion, riemann, integration, real-analysis, taylor-expansion, riemann-integration]
- A.27 [$e^{3i\pi/2}$, $e^{\pi i} = -1$, $e^{3\pi i/2}$, $(e^{\pi i})^{3/2}$, $(\sqrt{-1})^3 = i^3 = -i$, $\sqrt{(-1)^3} = \sqrt{-1} = i$, value, solving, value, know, confused, right, answer, evaluate, two, possible, answers, this, one, correct, answer, going, complex, numbers, exponentiation, complex-numbers]
- A.28 [$\sin(18^\circ) = \frac{a+\sqrt{b}}{c}$, $a + b + c$, $\sin(18) = \frac{a+\sqrt{b}}{c}$, $a + b + c$, $\sin(18)$, $\frac{x}{z}$, $x = a + \sqrt{b}$, $z = c$, $y = \sqrt{c^2 - (a + \sqrt{b})^2}$, $\cos(18) = \frac{y}{z} = \frac{\sqrt{c^2 - (a + \sqrt{b})^2}}{c}$, $b = (c \sin(18) - a)^2 = c^2 \sin^2(18) - 2ac \sin(18) + a^2$, $\sin(18) = \frac{-1 + \sqrt{5}}{4}$, A, B, C , $A \sin(18)^2 + B \sin(18) + C = 0$, $\sin(18)$, $Ax^2 + Bx + C$, $a = -B$, $b = B^2 - 4AC$, $c = 2A$, $\sin(18) = (-1 + \sqrt{5})/4$, sin, sin, form, sin, right, triangle, sides, front, corner, angle, degrees, hypotenuse, find, cos, found, sin, sin, sin, solution, says, sin, intuition, find, sin, sin, sin, root, Totally, This, question, prove, part, solution, algebra, precalculus, trigonometry, euclidean, geometry, contest, math, algebra-precalculus, euclidean-geometry, contest-math]
- A.29 [$3 + 2i$, $-4i$, $i = \sqrt{-1}$, $\frac{5i}{\infty} = 0$, $\frac{5}{\infty} \cdot \frac{\sqrt{-1}}{\infty} = 0 \cdot 0 = 0$, $\frac{5x}{\infty} = 0$, $\frac{5}{\infty} \cdot \frac{x}{\infty} = 0 \cdot 0 = 0$, Dividing, Complex, Numbers, Infinity, My, PreCalculus, properties, limits, before, test, stated, real, number, divided, infinity, equals, This, whether, complex, number, divided, infinity, equal, This, completely, was, theoretical, calculation, knowing, calculated, complex, number, since, using, properties, utilized, real, numbers, state, since, Is, this, theoretical, calculation, correct, concept, this, algebra, precalculus, limits, complex, numbers, infinity, algebra-precalculus, complex-numbers]
- A.30 [$a^3 + b^3 + c^3 - 3abc$, $a^3 + b^3 + c^3 - 3abc$, (a) 1, (b) 0, (c) -1, (d) -2, a, b, e^x , Find, binomial, theorem, Find, help, solve, this, added, It, expansion, know, use, binomial, theorem, binomial-theorem]
- A.31 training data

- A.32 [$Empty(x) \iff \nexists y(y \in x)$, $Empty(x) \iff \exists y(y \in x)$, definitions, axioms, very, elementary, definition, first, order, logical, example, say, Define, Is, definition, axiom, call, definitional, this, one, place, predicate, symbol, Empty, new, among, listed, primitives, say, Zermelo, has, only, identity, membership, primitive, So, stating, definitions, effect, stating, axioms, characterizing, primitive, definitional, axioms, complete, reference, specified, set, symbols, Is, correct, case, why, call, axiom, state, mean, why, say, example, Definitional, axiom, terminology, definition, first, order, logic, axioms, first-order-logic]
- A.33 [$f(t, x)$, $\frac{\partial^3 f}{\partial t^3}$, $\frac{\partial^3 f}{\partial x^3}$, Physical, meaning, significance, third, derivative, function, Given, physical, quantity, represented, function, meaning, third, derivative, physics]
- A.34† [$a, b > 0$, $a \uparrow b$, $a \downarrow\downarrow b$, $a \downarrow\downarrow\downarrow b$, $a \downarrow^n b$, $n \geq 3$, $a \downarrow\downarrow b = a + 1, b + 1$, $a \uparrow^n b$, $n \leq 0$, $a \downarrow\downarrow b = b + 1$, $a \downarrow b = a + b - 1$, $a > b$, $a \downarrow b$, $a \downarrow\downarrow\downarrow b$, $a > b$, $a \downarrow\downarrow b = b + 1 + \lfloor \frac{a}{b} \rfloor$, $a \downarrow\downarrow\downarrow b$, Extending, Knuth, non, positive, values, up-arrow/hyperoperations, non-positive, So, idea, extend, Knuth, arrow, notation, included, zero, negative, It, normally, defined, basically, hyperoperation, sequence, only, try, go, backwards, trivial, extension, letting, arrows, represent, negative, arrows, why, heck, coming, expression, does, Alternatively, way, defining, zero, arrows, does, So, question, Is, extension, Knuth, arrow, notation, exists, Edit, this, question, initially, was, correct, So, example, extension, modified, question, An, extension, define, satisfies, recursive, definition, Edit, turns, correct, this, imply, example, close, need, exception, case, does, show, evaluating, need, defined, try, extend, instance, abuse, case, allowed, let, finding, intractable, result, up-arrow, up-arrow, hyperoperation, ackermann, function, ackermann-function]
- A.35 [$\int e^{x^2} dx$, $\int e^{2x} dx$, does, function, antiderivative, know, this, question, sound, why, ca, write, does, antiderivative, In, light, this, question, sufficient, conditions, function, need, careful, examination, function, say, does, antiderivative, way, see, function, right, say, does, antiderivative, integration]
- A.36 [$\neg P$, true, $\neg P \rightarrow A_1 \rightarrow \dots \rightarrow A_n \rightarrow P$, $\neg P \rightarrow P \iff \neg(\neg P) \vee P \iff P$, $\neg P$, $\neg P$, $\neg P$, Proof, contradiction, status, initial, assumption, proof, complete, First, like, say, looked, answers, specific, question, found, existing, question, Say, need, prove, statement, method, Assuming, holds, using, list, statements, proven, hold, derived, proof, arrive, list, proven, statements, was, proofs, contain, lines, contradiction, proves, initial, assumption, was, holds, initial, assumption, proven, FALSE, why, sure, derived, holds, particular, holds, On, hand, derived, assumption, explain, why, this, type, argument, used, logic, proof, writing, proof-writing]

- A.37 $[f \circ g = g \circ f, f^{-1} \neq g, f \neq \text{id} \neq g, f \circ g = g \circ f?, f(x) = 2x, g(x) = 3x, f: \mathbb{R} \setminus \{1\} \rightarrow \mathbb{R}, f(x) = 2x/(1-x), g(x) = f^{-1} \circ g \circ f = \frac{g(\frac{2x}{1-x})}{2+g(\frac{2x}{1-x})}, g_0, g_{n+1} \mapsto f^{-1} \circ g_n \circ f, g_0, \text{Non, trivial, examples, real, valued, functions, inverses, identity, linear, Examples, trivial, this, given, function, one, go, example, function, commutes, this, similar, fixed, point, iteration, defined, need, function, function, finding, fixed, point, possible, strategy, ca, find, real-valued, real, analysis, functional, analysis, functions, real-analysis, functional-analysis}]$
- A.38 $[a, b, q, r : a = bq + r, \text{Choose}, q : qb \leq a, \text{Uses, Axiom, Choice, first, year, maths, student, drift, years, ZFC, axioms, first, time, college, stuff, far, nowhere, near, ZFC, terms, happened, use, axiom, choice, time, module, even, know, name, example, proof, non, negative, integers, exist, integers, known, restrictions, proof, like, this, largest, integer, Is, axiom, choice, allows, this, simple, important, step, couple, questions, name, simple, proofs, theorems, results, etc, axiom, choice, essential, read, has, long, topic, dispute, mathematicians, even, people, accept, alternative, axiomatic, systems, work, equally, well, without, needing, first-year, non-negative, set, theory, set-theory}]$
- A.39 $[2018^{2019}, 2019^{2018}, 2019 \log(2018), 2018 \log(2019), \log 2019 > \log 2018, 2019^{2018}, \text{know, value, logs, sides, log, log, know, log, log, does, this, mean, one, algebra, precalculus, logarithms, algebra-precalculus}]$
- A.40 $[a_1x_1 + a_2x_2 + a_3x_3 + \dots + a_nx_n =, f(x+y) = f(x) + f(y), f(cx) = cf(x), \text{meaning, term, linear, called, linear, equation, represents, equation, line, dimensional, So, linear, comes, word, line, higher, power, graph, functions, straight, called, linear, differential, equation, derivatives, power, equal, similar, above, definition, linear, function, called, linear, In, this, definition, linearity, function, does, word, linear, means, does, relate, straight, line, Finally, does, term, linear, means, case, linear, vector, spaces, reference, straight, line, So, whether, linear, word, used, different, contexts, Does, different, meaning, different, situation, Or, linearity, refers, relation, straight, line, At, Least, explain, linearity, function, linear, vector, space, relate, equation, line, linear, algebra, linear-algebra}]$
- A.41 $[n \leq m, \sum_{r=1}^n (-1)^{(n-r)} \binom{n}{r} (r)^m, \text{Confusion, find, number, onto, functions, two, sets, given, In, book, given, two, finite, sets, containing, elements, number, onto, functions, well, ca, combinations, combinatorics, functions, combinations}]$

- A.42 $[x^2 + 1 = 0, \sqrt{-1}, \text{simple, physical, situation, complex, numbers, emerge, naturally, trying, teach, middle, emergence, complex, numbers, motivate, this, this, mean, sort, real, world, problem, people, trying, solve, led, realize, needed, extend, real, numbers, instance, Greeks, forced, irrational, numbers, pure, mathematical, reasons, length, diagonal, square, unit, length, irrational, this, kind, geometrical, situation, dealing, similar, situation, lead, complex, numbers, terms, say, try, solve, equation, physical, abstract, sort, answer, like, object, define, certain, properties, turn, consistent, important, entirely, satisfying, complex, numbers, physics, applications, complex-numbers}]$
- A.43 $[\sum_{n \geq 1} \frac{1}{n^2+1} = \frac{\pi \coth \pi - 1}{2}, \sum_{n \geq 1} \frac{1}{n^2+1} = \frac{\pi \coth \pi - 1}{2}, S = \sum_{n \geq 1} \frac{1}{n^2+1}, x = \pi, e^\pi = \frac{\sinh \pi}{\pi} + \frac{2 \sinh \pi}{\pi} S, S = \frac{\pi e^\pi}{2 \sinh \pi} - \frac{1}{2}, \text{Prove, trying, prove, Letting, Fourier, series, exponential, function, sinh, sinh, cos, sin, Plugging, sinh, sinh, cos, sin, sinh, sinh, sinh, sinh, sinh, nowhere, near, prove, identity, real, analysis, sequences, series, real-analysis, sequences-and-series}]$
- A.44 $[A, B \in \mathcal{M}_{2 \times 2}(\mathbb{Q}), AB, (\mathcal{M}_{2 \times 2}(\mathbb{Q}), \times), AB, \text{finite, order, show, has, infinite, order, Let, group, Let, order, Let, order, Show, has, infinite, only, reasoning, possible, contradiction, tried, before, concrete, matrices, group, theory, cyclic, groups, group-theory, cyclic-groups}]$
- A.45 $[\sin(x), \sin(2x), \sin(3x), \dots, \sin(nx), \mathbb{R}, \sin(x), \sin(2x), \sin(3x), \dots, \sin(nx), \mathbb{R}, W \neq 0, W \neq 0, \text{prove, sin, sin, sin, sin, independent, Prove, sin, sin, sin, sin, independent, trial, know, trivial, solution, trivial, show, sin, sin, sin, sin, cos, cos, cos, cos, sin, sin, sin, sin, cos, cos, cos, cos, looks, like, Vandermonde, matrix, cant, prove, ordinary, differential, equations, ordinary-differential-equations}]$
- A.46† $[f, [0, 1], \int x^k f(x) dx = 0, f = 0 \text{ a.e.}, [0, 1], \int_0^1 x^k f(x) dx = 0, k = \{0, 1, 2, 3, \dots\}, f = 0, \int_0^1 f^2(x) dx = 0, \epsilon > 0, M = \sup_{x \in [0, 1]} |f(x)|, [0, 1], \int_0^1 x^k f(x) dx = 0, k = \{0, 1, 2, 3, \dots\}, f = 0, [0, 1], \int_0^1 x^k f(x) dx = 0, k = \{0, 1, 2, 3, \dots\}, f = 0, f = 0 \text{ a.e.}, \text{Lebesgue, integrable, function, satisfies, prove, indefinitely, differentiable, real, valued, function, satisfies, prove, My, prove, this, assertion, prove, approximation, theorem, find, polynomial, sup, It, need, condition, indefinitely, differentiable, continuous, hold, My, question, Lebesgue, integrable, real, valued, function, satisfies, prove, except, set, measure, TRUE, Riemann, integrable, real, valued, function, satisfies, prove, except, set, measure, EDIT, prove, prove, fourier, coefficient, equal, cos, cos, proof, complete, real, analysis, real-analysis}]$

- A.47 $[p, 0 < r < p - 1, q, rq \equiv 1 \pmod p, 0 < r < p - 1, rq \equiv 1 \pmod p, rq - 1 = kp, qr - kp = 1, \sum_{r=1}^{p-2} r = \frac{(p-2)(p-1)}{2} = p\frac{p-3}{2} + 1 \equiv 1 \pmod p, (p-1)! \equiv -1 \pmod p, (p-1)! + 2 \equiv 1 \pmod p, (p+1), (p+1)(p-1)! = -(p+1) \pmod p, (p+1)(p-1)! = 1 \pmod p, \text{Prove, given, prime, exists, mod, Prove, given, prime, exists, mod, only, taken, one, number, theory, course, years, this, popped, computer, science, class, was, assuming, this, proof, elementary, since, current, class, algorithm, cours, basic, tried, look, couple, approaches, reverse, engineer, arrive, conclusion, need, little, manipulation, looks, ca, see, sum, mod, looks, good, know, final, Wilson, Lagrange, vaguely, this, theorem, was, looking, old, book, was, see, arrived, prime, mod, multiplier, built, factorial, expression, was, adding, side, mod, dead, end, sure, was, multiplying, mod, results, mod, multiple, sure, valid, elementary, number, theory, proof, verification, elementary-number-theory, proof-verification}]$
- A.48 $[x, y \geq 0, (x+y)^k \geq x^k + y^k, k \geq 1, x, y \geq 0, (x+y)^k \geq x^k + y^k, k \in \mathbb{R}_{\geq 1}, x \leq x+y, x^k \leq (x+y)^k, k \geq 1, y^k \leq (x+y)^k, x^k + y^k \leq 2(x+y)^k, x, y \geq 0, f(k) = (x+y)^k, g(k) = x^k + y^k, (x+y)^r \geq x^r + y^r, r \in \mathbb{Z}^+, k = 1, (x+y)^r > x^r + y^r, r \geq 2, m \notin \mathbb{Z}, f(m) = g(m), k = 1, (x+y)^r > x^r + y^r, r \geq 2, k = 1, \text{showing, trying, show, So, far, tried, things, since, sides, inequality, positive, inequality, hold, Adding, inequalities, this, close, course, Alternatively, was, fix, let, let, show, using, binomial, theorem, show, intersect, only, better, proving, statement, since, positive, integers, real, number, property, exist, since, violate, only, continuity, positive, integers, course, this, dependent, showing, intersect, only, running, low, real, analysis, functions, inequality, real-analysis}]$
- A.49 $[\binom{2n}{n} = \sum_{k=0}^n \binom{n}{k}^2, (1+x)^{2n} = [(1+x)^n]^2, x^n, \binom{2n}{n}, \sum_{k=0}^n \binom{n}{k}^2, \text{Is, simple, combinatoric, interpretation, this, identity, across, exercise, prove, identity, exercise, It, use, prove, identity, expressions, identity, coefficients, expansions, expressions, course, number, was, whether, equivalent, counting, interpretation, It, clear, number, ways, half, elements, set, this, possible, interpret, equivalently, equivalent-counting, combinatorics, binomial, coefficients, binomial-coefficients}]$
- A.50 $[\sum \frac{1}{n^{2+\cos n}}, \sum \frac{1}{n^{2+\cos n}}, \epsilon, A_\epsilon, n, |2 + \cos n| \leq 1 + \epsilon, (n \in A_\epsilon \iff -1 \leq \cos n \leq -1 + \epsilon), A_\epsilon, \text{Divergent, series, cos, Show, cos, divergent, My, main, problem, infinitely, small, positive, real, number, define, set, cos, cos, divergence, come, sum, idea, handle, this, real, analysis, integration, sequences, series, analysis, real-analysis, sequences-and-series}]$
- A.51 $[\sum_{r=0}^n \binom{n+r}{r} \frac{1}{2^r} = 2^n, x^n, \text{Sum, series, binomial, coefficients, Prove, try, coefficient, solve, binomial, coefficients, binomial-coefficients}]$

- A.52 [$\forall n \in \mathbb{N}, \exists m \in \mathbb{N}, m > n, m, n_1, n_2, \dots, n_k \in \mathbb{N}, n = n_1 n_2 \dots n_k + 1, n_1, n_2, \dots, n_k, a, b \in \mathbb{N}, \mathbb{Z}^+, a = qb + r, 0 \leq r < q, k \in \mathbb{N}, n_1, n_2, \dots, n_k \geq 1, \forall i, n_i \nmid n = n_1 n_2 \dots n_k + 1, \exists n \in \mathbb{N}, \forall m \in \mathbb{N}, m \leq n$, Prove, prime, two, parts, Prove, least, divisible, numbers, Prove, truth, negation, leads, Use, theorem, exist, unique, quotient, remainder, part, given, show, set, sure, go, part, know, negation, prime, sure, proof, theorem, proof, writing, prime, numbers, proof-writing, prime-numbers]
- A.53 [$g \in G, gh = 1, hg = 1, AB = 1 \not\Rightarrow BA = 1, AB = 1 \Rightarrow BA = 1$, Show, one, sided, inverse, square, matrix, TRUE, inverse, one-sided, know, group, element, does, mean, In, case, matrices, linear, maps, vector, spaces, TRUE, This, happens, square, matrices, case, even, form, group, multiplication, restrict, square, matrices, simple, proof, this, avoids, chasing, entries, makes, use, simply, vector, space, structure, linear, transformations, In, prove, this, this, imply, group, one, sided, two, sided, inverses, has, infinite, since, finite, group, finite, dimensional, representation, one-sided(but, two-sided), linear, algebra, group, theory, linear-algebra, group-theory]
- A.54 [$P(N) = (S|S \subseteq N), P(N) = (S|S \subseteq N)$, using, diagonal, argument, show, uncountable, tips, solutions, this, one, using, diagonal, argument, show, uncountable, discrete, mathematics, elementary, set, theory, discrete-mathematics, elementary-set-theory]
- A.55 [$\frac{1}{\sqrt{-1}} = \sqrt{-1}, \sqrt{-1}, (\sqrt{-1})^2 = -1, \frac{1}{\sqrt{-1}} = \sqrt{\frac{1}{-1}} = \sqrt{-1}$, calculation, set, new, number, property, write, know, this, correct, result, missing, complex, numbers, definition, complex-numbers]
- A.56 [$\exists p$ (pisprime $\rightarrow \forall x$ (xisprime)), curious, logical, formula, involving, prime, numbers, Let, set, natural, Is, formula, TRUE, FALSE, know, answer, this, question, shortest, way, arrive, conclusion, using, deduction, system, logic, first, order, logic, first-order-logic]
- A.57† [$\mathbb{R}^n, f : B \rightarrow \mathbb{R}^m, f^{-1}, f(B), f(X), \mathbb{R}^n, f : X \rightarrow \mathbb{R}^m, f^{-1} : f(X) \mapsto X, f(X)$, Preimage, continuous, one, one, function, connected, domain, continuous, one-to-one, know, given, compact, subset, continuous, injective, one, one, function, continuous, This, TRUE, know, image, connected, subset, connected, continuous, let, connected, non, compact, subset, continuous, injective, one, one, trying, counterexample, mapping, continuous, parametrized, example, advance, (one-to-one), (non-compact), (one-to-one), real, analysis, general, topology, analysis, continuity, metric, spaces, real-analysis, general-topology, metric-spaces]
- A.58 [$3 \arcsin \frac{1}{4} + \arccos \frac{11}{16} = \frac{\pi}{2}, 3 \arcsin \frac{1}{4} + \arccos \frac{11}{16} = \frac{\pi}{2}$, Prove, arcsin, help, this, exercise, know, prove, answer, fully, Exercise, Prove, arcsin, trigonometry]
- A.59 [$\sum_{d|n} \phi(d) = n, \phi(n), \sum_{d|n} \phi(d) = n, n = \prod_{k=1}^m p_k^{\alpha_k}, d = \prod_{k=1}^m p_k^{\beta_k}, 0 \leq \beta_k \leq \alpha_k$, Multiple, proofs, looking, multiple, proofs, statement, denotes, Euler, totient, one, unique, factorisation, theorem, align, group, theory, number, theory, alternative, proof, big, list, group-theory, number-theory, alternative-proof, big-list]

- A.60 $[a_n = \left(1 - \frac{1}{\sqrt{2}}\right) \dots \left(1 - \frac{1}{\sqrt{n+1}}\right), n \geq 1, \lim_{n \rightarrow \infty} a_n, \frac{1}{\sqrt{\pi}}, a_n, (0.293) * (0.423) * (0.5) * (0.553) * (0.622) * (0.647) * (0.667) * \dots, (D), a_n, \text{Limiting, value, sequence, tends, infinity, Let, equals, does, exist, equals, equals, My, Approach, particular, direction, procedure, find, value, tends, So, tried, like, this, simple, way, substitute, values, trying, find, limiting, value, So, value, tending, option, tried, like, this, find, value, converges, tends, infinity, help, procedure, solve, this, question, calculus, sequences, series, limits, products, sequences-and-series}]$
- A.61 $[i, j \in \mathbb{N}, n = 3i + 5j, n \geq 8, i, j \in \mathbb{N}, n = 3i + 5j, n \geq 8, n = 8 \implies 8 = 3 \cdot 1 + 5 \cdot 1, n = 9 \implies 9 = 3 \cdot 3 + 5 \cdot 0, n = 10 \implies 10 = 3 \cdot 0 + 5 \cdot 2, n = h, n = h + 1, k + 1 = 3i + 5j, \text{exists, Prove, exists, hard, time, this, exercise, trying, prove, induction, Basis, step, Induction, step, TRUE, TRUE, So, know, proving, elementary, number, theory, discrete, mathematics, induction, diophantine, equations, elementary-number-theory, discrete-mathematics, diophantine-equations}]$
- A.62 $[\mathbb{Q}, \mathbb{Z}, |\mathbb{Q}| = |\mathbb{Z}|, \mathbb{Q}, \mathbb{Z} \times \mathbb{Z}, |\mathbb{Z} \times \mathbb{Z}| = |\mathbb{Z}|, \text{Prove, cardinality, set, rational, numbers, set, integers, equal, learned, cardinality, discrete, class, days, this, This, confusing, entirely, sure, even, full, question, Let, denote, set, rational, numbers, denote, set, Prove, saying, element, element, know, prove, bijection, even, prove, help, discrete, mathematics, discrete-mathematics}]$
- A.63 $[\text{gcd, lcm, 2, } n_1, n_2, \text{lcm}(n_1, n_2) = \frac{n_1 n_2}{\text{gcd}(n_1, n_2)}, n_1, n_2, n_3, \dots, n_r, \text{gcd, positive, integers, two, positive, integers, relationship, greatest, common, divisor, least, common, multiple, given, gcd, set, positive, integers, does, relationship, hold, Is, TRUE, like, this, prove, handle, proof, explanation, greatest, common, divisor, least, common, multiple, proof-explanation, greatest-common-divisor, least-common-multiple}]$
- A.64† $[f : [a, b] \rightarrow \mathbb{R}, f([a, b]) \subset [a, b], c \in [a, b], f(c) = c, f(a) = a, f(b) = b, f(a) = a, f(b) = b, [a, b], f(a) > a, f(b) < b, f(a) > a, f(b) < b, x, y \in [a, b], f(a) = x, f(b) = y, f[a, b] = [x, y], [x, y] \subset [a, b], [x, y], c \in [a, b], f(c) = c, \text{continuous, Prove, exists, point, satisfying, continuous, Prove, exists, point, satisfying, left, show, well, assume, Since, values, this, assuming, So, far, Assume, Let, means, Notice, Since, continuous, exists, equal, equal, This, assume, Intermediate, Value, Theorem, real, analysis, continuity, proof, explanation, real-analysis, proof-explanation}]$
- A.65 $[e^{-2\lambda t} \lambda^2 \leq \frac{1}{e^{2t^2}}, \lambda, t \geq 0, e^{-2\lambda t} \lambda^2 \leq \frac{1}{e^{2t^2}} 1, \lambda, t \geq 0, \ln, (1), t\lambda \leq e^{t\lambda-1} 2, x \leq e^{x-1}, x \geq 0, \text{show, show, Applying, sides, yields, equivalent, So, show, this, calculus, inequality, exponential, function, exponential-function}]$
- A.66 $[x, h \in \mathbb{R}^d, A \in \mathbb{R}^{d \times d}, (x^T A h)^T = h^T A^T x, x, h \in \mathbb{R}^d, A \in \mathbb{R}^{d \times d}, (x^T A h)^T = h^T A^T x, \text{possible, possible, linear, algebra, transpose, linear-algebra}]$
- A.67 $[k \times k, k \times l, l \times l, \text{Combination, matrixes, matrix, matrix, matrix, prove, matrix, elements, equal, know, rules, calculating, determinants, know, this, question, calculus, determinant}]$

- A.68 $[a^n + 1, a + 1, n, a^n + 1, a + 1, -1, n \in \mathbb{N}, 2k + 1, a^{2(k+1)+1} + 1, = a^{2k+3} + 1, = a^3 \cdot a^{2k} + 1, = (a^3 + 1) \cdot a^{2k} - a^{2k} + 1, a^{2k}, a^n + 1, a + 1, \text{Prove, divisible, odd, Prove, divisible, odd, know, Since, odd, rewrite, assume, holds, prove, holds, next, sure, Since, means, exponential, term, even, cant, use, divisible, odd, polynomials, induction, divisibility}]$
- A.69 $[\binom{s}{s} + \binom{s+1}{s} + \dots + \binom{n}{s} = \binom{n+1}{s+1}, n \geq s, 0 \leq s \leq n, \text{Induction, two, variable, parameters, So, was, assigned, this, problem, tried, professor, explanations, My, professor, saying, statement, need, prove, formula, correct, need, use, induction, variables, sure, help, combinatorics}]$
- A.70† $[\sum_{j=0}^{N-1} \cos \frac{(2j+1)\pi}{2N} = 0, l \in \mathbb{Z}, N \in \mathbb{N}, \sum_{j=0}^{N-1} \cos \left(l \frac{(2j+1)\pi}{2N} \right) = 0, \text{Proving, cos, Let, need, prove, cos, tried, use, Euler, formula, sum, first, terms, geometric, serie, ideas, trigonometry, summation}]$
- A.71† $[1^2 + 2^2 + \dots + n^2 = \frac{n(n+1)(2n+1)}{6}, 1^2 + 2^2 + \dots + n^2 = \frac{n(n+1)(2n+1)}{6}, LHS = 1^2, RHS = \frac{(1+1)(2+1)}{6} = \frac{2 \cdot 3}{6} = 1, LHS_p = 1^2 + 2^2 + \dots + p^2, RHS_p = \frac{p(p+1)(2p+1)}{6}, LHS_{p+1} = 1^2 + 2^2 + \dots + p^2 + (p+1)^2, RHS_{p+1} = \frac{(p+1)((p+1)+1)(2(p+1)+1)}{6}, RHS_{p+1} = RHS_p + (p+1)^2, RHS_{p+1} = \frac{p(p+1)(2p+1)}{6} + (p+1)^2, RHS_{p+1} = \frac{(p+1)((p+1)+1)(2(p+1)+1)}{6}, \frac{p(p+1)(2p+1)}{6} + (p+1)^2, \text{Show, induction, Show, induction, My, Case, Case, Case, show, this, induction, need, show, So, need, rewrite, equal, Anyone, see, Or, solution, induction}]$
- A.72 $[\mathcal{X} = \mathcal{Y}, \mathcal{X} \in \mathcal{Y}, \mathcal{X} = \mathcal{Y}, \mathcal{X} \in \mathcal{Y}, \text{Is, possible, Is, possible, set, equal, set, set, element, set, set, theory, axioms, set-theory}]$
- A.73† $[\binom{n}{0}^2 + \binom{n}{1}^2 + \dots + \binom{n}{n}^2 = \binom{2n}{n}, \binom{n}{0}^2 + \binom{n}{1}^2 + \dots + \binom{n}{n}^2 = \binom{2n}{n}, (1+x)^n(1+x)^n = (1+x)^{2n}, x^n, \binom{n}{0}1 + \binom{n}{1}x + \dots + \binom{n}{r}x^r + \dots + \binom{n}{n}x^n, x^n, x^i, x^{n-i}, x^n, \binom{n}{n-r} = \binom{n}{r}, \binom{n}{0}^2 + \binom{n}{1}^2 + \dots + \binom{n}{n}^2, x^n, \binom{2n}{n}, x^n, \binom{2n}{n}, \binom{n}{0}^2 + \binom{n}{1}^2 + \dots + \binom{n}{n}^2 = \binom{2n}{n}, x^n, x^n, x^n, \text{Help, proof, proof, required, made, binomial, expose, was, forward, questions, exposing, see, question, marks, like, this, one, points, quite, numeration, This, Prove, use, equality, call, result, proved, finding, coefficient, terms, this, equality, binomial, theorem, left, hand, side, this, equation, product, two, factors, equal, factors, multiply, term, term, first, factor, has, term, second, factor, has, coefficients, Since, summation, equal, So, left, hand, side, equation, coefficient, expand, right, hand, side, equation, find, coefficient, left, hand, side, equation, prove, equal, In, This, was, My, one, heck, does, this, equation, come, know, equation, come, requested, prove, different, equality, two, why, solution, first, equation, finding, coefficients, one, made, see, one, three, why, coefficient, left, hand, side, equation, made, coefficient, right, hand, side, this, equation, well, prove, original, equation, one, was, prove, first, place, know, many, guys, help, long, post, long, (?-n), (?-1), (?-2)., left-hand, right-hand, (?-3), left-hand, (?-1), (?-2), (?-3), right-hand, discrete, mathematics, binomial, coefficients, binomial, theorem, discrete-mathematics, binomial-coefficients, binomial-theorem}]$

- A.74 $[f : (0, \infty) \rightarrow \mathbb{R}, f(x) = x + \frac{1}{x}, [2, \infty), f : (0, \infty) \rightarrow \mathbb{R}, f(x) = x + \frac{1}{x}, [2, \infty), x = 1, f(1) = 2, [2, \infty), \text{Show, image, function, interval, Show, image, function, interval, So, show, function, interval, functions, elementary, set, theory, elementary-set-theory}]$
- A.75 $[m, \lim_{u \rightarrow \infty} \frac{u^m}{e^u} = 0, \lim_{u \rightarrow \infty} \frac{u^m}{e^u} = 0, e^u, >, \frac{u^{m+1}}{(m+1)!}, \text{Prove, integer, show, integer, Looking, solutions, sure, this, logical, step, real, analysis, calculus, limits, real-analysis}]$
- A.76† $[\mathbb{Z}, a + b\mathbb{Z} = \{z \in \mathbb{Z} \mid z = a + bk \text{ for some } k \in \mathbb{Z}\}, a \in \mathbb{Z}, b \in \mathbb{Z} - \{0\}, \{a_i + b_i\mathbb{Z} \mid i \in \mathbb{N}\}, \bigcup_{i \in \mathbb{N}} (a_i + b_i\mathbb{Z}) = \mathbb{Z}, I \subset \mathbb{N}, \bigcup_{i \in I} (a_i + b_i\mathbb{Z}) = \mathbb{Z}, \{p_k\} = \{2, 3, 5, \dots\}, \ell_1, \ell_2, \ell_1 = \ell_2 = 5, \{0 + p_{k_j}\}, j = 1, \dots, n, p > \max\{p_{k_1}, \dots, p_{k_n}\}, p \notin \bigcup_{1 \leq j \leq n} (0 + p_{k_j}\mathbb{Z}), p \notin (-1 + 5\mathbb{Z}) \cup (1 + 5\mathbb{Z}), 5 \nmid p - 1, 5 \nmid p + 1, \text{Covering, arithmetic, progressions, solving, problems, old, exam, topology, translated, problem, algebraic, terms, problem, Let, collection, sets, satisfying, Show, whether, always, possible, extract, finite, lot, elementary, algebra, Let, set, construct, non, negative, integers, instance, pick, finite, sub, collection, assume, prime, max, run, This, course, possible, prime, last, this, approach, means, prove, infinitely, many, primes, ending, like, thing, prove, simple, problem, like, Surely, way, solving, this, EDIT, interested, solution, topology, whether, solution, like, solution, works, non-negative, sub-collection, general, topology, elementary, number, theory, general-topology, elementary-number-theory}]$
- A.77 $[(-1)(-1) = 1, (-1)(-1) = 1, 1 \cdot 1 = 1, 1 \cdot x = x, \text{Show, relation, consequence, distributive, law, Show, relation, consequence, distributive, This, question, first, problem, Theory, Andre, point, tried, using, help, work, elementary, number, theory, elementary-number-theory}]$
- A.78 **training data**
- A.79 $[\left| \frac{e^{-ixu} - 1}{u} \right| \leq |x|, u \neq 0, \text{Inequality, complex, exponential, Rudin, Real, Complex, Analysis, uses, this, proof, near, chapter, real, Why, this, TRUE, Edit, real, inequality, exponential, function, fourier, transform, exponential-function, fourier-transform}]$
- A.80 $[\mathbb{N}, \emptyset, \{1\}, \{2\}, \{1, 2\}, \{3\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}, \{4\}, \dots, \{1\}, \{2\}, [n] = \{1, 2, \dots, n\}, [n], 2^n, \text{Why, does, this, proof, set, finite, subsets, countable, set, work, set, subsets, found, this, proof, thread, found, simple, answers, sort, formula, like, trying, way, see, set, finite, subsets, countable, probably, list, elements, set, one, coming, first, next, shows, set, pattern, see, least, In, words, first, comes, comes, time, new, integer, list, subsets, contain, ones, contain, showed, subsets, show, first, elements, this, applies, finite, subsets, cant, why, apply, set, subsets, continue, this, scheme, assume, way, infinity, quite, help, analysis, elementary, set, theory, proof, explanation, elementary-set-theory, proof-explanation}]$

- A.81† [$n \in \mathbb{N}$, $\{1, \dots, n\}$, $M \neq \emptyset$, $x \in M$, $f(1)$, $f(1) := x$, $\{1\}$, $\{1, \dots, n\}$, $M - \{f(1), \dots, f(n)\}$, $x \in M - \{f(1), \dots, f(n)\}$, $g(1), \dots, g(n+1)$, $g(1) := f(1), \dots, g(n) := f(n)$, $g(n+1) := x$, $\{1, \dots, n+1\}$, n_1 , $h(n_1)$, $\{1, \dots, n_1\}$, $g(n_1)$, $h(n_1)$, $(1, g(1)), \dots, (n_1, g(n_1))$, $h(n_2)$, $n_2 \leq n_1$, $g(n_2)$, $g(n_2)$, $h(n_2)$, $h(n_3)$, $n_3 > n_1$, $(n_1 + 1, g(n_1 + 1)), \dots, (n_3, g(n_3))$, $g(n_3)$, $h(n_3)$, $h(n)$, $n \in \mathbb{N}$, $h : \mathbb{N} \rightarrow M$, $h : \mathbb{N} \rightarrow M$, $h : \mathbb{N} \rightarrow M$, infinite, set, contains, countable, Why, proof, Axiom, Choice, Let, infinite, Proposition, exists, injection, Since, exists, Define, injection, exists, injection, Since, infinite, set, empty, So, exists, Define, injection, Let, arbitrary, natural, calculate, injection, Proposition, return, value, pairs, calculate, search, database, value, database, return, value, calculate, pairs, database, Proposition, return, value, calculate, above, injection, Why, proof, man, know, injection, man, know, injection, elementary, set, theory, axiom, choice, elementary-set-theory, axiom-of-choice]
- A.82† [$f(x)$, $[0, 2\pi]$, $f(x)$, $[0, 2\pi]$, $A = \int_0^{2\pi} f(x)dx$, $g(x)$, $f(x) = g(x) \cdot \cos(x)$, $f(x)$, $A = \int_0^{2\pi} g(x) \cdot \cos(x)dx$, $t = \sin(x)$, $dt = \cos(x)dx$, $x = \arcsin(t)$, $t = \sin(x)$, $\sin(0) = 0$, 2π , $\sin(2\pi) = 0$, $A = \int_0^0 g(\arcsin(t))dt$, $A = 0$, definite, integrals, evaluate, using, periodic, functions, know, reasoning, know, went, discuss, this, Maths, even, find, Let, assuming, function, continuous, has, antiderivative, interval, Let, area, curve, interval, exist, function, cos, Substituting, value, cos, Using, substitution, Let, sin, cos, arcsin, Changing, limits, sin, sin, sin, Substituting, definite, integral, arcsin, Definite, Integral, lower, upper, bounds, So, help, calculus, integration, trigonometry, definite, integrals, inverse, function, definite-integrals, inverse-function]
- A.83 [$a_n \leq M$, Is, sequence, sums, inverse, natural, numbers, bounded, reading, Infinite, Sequences, right, trivial, matter, determine, boundedness, mind, sequence, before, learn, sequence, know, sequence, bounded, above, number, calculus, sequences, series, harmonic, numbers, sequences-and-series, harmonic-numbers]
- A.84† [$4, x$, $Z[x]$, $I = \langle p, x \rangle$, $Z[x]$, $I = \langle p, x \rangle$, $Z[x]$, $4, x$, $Z[x]$, Is, ideal, generated, principal, ideal, principal, ideal, My, question, Is, principal, ideal, prime, ideal, generated, principal, ideal, abstract, algebra, ring, theory, ideals, principal, ideal, domains, abstract-algebra, ring-theory, principal-ideal-domains]

- A.85 $[N, (+1), 1/2, 1/2, p_n = \frac{1}{2}p_{n-1}, p_n, p_N = 1, p_{N-1} = 2, p_0 = 2^N,$
 Expected, number, steps, bug, reach, position, bug, time, position, At,
 step, bug, moves, right, step, probability, returns, origin, probability, ex-
 pected, number, steps, this, bug, reach, position, tried, first, find, possi-
 bility, this, bug, reaches, number, steps, recurrence, equation, find, possi-
 bility, bug, position, reach, boundary, condition, see, does, make, sense,
 sort, value, probability, first, number, expected, steps, sure, recurrence,
 equation, markov, chains, random, walk, markov-chains, random-walk]
- A.86 $[\sum_{k=0}^n k \cdot \binom{n}{k}] = O(2^{n \log_3 n})?$, Is, TRUE, log, Problem, Is, TRUE, log,
 My, solution, this, upper, bound, way, large, ca, find, solution, combina-
 torics, elementary, number, theory, discrete, mathematics, elementary-
 number-theory, discrete-mathematics]
- A.87 $[\forall n \in \mathbb{N} : (\sum_{i=1}^n a_i)(\sum_{i=1}^n \frac{1}{a_i}) \geq n^2, a_i, \forall i \in \mathbb{N} : a_i \in \mathbb{R}^+, n = 1,$
 $n = 2, n = 3, a, b \in \mathbb{R}^+, ab = 1, a + b \geq 2, n = 3, a, b, c \in \mathbb{R}^+,$
 $(\frac{a}{b} + \frac{b}{a}) \geq 2, (\frac{a}{c} + \frac{c}{a}) \geq 2, (\frac{b}{c} + \frac{c}{b}) \geq 2, P(1), P(n),$ Is, TRUE, positive,
 TRUE, prove, this, holds, using, lemma, Lemma, Let, example, case,
 proven, like, this, Let, lemma, sure, generalized, version, natural, ca,
 come, counterexample, try, prove, induction, Let, Base, case, Inductive,
 hypothesis, assume, Inductive, step, know, Is, this, inequality, TRUE,
 prove, anyone, show, counterexample, algebra, precalculus, inequality,
 algebra-precalculus]
- A.88 $[x^4 + 10x^2 + 1, \mathbb{Z}[x], x^4 + 10x^2 + 1, \mathbb{Z}[x],$ Is, polynomial, reducible,
 Is, polynomial, reducible, abstract, algebra, ring, theory, field, the-
 ory, irreducible, polynomials, abstract-algebra, ring-theory, field-theory,
 irreducible-polynomials]
- A.89 $[A^2 + B^2 = C^2 + D^2, A, B, C, D,$ Parametrization, pythagorean, like,
 equation, pythagorean-like, Is, known, complete, parametrization, Dio-
 phantine, equation, positive, rational, numbers, equivalently, integers,
 number, theory, diophantine, equations, number-theory, diophantine-
 equations]
- A.90 $[n \times n, n \times n, A^{-1}, A^{-1}A = \mathbb{I}_n \wedge AA^{-1} = \mathbb{I}_n \quad (1), \mathbb{I}_n, n \times n, A^{-1}A = \mathbb{I}_n,$
 $AA^{-1} \neq \mathbb{I}_n,$ Question, definition, Inverse, matrix, definition, matrix,
 inverse, matrix, property, identity, cases, way, around, making, FALSE,
 statement, linear, algebra, matrices, linear-algebra]
- A.91† $[\mathbb{R} \rightarrow \mathbb{R}, x^2,$ Continuous, function, reaches, value, range, times, Is, con-
 tinuous, function, reaches, possible, values, value, range, times, example,
 perfect, was, question, almost, certain, functions, knows, bunch, real,
 analysis, calculus, real-analysis]

- A.92† $[\mathbb{F}_q[X, Y], K[X, Y], \mathbb{F}_q[X, Y], I = (P(X, Y)), Q(X), \mathbb{F}_q[X, Y]/(Q(X)) \cong (\mathbb{F}_q[X]/(Q(X)))[Y], P(X, Y) = \sum_{i=0}^n P_i(X)Y^i, d \geq 1, P_i, P_n, \mathbb{F}_q, P_n, \prod_{\alpha \in \mathbb{F}_q} (X - \alpha), \alpha \in \mathbb{F}_q, P_n(\alpha) \neq 0, I := (P, X - \alpha), (P), (P), \mathbb{F}_q[X, Y], 1 \in I, X = \alpha, n = \deg_Y(P) = 0, P(X, Y) = (X^{p-1} - 1)XY - 1, P(X, Y) = (X^{p-1} - 1)XY - X - 1, \mathbb{F}_p[X, Y], \text{Is, principal, maximal, ideal, Given, infinite, field, one, prove, maximal, ideal, ca, In, non, principal, prime, ideal, maximal, ideal, generated, two, whether, result, holds, find, principal, ideal, irreducible, polynomial, maximal, ideal, polynomial, positive, degrees, given, irreducible, polynomial, only, one, variable, quotient, ring, polynomials, field, polynomial, form, polynomials, polynomial, vanishes, identically, divisible, ideal, contain, maximal, whole, Writing, evaluating, leads, conclusion, This, infer, above, tried, looking, prime, number, determining, whether, quotient, field, know, answer, problem, proof, counter, example, very, advance, non-principal, counter-example, polynomials, commutative, algebra, field, theory, finite, fields, commutative-algebra, field-theory, finite-fields}]$
- A.93 $[AB =, BA, A, B, \mathbb{R}, nxn, \det(xI - AB) = \det(xI - BA), \text{Characteristic, Polynomial, characteristic, polynomial, Let, matrix, size, prove, singular, matrix, linear, algebra, polynomials, linear-algebra}]$
- A.94† $[2^{\aleph_0}, 2^{\aleph_0}, \text{sets, natural, numbers, two, finite, intersection, Question, sets, natural, numbers, two, finite, read, infinite, families, need, ignore, Property, family, finitely, many, sets, family, sets, infinite, probably, find, ones, property, Property, family, almost, tower, pick, two, sets, family, one, almost, contained, probably, meaning, find, intersection, finite, contained, find, rational, numbers, natural, numbers, rational, numbers, paired, natural, ones, solving, this, problem, families, rational, numbers, solving, families, natural, need, various, ways, defining, real, help, "almost-tower":, almost-contained, cardinals, rational, numbers, natural, numbers, rational-numbers, natural-numbers}]$
- A.95† $[\sin x, \sin x, \angle B = x, \lim_{x \rightarrow 0} \frac{\sin x}{x} = 1, \sin x, \sin x < x < \tan x, \sin x < x < \sin x + 1 - \cos x, \sin x, \mathbb{R}^2, F \subset G, |\partial F| < |\partial G|, |\partial F|, \int_s^t \sqrt{1 + y'(u)^2} du, y'', y(s) = y(t), y', y = \sqrt{1 - x^2}, -1 \leq x \leq 1, \sin x, \text{Length, convex, paths, bounding, sin, problem, Defining, sin, leg, right, triangle, radians, hypotenuse, prove, sin, motivation, find, derivative, sin, elementary, pre, Taylor, pre, series, context, many, times, proof, based, length, arc, satisfies, sin, sin, sin, cos, lower, bound, clear, sin, length, straight, segment, length, curved, segment, find, upper, bound, based, this, intuitive, two, convex, subsets, length, boundary, proof, tried, trying, bound, example, negative, realize, ca, bounded, note, example, Question, Is, possible, derivative, sin, inequality, lengths, bounds, Is, calculus, text, this, approach, similar, define, functions, pre-Taylor, pre-series, real, analysis, derivatives, trigonometry, reference, request, arc, length, real-analysis, reference-request, arc-length}]$

- A.96 $[\sum_i a_i, a_i, \sum_i \frac{a_i}{a_i+a_{i+1}+a_{i+2}+\dots}, \sum_i a_i, a_i, s_n = \sum_{i=n}^{\infty} a_i, \sum_i \frac{a_i}{s_i}, a_i = r^i, a_i = 1/i^2, \text{Let, convergent, sum, positive, Does, always, diverge, Let, convergent, sum, Let, Does, always, diverge, tried, examples, geometric, series, always, diverge, sequences, series, analysis, sequences-and-series}]$
- A.97† $[\mathbb{R}, \mathbb{Q}, \mathbb{C}, \mathbb{R}, \mathbb{C} = \text{span}\{1, i\}, \mathbb{R}, \mathbb{Q}, m := n + 1, x_1, \dots, x_m, \frac{a_1}{b_1}x_1 + \dots + \frac{a_m}{b_m}x_m = 0, \text{lcm}(b_1, \dots, b_m), \text{dimension, vector, space, field, look, vector, space, dimension, question, dimension, vector, space, like, answer, infinity, dimension, was, finite, say, real, numbers, was, linear, combination, rational, coefficients, Multiplying, real, numbers, linear, combination, natural, coefficients, FALSE, prove, linear, algebra, vector, spaces, linear-algebra, vector-spaces}]$
- A.98 $[R : S^1 \rightarrow S^1, \{R^n([x])\}, S^1, \alpha, R : S^1 \rightarrow S^1, [x] \rightarrow [x + \alpha], [x] \in S^1, \{R^n([x])\}, S^1, R(e^{2\pi i x}) = e^{2\pi i(x+\alpha)}, R^n(x), n \in \mathbb{Z}, R^n(e^{2\pi i x}) = e^{2\pi i(x+n\alpha)}, e^{2\pi i x} \in S^1, [y] = e^{2\pi i y} \in S^1, [y], [z_y] = e^{2\pi i z}, [z_y] = R^n([x]), n \in \mathbb{Z}, \text{irrational, rotation, dense, points, Let, irrational, number, irrational, rotation, need, prove, set, dense, First, write, So, write, need, prove, neighborhood, contains, point, Is, way, correct, does, does, need, this, question, showing, proof, different, proofs, general, topology, circles, rotations, irrational, numbers, general-topology, irrational-numbers}]$
- A.99 $[f : \mathbb{R} \rightarrow \mathbb{R}, \mathbb{Q}, \text{Rationals, set, continuity, function, functions, discontinuities, rationals, continuities, irrationals, function, continuities, rationals, Or, words, set, continuities, function, real, analysis, calculus, continuity, real-analysis}]$
- A.100† $[E = (0, 1], E \subseteq \mathbb{R}, \text{Is, this, set, closed, Is, correct, say, this, set, set, real, numbers, closed, real, analysis, real-analysis}]$
- A.101 **training data**