NTCIR-10 Math Pilot Task Overview

Akiko Aizawa & Michael Kohlhase & Iadh Ounis

http://kwarc.info/kohlhase
Center for Advanced Systems Engineering
Jacobs University Bremen, Germany

NTCIR-10, June 19. 2013
Introduction & Motivation for a Math Pilot Task
Introduction/Background

- **Mathematics** plays a fundamental role in Science, Technology, and Engineering (learn from Math, apply for STEM)

- Mathematical knowledge is rich in content, sophisticated in structure, and technical in presentation!
Introduction/Background

- **Mathematics** plays a fundamental role in Science, Technology, and Engineering (learn from Math, apply for STEM)

- Mathematical knowledge is rich in content, sophisticated in structure, and technical in presentation!

- There is a lot of documents with maths
  - there are 120,000 journal articles per year in pure/applied math, 3.5 Million overall
  - 50 million science articles in 2010 [Jin10] with a doubling time of 8-15 years [LvI10]
  - And this excludes gray literature, engineering, and school textbooks.
  - Even in the Renaissance, polymaths like Leonardo de Vinci were a rare exception.
Introduction/Background

- **Mathematics** plays a fundamental role in Science, Technology, and Engineering (learn from Math, apply for STEM)
- Mathematical knowledge is rich in content, sophisticated in structure, and technical in presentation!
- There is a lot of documents with maths
  - there are **120,000 journal articles per year** in pure/applied math, **3.5 Million overall**
  - **50 million science articles** in 2010 [Jin10] with a **doubling time of 8-15 years** [LvI10]
  And this excludes gray literature, engineering, and school textbooks.
- Even in the Renaissance, polymaths like Leonardo de Vinci were a rare exception.
- We need IR support to deal with this! (⇝ NTCIR-10 Math Pilot Task)
Mathematics Resources on the Web

Example 1 (The Wolfram Functions Site) contains \( \geq 307k \) Formuae

\[
\exp \text{ Exp} \]

Exp
Exponential function

Mathematica Notation: \( \text{Exp}[z] \)

Traditional Notation: \( \exp(z) = e^z \)

Fourier transformation and Parseval relation (1 formula)

\[
\mathcal{F}(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{ixy} \, dx \iff f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathcal{F}(y) e^{-iyx} \, dy;
\]

\[
\int_{-\infty}^{\infty} f_1(t) f_2(x-t) \, dt = \int_{-\infty}^{\infty} \mathcal{F}_1(y) \mathcal{F}_2(y) e^{-iyx} \, dy.
\]
More Mathematics on the Web

- The Connexions project  
  (http://cnx.org)
- Wolfram Inc.  
  (http://functions.wolfram.com)
- Eric Weisstein’s MathWorld  
  (http://mathworld.wolfram.com)
- Digital Library of Mathematical Functions  
  (http://dlmf.nist.gov)
- Cornell ePrint arXiv  
  (http://www.arxiv.org)
- Zentralblatt Math  
  (http://www.zentralblatt-math.org)
- ...Engineering Company Intranets, ...

- **Question**: How will we find content that is relevant to our needs
- **Idea**: try Google  
  (like we always do)
- **Scenario**: Try finding the distributivity property for \( \mathbb{Z} \)
  \[ (\forall k, l, m \in \mathbb{Z}. k \cdot (l + m) = (k \cdot l) + (k + m)) \]
Searching for Distributivity

Google search with the query: "forall k,l,m:Z. k * (l + m) = k*l + k*m"

Tip: Try removing quotes from your search to get more results.

Your search - "forall k,l,m:Z. k * (l + m) = k*l + k*m" - did not match any documents.

Suggestions:

- Make sure all words are spelled correctly.
- Try different keywords.
- Try more general keywords.
Searching for Distributivity

\[ \forall x, y, z : \mathbb{Z}. \ x \cdot (y + z) = x \cdot y + x \cdot z \]

Untitled Document

... theorem distributive \_Ztimes\_Zplus: distributive \_Ztimes\_Zplus. change with (\\forall x, y, z : \mathbb{Z}. x \cdot (y + z) = x \cdot y + x \cdot z). intros.elim x. ...

matita.cs.unibo.it/library/\mathbb{Z}/times.ma - 21k - Cached - Similar pages
Searching for Distributivity

Mathematica - Setting up equations
Try *Reduce* rather than *Solve* and use *ForAll* to put a condition on x, y, and z. In[1]:=
Reduce[ForAll[{x, y, z}, 5x + 6y + 7z == a*x + b*y + c*z], ...
www.codecomments.com/archive382-2006-4-904844.html - 18k - Supplemental Result -
Cached - Similar pages

File Format: PDF/Adobe Acrobat - View as HTML
7.2 Appendix B. Elliptic constants related to g(l(N,C). ... 1 for all s ≤ j}. (4.14). The first condition means that the traces (4.13) of the Lax operator ...
www.citebase.org/cgi-bin/fulltext?format=application/pdf&identifier=oai:arXiv.org:nlin/0309017 -
Supplemental Result - Similar pages

\documentclass[article] \usepackage{axiom} \usepackage{amssymb ...
Of course Google cannot work out of the box

- Formulae are not words:
  - $a, b, c, k, l, m, x, y, \text{ and } z$ are (bound) variables.
    (do not behave like words/symbols)
  - where are the word boundaries for “bag-of-words” methods?
Of course Google cannot work out of the box

- Formulae are not words:
  - $a, b, c, k, l, m, x, y,$ and $z$ are (bound) variables. (do not behave like words/symbols)
  - where are the word boundaries for “bag-of-words” methods?

- Idea: Need a special treatment for formulae (translate into “special words”) Indeed this is done ([MY03, MM06, LM06, MG11]) . . . and works surprisingly well (using Lucene as an indexing engine)

- Idea: Use database techniques (extract metadata and index it) Indeed this is done for the Coq/HELM corpus ([AGC+06])

- Idea: Use Automated Reasoning Techniques (Term Indexing [Nor06, KŠ06, KMP12])

- Idea: Use standard IR techniques (Learn from the NTCIR crowd?)
Of course Google cannot work out of the box

- Formulae are not words:
  - $a, b, c, k, l, m, x, y,$ and $z$ are (bound) variables. (do not behave like words/symbols)
  - where are the word boundaries for “bag-of-words” methods?

- Idea: Need a special treatment for formulae (translate into “special words”)
  - Indeed this is done ([MY03, MM06, LM06, MG11])
  - ...and works surprisingly well (using Lucene as an indexing engine)

- Idea: Use database techniques (extract metadata and index it)
  - Indeed this is done for the Coq/HELM corpus ([AGC+06])

- Idea: Use Automated Reasoning Techniques (Term Indexing [Nor06, KŠ06, KMP12])

- Idea: Use standard IR techniques (Learn from the NTCIR crowd?)

- Which one is best?: We do not really know, evaluation is very difficult

- Future: maybe even mix/integrate the respective best features (once we know)
MathML3 is a W3C Recommendation for representing Formulae [ABC+10]

**Idea:** Combine the presentation and content markup and cross-reference

use e.g. for semantic copy and paste.

(click on presentation, follow link and copy content)
MathML3 is a W3C Recommendation for representing Formulae \([ABC^{+10}]\)

**Idea:** Combine the presentation and content markup and cross-reference

\[
\frac{3}{(x+2)}
\]

**Solution:** Write \(\LaTeX\), convert to HTML5 \(\Rightarrow\) HTML+MathML+SVG

use e.g. for semantic copy and paste.

\(\text{click on presentation, follow link and copy content}\)

**But:** Formulae are mostly written in \(\LaTeX\), e.g. \(\backslash{\text{frac}}\{3\}\{(x+2)\}\)
Parallel Markup Markup in MathML

- **Concrete Realization in MathML**: `semantics` element with presentation as first child and content in `annotation-xml` child

```xml
<semantics>...
<annotation-xml>...
<mfrac id="M">...
<mn id="3">3</mn>
<mfenced id="f">...
<mi id="x">x</mi>
<mo id="p">+</mo>
<mn id="2">2</mn>
<apply href="M">...
<apply href="f">...
<apply href="p"/>
<ci href="x">x</ci>
<ci href="3">3</ci>
<divide/>
<plus/>
<cn href="2">2</cn>
</apply>
</mfenced>
</mfrac>
</annotation-xml>
</semantics>
```
Task Description
Task Overview

- **Math Retrieval Subtask:** Given a document collection, retrieve relevant mathematical formulae or documents for a given query.

- **Math Understanding Subtask:** Extract natural language definitions of mathematical expressions in a document for their semantic interpretation.
NTCIR-Math Pilot Task: Task Design & State

- **NTCIR-10 Math Dataset:** 100,000 Documents transformed to HTML5 from http://arxiv.org (10,000 for dry run)
  - 63 GiB overall size, 35 MFormulae, 297 MSubformulae (size challenge for systems)
  - every formula given in content MathML, presentation MathML, and \texttt{\LaTeX} (23 GiB)

- **NTCIR Topics:**
  - Three Math Retrieval Subtasks: (write \texttt{\LaTeX}+?, transform to MathML Query)
    - Formula Search (FS; automated): Formulae with named wildcards e.g. $\int \, f(x)^2 \, dx$
    - Full Text Search (FT; automated): (Formulae and keywords) e.g. Bell curve in the form of $\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
    - Open Math IR (OMIR; semi-automated): human-readable questions e.g. For which $n$ and $k$ is $\text{PSL}(n, k)$ not commutative?

- **Math Understanding Task:** Manually created content MathML as gold standard

- **State of Play:** Establishing community (16/6 Teams), ran successful Task (made mistakes, learnt a lot)
The uncertainty principle requires that sequential measurements of non-commuting spin components cannot achieve a resolution of $\epsilon = 1$ at zero back-action. For orthogonal spin components, the quantitative limit can be expressed in terms of the uncertainty relation [18, 19]

$$\epsilon^2 + (1 - \eta)^2 \leq 1.$$  \hspace{1cm} (5)

It is therefore impossible to measure the intrinsic joint probabilities $P_{\psi}(s_2, s_3)$ directly. However, the spin flip model allows us to reconstruct this joint probability from the experimentally observed distribution of sequential outcomes, $P_{\text{exp}}(s_2, s_3)$. Due to the spin flip errors, each measurement outcome $(s_2, s_3)$ can also originate from different spin values, with probabilities determined by the spin flip probabilities of $(1 - \epsilon)/2$ and $\eta/2$. The relation between the experimental probabilities and the intrinsic probabilities is then given by

$$P_{\text{exp}}(s_2, s_3) = \left(\frac{1 + \epsilon}{2}\right)\left(1 - \frac{\eta}{2}\right)P_{\psi}(s_2, s_3) + \left(\frac{1 - \epsilon}{2}\right)\left(1 - \frac{\eta}{2}\right)P_{\psi}(-s_2, s_3)$$

$$+ \left(\frac{1 + \epsilon}{2}\right)\left(\frac{\eta}{2}\right)P_{\psi}(s_2, -s_3) + \left(\frac{1 - \epsilon}{2}\right)\left(\frac{\eta}{2}\right)P_{\psi}(-s_2, -s_3).$$  \hspace{1cm} (6)

This linear map can be inverted to reconstruct the intrinsic joint probabilities $P_{\psi}(s_2, s_3)$ from the experimentally observed joint probabilities $P_{\text{exp}}(s_2, s_3)$. If the measurement resolution and the back-action are known, the same joint probabilities $P_{\psi}(s_2, s_3)$ should be obtained at any measurement strength. The relations that describe the reconstruction of intrinsic joint probabilities from the measurement data are given by

$$P_{\psi}(s_2, s_3) = \frac{(1 + \epsilon)(2 - \eta)}{4\epsilon(1 - \eta)}P_{\text{exp}}(s_2, s_3) - \frac{(1 - \epsilon)(2 - \eta)}{4\epsilon(1 - \eta)}P_{\text{exp}}(-s_2, s_3)$$

$$- \frac{(1 + \epsilon)\eta}{4\epsilon(1 - \eta)}P_{\text{exp}}(s_2, -s_3) + \frac{(1 - \epsilon)\eta}{4\epsilon(1 - \eta)}P_{\text{exp}}(-s_2, -s_3).$$  \hspace{1cm} (7)

Note that the spin flip model used to reconstruct the intrinsic joint probabilities of the quantum state does not require any assumptions from quantum theory and is based entirely on the experimentally observable spin flip rates $(1 - \epsilon)/2$ and $\eta/2$. 

---

NTCIR-Math Pilot Task: Task Design & State

- **NTCIR-10 Math Dataset**: 100,000 Documents transformed to HTML5 from http://arxiv.org (10,000 for dry run)
  - 63 GiB overall size, 35 MFormulae, 297 MSubformulae (size challenge for systems)
  - every formula given in content MathML, presentation MathML, and LaTeX (23 GiB)

- **NTCIR-10 Topics**:
  Three Math Retrieval Subtasks: (write LaTeX+?, transform to MathML Query)
  - **Formula Search (FS; automated)**: Formulae with named wildcards e.g. \( \int_{\theta_1}^{\theta_2} \sin^2(x) \, dx \)
  - **Full Text Search (FT; automated)**: (Formulae and keywords)
    e.g. Bell curve in the form of \( \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \)
  - **Open Math IR (OMIR; semi-automated)**: human-readable questions
    e.g. For which \( n \) and \( k \) is \( \text{PSL}(n, k) \) not commutative?

Math Understanding Task: Manually created content MathML as gold standard
NTCIR-Math Pilot Task: Task Design & State

- **NTCIR-10 Math Dataset**: 100,000 Documents transformed to HTML5 from http://arxiv.org (10,000 for dry run)
  - 63 GiB overall size, 35 MFormulae, 297 MSubformulae (size challenge for systems)
  - every formula given in content MathML, presentation MathML, and \LaTeX\ (23 GiB)

- **NTCIR-10 Topics**:
  - Three Math Retrieval Subtasks: (write \LaTeX+?, transform to MathML Query)
    - Formula Search (FS; automated): Formulae with named wildcards e.g. $\int_{?l}^{?h} ?f(x)^2 dx$
    - Full Text Search (FT; automated): (Formulae and keywords)
      - e.g. Bell curve in the form of $\frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
    - Open Math IR (OMIR; semi-automated): human-readable questions
      - e.g. For which $n$ and $k$ is $PSL(n, k)$ not commutative?

- **Math Understanding Task**: Manually created content MathML as gold standard

- **State of Play**: Establishing community (16/6 Teams), ran successful Task (made mistakes, learnt a lot)
Participation, Evaluation & Results
### NTCIR-10 Math Pilot Task Participants

<table>
<thead>
<tr>
<th>Group ID</th>
<th>Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRKLY</td>
<td>University of California, USA</td>
</tr>
<tr>
<td>FSE</td>
<td>Technische Universität Berlin, Germany</td>
</tr>
<tr>
<td>KWARC</td>
<td>Jacobs University, Germany</td>
</tr>
<tr>
<td>MCAT</td>
<td>National Institute of Informatics, Japan</td>
</tr>
<tr>
<td>MIRMU</td>
<td>Masaryk University, Czech Republic</td>
</tr>
<tr>
<td>NAK</td>
<td>Keio University, Japan</td>
</tr>
</tbody>
</table>
Number of runs for each subtask category.

<table>
<thead>
<tr>
<th>Group ID</th>
<th>Subtasks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MIR/FS</td>
</tr>
<tr>
<td>BRKLY</td>
<td>4</td>
</tr>
<tr>
<td>FSE</td>
<td>1</td>
</tr>
<tr>
<td>KWARC</td>
<td>1</td>
</tr>
<tr>
<td>MCAT</td>
<td>1</td>
</tr>
<tr>
<td>MIRMU</td>
<td>4</td>
</tr>
<tr>
<td>NAK</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>12</td>
</tr>
</tbody>
</table>

* Reported only document URIs without formula IDs and were not included in the relevance judgment pool.
<table>
<thead>
<tr>
<th>Query type</th>
<th>Distributed</th>
<th>Evaluated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formula Search</td>
<td>22</td>
<td>21</td>
</tr>
<tr>
<td>Full Text Search</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Open Search</td>
<td>19</td>
<td>0</td>
</tr>
</tbody>
</table>
Assessment: Math Extension for SEPIA

Task: NTCIR-10 Math Task (EN), Username: admin

Select a pool and then topic and you will see a list of potentially relevant documents to judge. For each document, judge relevance.

Topic: Derivative approximation

Question (TrgLang) Derivative approximation
Information Need (TrgLang) $\frac{f(x+h)-f(x)}{h}$
Query words
Answer Type Formula Search Query

Document List:

| [x] f095933#d79338 |
| [x] f005076#d16105712 |
| [x] f056009#d67008 |
| [x] f084809#d120008 |
| [x] f056393#d60623 |
| [x] f093556#d81652 |
| [x] f050214#d54091 |
| [x] f008232#d60483 |
| [x] f022048#d53712 |
| [x] f003698#d63751 |
| [x] f074593#d61838 |
| [x] f021585#d66555 |
| [x] f098185#d6999 |
| [x] f086627#d130041 |
| [x] f008946#d53678 |
| [x] f075613#d86622 |
| [x] f019088#d71630 |
| [x] f038931#d87832 |
| [x] f018041#d55519 |

Relevance Judgment

<DOC>
<DOCNO>f095933#id79338</DOCNO>
<URL>0207/cond-mat.0207603/cond-mat.0207603.xhtml#id79338</URL>
<br>

$h$ and the optical conductivity is obtained by a convolution of two full Green functions:

\[
\int d\omega \left( -\frac{f(\omega + \Omega) - f(\omega)}{\Omega} \right) \int d\epsilon \rho(\epsilon) \rho(\epsilon, \omega) \rho(\epsilon, \omega + \Omega)
\]

</DOC>

Evidence:

- Relevant
- Partially Relevant
- Not Relevant
Math Retrieval Subtask: Pooling

- Select formulae as evenly as possible from all the runs.
- The current top ranked formulae were taken from all the ranking lists, and added to the pool if they were not found.
- This process was repeated until the total size of the pool becomes equal or greater than 100.
### Pooling Results

<table>
<thead>
<tr>
<th>Query ID</th>
<th>Relevance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td>FS-1</td>
<td>0</td>
</tr>
<tr>
<td>FS-2</td>
<td>0</td>
</tr>
<tr>
<td>FS-3</td>
<td>10</td>
</tr>
<tr>
<td>FS-4</td>
<td>8</td>
</tr>
<tr>
<td>FS-5</td>
<td>38</td>
</tr>
<tr>
<td>FS-6</td>
<td>0</td>
</tr>
<tr>
<td>FS-7</td>
<td>10</td>
</tr>
<tr>
<td>FS-8</td>
<td>48</td>
</tr>
<tr>
<td>FS-9</td>
<td>0</td>
</tr>
<tr>
<td>FS-10</td>
<td>0</td>
</tr>
<tr>
<td>FS-11</td>
<td>0</td>
</tr>
<tr>
<td>FS-12</td>
<td>2</td>
</tr>
<tr>
<td>FS-13</td>
<td></td>
</tr>
<tr>
<td>FS-14</td>
<td>1</td>
</tr>
<tr>
<td>FS-15</td>
<td>3</td>
</tr>
<tr>
<td>FS-16</td>
<td>19</td>
</tr>
<tr>
<td>FS-17</td>
<td>44</td>
</tr>
<tr>
<td>FS-18</td>
<td>32</td>
</tr>
<tr>
<td>FS-19</td>
<td>27</td>
</tr>
<tr>
<td>FS-20</td>
<td></td>
</tr>
<tr>
<td>FS-21</td>
<td></td>
</tr>
<tr>
<td>FS-22</td>
<td></td>
</tr>
</tbody>
</table>

- **(Easy)** R: 45  PR: 6  N: 50
  - NTCIR10-FS-8: \(? a \? x^2 + ? b \? x + ? c\)
- **(Hard)** R: 0  PR: 13  N: 87
  - NTCIR10-FS-10: \(? f \? n(\? z) \? f(\? k)(\? a \? z) 6 = ? c\)

12% relevant formulae, 23% partially-relevant formulae
## Relevance judgment statistics (Formula Search).

<table>
<thead>
<tr>
<th>Query ID</th>
<th>Relevance score</th>
<th>Total judged</th>
<th>Total hit</th>
<th>Uniq ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4 3 2 1 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FS-1</td>
<td>0 1 1 30 69</td>
<td>101</td>
<td>155</td>
<td>0.30</td>
</tr>
<tr>
<td>FS-2</td>
<td>0 0 1 1 102</td>
<td>104</td>
<td>453</td>
<td>0.25</td>
</tr>
<tr>
<td>FS-3</td>
<td>10 3 12 10 66</td>
<td>101</td>
<td>284</td>
<td>0.33</td>
</tr>
<tr>
<td>FS-4</td>
<td>8 6 17 19 52</td>
<td>102</td>
<td>278</td>
<td>0.56</td>
</tr>
<tr>
<td>FS-5</td>
<td>38 0 25 0 38</td>
<td>101</td>
<td>274</td>
<td>0.34</td>
</tr>
<tr>
<td>FS-6</td>
<td>0 0 25 0 77</td>
<td>102</td>
<td>261</td>
<td>0.53</td>
</tr>
<tr>
<td>FS-7</td>
<td>10 0 27 0 68</td>
<td>105</td>
<td>382</td>
<td>0.46</td>
</tr>
<tr>
<td>FS-8</td>
<td>45 0 6 0 50</td>
<td>101</td>
<td>993</td>
<td>0.77</td>
</tr>
<tr>
<td>FS-9</td>
<td>0 0 40 0 63</td>
<td>103</td>
<td>361</td>
<td>0.58</td>
</tr>
<tr>
<td>FS-10</td>
<td>0 0 13 0 87</td>
<td>100</td>
<td>281</td>
<td>0.49</td>
</tr>
<tr>
<td>FS-11</td>
<td>0 0 42 0 58</td>
<td>100</td>
<td>161</td>
<td>0.29</td>
</tr>
<tr>
<td>FS-12</td>
<td>0 0 26 0 74</td>
<td>100</td>
<td>135</td>
<td>0.26</td>
</tr>
<tr>
<td>FS-13</td>
<td>2 0 0 0 98</td>
<td>100</td>
<td>245</td>
<td>0.49</td>
</tr>
<tr>
<td>FS-14</td>
<td>1 0 34 0 65</td>
<td>100</td>
<td>231</td>
<td>0.40</td>
</tr>
<tr>
<td>FS-15</td>
<td>3 0 0 0 98</td>
<td>101</td>
<td>304</td>
<td>0.23</td>
</tr>
<tr>
<td>FS-16</td>
<td>19 0 2 0 81</td>
<td>102</td>
<td>357</td>
<td>0.38</td>
</tr>
<tr>
<td>FS-18</td>
<td>44 0 32 0 28</td>
<td>104</td>
<td>610</td>
<td>0.58</td>
</tr>
<tr>
<td>FS-19</td>
<td>0 0 24 0 76</td>
<td>100</td>
<td>195</td>
<td>0.29</td>
</tr>
<tr>
<td>FS-20</td>
<td>32 0 27 0 41</td>
<td>100</td>
<td>100</td>
<td>0.00</td>
</tr>
<tr>
<td>FS-21</td>
<td>27 0 12 0 61</td>
<td>100</td>
<td>178</td>
<td>0.31</td>
</tr>
<tr>
<td>FS-22</td>
<td>0 0 72 0 29</td>
<td>101</td>
<td>128</td>
<td>0.22</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>239 10 438 60 1,381</strong></td>
<td><strong>2,128</strong></td>
<td><strong>6,496</strong></td>
<td><strong>0.45</strong></td>
</tr>
</tbody>
</table>
Evaluation Measure & Results

- **Formula-based evaluation**: It turned out document-based evaluation cost too much for human assessors!

- **Evaluation measures**:
  - Trec-eval (MAP, P-5, P-10) (similarity-based systems)
  - P-hit: The ratio of the relevant and the submitted hits for all the queries. (matching-based systems)
Math Understanding Task: Dataset & Participants

- **Development Data:** 35 papers selected from ArXiv.org dataset which were also used in Math Retrieval Task.

- **Data for Formal Run:** 10 papers selected from ArXiv.org dataset which were also used in Math Retrieval Task.

- There was only one participant

- The achieved performance was
  - 0.45-0.55 (F1-measure) for strict matching
  - 0.55-0.65 (F1-measure) for soft matching

- The best precision for soft matching was 0.87
Lessons Learnt & Conclusions
Four Types of Approaches

- **Math-agnostic IR systems:** (BRKLY)
  - Keyword-based search
  - Did not perform well

- **Batch Math processors:** (FSE)
  - Distributed system, but does not use a search index
  - Not suitable for interactive IR

- **Matching/Unification-based Math IR systems:** (KWARC, NAK)
  - Return exact instances of the query

- **Similarity-search Math IR systems:** (MIRMU, MCAT)
  - Return large sets of similar formulae scored by closeness.
  - Query variables are similar to any sub-formula
Analysis of Retrieval Performances

- Without explicitly dealing with mathematical formulae, it is very difficult (impossible?) to achieve high effectiveness
  - Math-agnostic systems did not perform well
- Similarity-search Math IR systems did better overall
  - Especially in terms of MAP
  - Investing into partial matches can be rewarding
- Matching /Unification-based Math IR systems perform better if unanswered topics are discarded
Relating the Subtasks

- Math Search has normally two components:
  - Math Understanding (semantic extraction)
  - Semantic Search

- Math Understanding can therefore be seen as a step towards more intelligent Math Search

- In the Formulae Search Subtask, all groups used a possible baseline (\LaTeX XML) to approximate Math understanding

- In the future, the efforts in creating more sophisticated and effective Math understanding systems can directly feed into Formulae Search enriching the indexing of documents
Conclusions: Achievements

- First time a task dedicated to Math IR was run as part of an evaluation forum
- The NTCIR-9 Pilot Math Task has been successful in creating an experimental platform for conducting Math Retrieval experiments:
  - The development of a new collection of 100K of documents and over 35M of formulae
  - The definition of 2 natural Math Search Subtasks
  - The development of a reusable relevance assessment system for Math Tasks
Conclusions: Assessment of the state

- A great deal of work has been done in the first NTCIR-10 Math Pilot Task:
  - Identification of reasonable baselines
  - Shaping the details of the tackled Math Subtasks
  - Facilitating the formation of a pluri-disciplinary community of researchers

- A great deal of work is still needed:
  - Refinement of the topic development process (e.g. easy vs hard topics)
  - Conducting an inter-assessor agreement study
  - Developing common/standard baselines
  - Converging perspectives from the two main types of participating groups: IR scientists and Mathematicians/Logicians
Conclusions: Assessment of the state

- A great deal of work has been done in the first NTCIR-10 Math Pilot Task:
  - Identification of reasonable baselines
  - Shaping the details of the tackled Math Subtasks
  - Facilitating the formation of a pluri-disciplinary community of researchers

- A great more deal of work is still needed:
  - Refinement of the topic development process (e.g. easy vs hard topics)
  - Conducting an inter-assessor agreement study
  - Developing common/standard baselines
  - ... Converging perspectives from the two main types of participating groups: IR scientists and Mathematicians/Logicians
Looking Forward

- It is our intention to run a new iteration of Math IR Task in NTCIR-11
  - Using the same created collection in NTCIR-10
  - Focussing more on the Formulae Search Subtask
  - Achieving a reusable test collection for the Formulae Search Subtask
  - Developing the Math Understanding Subtask

- ...Growing and supporting the Math IR community

Visit the new community portal: https://trac.mathweb.org/NTCIR-Math/

Mathematical Markup Language (MathML) version 3.0.

Andrea Asperti, Ferruccio Guidi, Claudio Sacerdoti Coen, Enrico Tassi, and Stefano Zacchiroli.

A content based mathematical search engine: Whelp.

Arif Jinha.


Michael Kohlhase, Bogdan A. Matican, and Corneliu C. Prodescu.
MathWebSearch 0.5 – Scaling an Open Formula Search Engine.
Michael Kohlhase and Ioan Sucan.
A search engine for mathematical formulae.

Paul Libbrecht and Erica Melis.
Methods for Access and Retrieval of Mathematical Content in ActiveMath.
http://www.activemath.org/publications/

Peder Olesen Larsen and Markus von Ins.
The rate of growth in scientific publication and the decline in coverage provided by science citation index.

Jozef Misutka and Leo Galambos.
System description: Egomath2 as a tool for mathematical searching on wikipedia.org.

Rajesh Munavalli and Robert Miner.
Mathfind: a math-aware search engine.

Bruce R. Miller and Abdou Youssef.
Technical aspects of the digital library of mathematical functions.

Immanuel Normann.
Extended normalization for e-retrieval of formulae.
2006.