

NTCIR-10 MATH PILOT TASK

The Goal of NTCIR-10 Math Task

- NTCIR Math Task aims at exploring methods for mathematical content access through its task design and the construction of the evaluation dataset.



[Formula]

a mathematical relationship or rule expressed in symbols
(Oxford Dictionary)

In science, a formula is a concise way of expressing information, or a general relationship between quantities.
(Wikipedia)

**INFORMATION
ACCESS TO
MATHEMATICAL
CONTENT**

?

Math Information Access

Representations

Embedded image (png, gif, ...)

$$\log(z_1) + \log(z_2) = \log(z_1 z_2) \ ; \ z_1 + z_2 \geq 0$$

Character sequence (latex source)

$$\log(z_1) + \log(z_2) == \log(z_1 z_2) \ ; \ z_1 + z_2 \geq 0$$

Web-browsable XML

```
<math xmlns='http://www.w3.org/1998/Math/MathML' mathematica:form='TraditionalForm' xmlns:mathematica='http://www.wolfram.com/XML/'>
<semantics>
<mrow>
<mrow>
<mrow>
<mrow>
<mi>log</mi>
<mo>+</mo>
<mi>z</mi>
<mi>+</mi>
<mi>z</mi>
<mi>≥</mi>
0
</mrow>
</mrow>
</mrow>
</mrow>
</semantics>
```

XML for math semantics

```
<annotation-xml encoding='MathML-Content'>
<apply>
<ci>Condition</ci>
<apply>
<eq />
<apply>
<plus />
<apply>
<ln />
<apply>
<ci>Subscript</ci>
<ci>z</ci>
<cn type='integer'>1</cn>
</apply>
</apply>
<apply>
<ln />
<apply>
<ci>Subscript</ci>
```

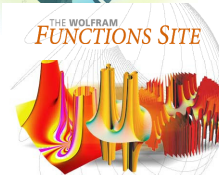
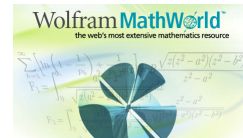
Resources

mathematical knowledge-base and math ontology

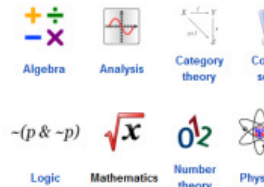


Strict Content MathML (W3C recommendation)

Wolfram MathWorld:
13,081 entries (Sep. 13, 2011)
Wolfram Functions site:
307,409 formulas (Sep. 15, 2011)



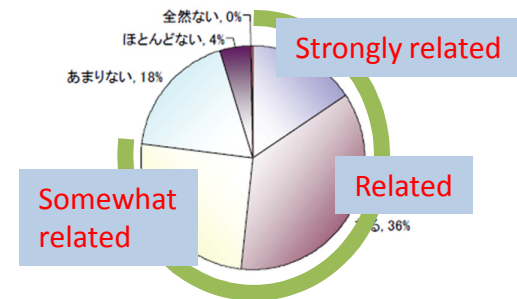
Wikipedia:
26,566 mathematics articles



Requirement

NISTEP Policy Study
Mathematics as deserted science in Japanese S&T policy
- Current situation on mathematical sciences research in major countries and need for mathematical sciences from the science in Japan (2006.5)

Q. Is mathematics related to your research?



77% researchers across diversity of disciplines answered 'YES'.

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.

Math Retrieval Subtask

- **Dataset (scheduled)**

- Scientific Articles from ArXiv e-print server

<http://arxiv.org/>

- Converted into XML+MathML by arXMLiv project

<http://kwarc.info/projects/arXMLiv/>

- 10,000 docs for a dry run, additional 100,000 docs for a formal run

- **Search Types**

- The Math retrieval task uses the above 110K docs and can be envisaged in three different search scenarios

- Formula Search

- Search for formula queries within the formulae database of the used dataset

- Full-Text Search

- Search the document collection using formula queries. Combinations of keywords and formulae.

- Open Information Retrieval

- Search the document collection using free textual queries.

Math Retrieval Subtask : Dataset Example

```

<m:mi id="id57134">S</m:mi>
<m:mi id="id57136">|</m:mi>
<m:mo id="id57138">&prime;</m:mo>
</m:msubsup>
</m:mrow>
<m:mo id="id57141">&ne;</m:mo>
<m:msup id="id57143">
<m:mi id="id57144">G</m:mi>
<m:mo id="id57146">&prime;</m:mo>
</m:msup>
</m:mrow>
<m:mo id="id57149">,</m:mo>
</m:mrow>
<m:annotation-xml id="id57151"
  <m:apply id="id57154">
  <m:neq id="id57155"/>
  <m:apply id="id57156">
  <m:apply id="id57157">
  <m:csymbol id="id57158" cd="ambiguous">superscript</m:
  <m:apply id="id57163">

```

math representation (MathML)

Let $\{G_\gamma \mid \gamma \in \Gamma\}$ be a family of abelian groups. If G_γ is not a proper union of then $G = \bigoplus_{\gamma \in \Gamma} G_\gamma$ is also not a proper union of finitely many cosets.

Demonstration Proof

To prove it by transfinite induction we have two cases to distinguish. If Γ is with some Γ' and for Γ' the statement is true. Then we get $G = G_\gamma \oplus G'$, with coset of G with respect to a subgroup H such that $b + G_\gamma \subseteq S$ with some b of form $S = G_\gamma + S'$, where S' is a proper coset of G' .

Suppose that G is a proper union of the cosets S_1, \dots, S_n . If S_1 contains a coset written as $G_\gamma + S'_1$, otherwise, S'_1 is the empty set. By induction

$$\bigcup_{i=1}^n S'_i \neq G',$$

therefore, there is a $d \in G'$, such that $d + G_\gamma$ is not contained in any S'_i . Moreover then it contains an $r_1 + d$ and $S_1 = r_1 + d + G_1$, where $r_1 \in G_\gamma$ and G_1 is a subgroup

$$S_1 \cap (d + G_\gamma) = (r_1 + d + G_1) \cap (r_1 + d + G_\gamma) = (r_1 + d) + G_1 \cap G_\gamma$$

and


imply that G_γ is a proper union of some of the cosets $r_1 + (G_1 \cap G_\gamma)$, which contradicts the induction hypothesis.

In the second case Γ is a limit ordinal. For a $\Gamma' < \Gamma$ set

$$G_{\Gamma'} = \bigoplus_{\alpha \in \Gamma'} G_\alpha.$$

Assuming G is a proper union of the cosets T_1, \dots, T_k we obtain

$$G_{\Gamma'} = \bigcup_{i=1}^k (G_{\Gamma'} \cap T_i).$$


Cornell University Library
We gratefully acknowledge supporting institutions

arXiv.org > math > arXiv:0801.0652
Search or Article-id (Help | Advanced search)

Mathematics > Rings and Algebras

Covering theorems for Artinian rings

A. Borbely, V. Bovdi, B. Brindza, T. Krausz

(Submitted on 4 Jan 2008)

The covering properties of Artinian rings which depend on their additive structure only, are investigated.

Comments: 5 pages
Subjects: Rings and Algebras (math.RA)
MSC classes: 16P20
Journal reference: Publ. Math. Debrecen, 51/3-4, 1997, p.323-329
Cite as: arXiv:0801.0652v1 [math.RA]

Submission history
From: Victor Bovdi [view email]
[v1] Fri, 4 Jan 2008 10:27:23 GMT (5kb)

Which authors of this paper are endorsers?

Link back to: arXiv, form interface, contact.

Download:

- Source
- PDF
- PostScript
- Other formats

Current browse context:

math.RA
< prev | next >
new | recent | 0801









Change to browse by:

math

References & Citations

- NASA ADS

Bookmark (what is this?)

papers from arXiv.org

xhtml/xml



Math Retrieval Subtask :

Query example

Formula search

1.

$$\int_{\boxed{l}}^{\boxed{h}} \boxed{f}(x)^2 dx$$

2.

$$\boxed{a} + \boxed{b} = \boxed{b} + \boxed{a}$$

3.

$$\lim_{\boxed{x} \rightarrow \infty} \boxed{a} + \boxed{b} = \boxed{c}$$

Full-text search

1.

Pythagorean Theorem

2.

Bell curve in the form of

$$\frac{1}{\boxed{\sigma}\sqrt{2\pi}} \exp^{-\frac{(\boxed{x}-\boxed{\mu})^2}{2\boxed{\sigma}^2}}$$

3.

NOT Bayes Theorem,
but written as

$$P(\boxed{X} | \boxed{Y})$$

Open IR

1.

What is the fifth
summand in the
Taylor expansion of
sinus hyperbolicus?

2.

For which n and
 k is $PSL(n,k)$ not
commutative.?

Math Retrieval Subtask : Formula Search Example

Query formula

The screenshot shows a search interface with a blue header containing navigation links: Questions, Activity, Sign in, Books, Articles, and MWS Engine BETA. A search input field contains the LaTeX code $\lim_{x \rightarrow 0} y$. Below the input field, a dropdown menu is open, displaying several search suggestions:

- Examples - LaTeX queries
- Generic subscript search
- Specific subscript search
- Specific integral search
- Physical constant search
- All limits approaching zero

The suggestion "All limits approaching zero" is highlighted in yellow. Below the suggestions, there is a search button labeled "Search" and a footer that says "Powered by Vanilla".

Retrieved formula

$$\chi(t, t_w) = \lim_{h_0 \rightarrow 0} \frac{m[h](t)}{h_0}. \quad (8)$$

Then we have that

$$\chi(t, t_w) = \beta \int_{t_w}^t dt' X[C(t, t')] \frac{\partial C(t, t')}{\partial t'}, \quad (9)$$

and by performing the change of variables $u = C(t, t')$ we finally obtain the key equation

$$\chi(t, t_w) = \beta \int_{C(t, t_w)}^1 du X(u), \quad (10)$$

where we have used the fact that $C(t, t) \equiv 1$ in Ising models.

It can be easily extracted simply measuring the integrated response to a small external field

$$w) = S[C(t, t_w)], \quad (11)$$

$$= \int_C^1 du X(u). \quad (12)$$

which is encoded in the shape of the function $S(C)$.

The violation factor is equal to one and the relation becomes

$$C(t, t_w) \text{ or } S(C) = 1 - C. \quad (13)$$

2.2 LINK between the statics and the dynamics

To get information on the thermodynamical properties of the model we should match the violation factor $X(C)$ to some static observable. This can be done using the following

Math Understanding Subtask

- **Task definition**
 - Extract natural language definitions of mathematical expressions in a document (Basic Task) with their semantic interpretation (Challenge Task, TBA).
- **Dataset (scheduled)**
 - Development Data
 - 10 papers selected from ACL-Anthology Reference Corpus
 - 30 papers selected from ArXiv.org dataset which will be also used in Math Retrieval Task.
 - Data for Formal Run (submission period: five days)
 - 10 papers selected from ACL-Anthology Reference Corpus
 - 10 papers selected from ArXiv.org dataset which will be also used in Math Retrieval Task.

Math Understanding Subtask : Dataset Example

Original text (pdf)

The similarity between G_1 and G_2 is then defined as follows:

$$groupSim(G_1, G_2) = RI(G_1, G_2) \times RC(G_1, G_2)$$

where

$$RI(G_1, G_2) = \frac{2AI(G_1, G_2)}{II(G_1) + II(G_2)}$$

is the relative interconnectivity and

$$RC(G_1, G_2) = \frac{AC(G_1, G_2)}{\frac{|G_1|}{|G_1| + |G_2|} IC(G_1) + \frac{|G_2|}{|G_1| + |G_2|} IC(G_2)}$$

is the relative closeness.

Example of Presentation MathML

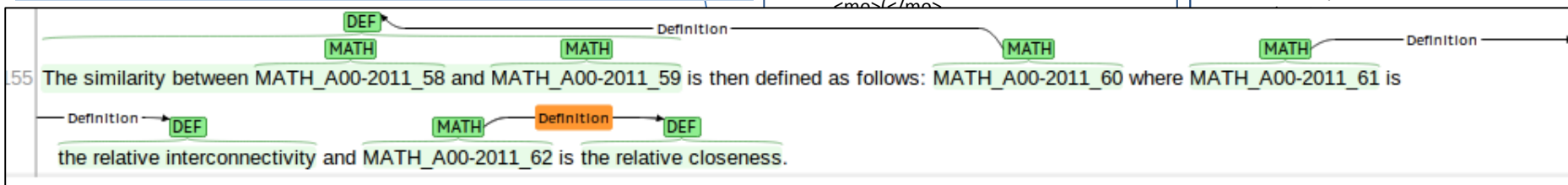
```
<mrow>
  <mi mathvariant="italic">groupSim</mi>
  <mo>?</mo>
  <mrow>
    <mo></mo>
    <mrow>
      <msub>
        <mi>G</mi>
        <mn>1</mn>
      </msub>
      <mo>,</mo>
      <msub>
        <mi>G</mi>
        <mn>2</mn>
      </msub>
    </mrow>
  </mrow>
  <mo>=</mo>
  <mrow id="math_b6a1eb64.2">
    <mrow>
      <mi mathvariant="italic">RI</mi>
      <mo>?</mo>
      <mrow>
        <msub>
          <mi>G</mi>
          <mn>1</mn>
        </msub>
        <mo>,</mo>
        <msub>
          <mi>G</mi>
          <mn>2</mn>
        </msub>
      </mrow>
    </mrow>
  </mrow>
```

MathML parallel markup

Example of Content MathML

```
<annotation-xml encoding="MathML-Content">
  <apply>
    <eq />
    <apply xref="math_b6a1eb64.1">
      <ci>groupSim</ci>
      <apply>
        <selector />
        <ci>G</ci>
        <cn>1</cn>
      </apply>
      <apply>
        <selector />
        <ci>G</ci>
        <cn>2</cn>
      </apply>
    </apply>
    <apply xref="math_b6a1eb64.2">
      <times />
      <apply>
        <ci>RI</ci>
        <apply>
          <selector />
          <ci>G</ci>
          <cn>1</cn>
        </apply>
        <apply>
          <selector />
          <ci>G</ci>
          <cn>2</cn>
        </apply>
      </apply>
    </apply>
```

Definitions of mathematical objects in text



* Annotation tool (<http://www.nactem.ac.uk/brat-annotation/>)

```
<mi>G</mi>
<mn>2</mn>
</msub>
</mrow>
<mo>)</mo>
```

Schedule

February-April, 2012	Task framework development
April, 2012	Call for participation
May, 2012	Dataset and example topics release
Early-mid October, 2012	Topic release for Open IR search type
Late October, 2012	Topic release for Formula, Full-text search types, and Math understanding subtask
Early November, 2012	Results submissions due for Formula, Full-text search types, and Math Understanding subtask
Mid November, 2012	Results submissions due for Open IR search type
February, 2013	Evaluation Results Released
March, 2013	Draft papers for NTCIR-10 Proceedings Due
May, 2013	Camera ready for NTCIR-10 Proceedings Due
June, 2013	NTCIR-10 Meeting

Task information

- **Contact**
 - ntcir10adm-math@nii.ac.jp
- **Task Web page**
 - <http://ntcir-math.nii.ac.jp/>
- **Task Organizers**
 - Akiko Aizawa (National Institute of Informatics, Japan)
 - Michael Kohlhase (Jacobs University Bremen)
 - Iadh Ounis (University of Glasgow)
- **Task Advisors**
 - Noriko Kando (National Institute of Informatics, Japan)
 - Fredric C. Gey (University of California, Berkeley)