NTCIR-11 Math-2

O What is Math Search?

O Goal of NTCIR-Math

O Task outline

O Technical challenges

O Contact information

What is Math Search?

INFORMATION ACCESS TO MATHEMATICAL CONTENT

$$\sum_{n=1}^{\infty} \frac{\sin(n)}{n}$$

infinite series conditionally convergent

NTCIR Math-1
Topic FT-6

What is "math" ?

```
<title>Conditional Convergence</title>
<query>
  <TeXquery>¥sum _{{n=1}}^{{¥infty}}¥frac{¥sin(n)}{n}</TeXquery>
  <pquery>
    <m:math>
      <m:mrow xml:id="m8.1.5.pmml" xref="m8.1.5">
        <m:msubsup xml:id="m8.1.5.1.pmml" xref="m8.1.5.1">
          <m:mo xml:id="m8.1.1.pmml" xref="m8.1.1">∑</m:mo>
          <m:mrow xml:id="m8.1.2.1.pmml" xref="m8.1.2.1">
            <m:mi xml:id="m8.1.2.1.1.pmml" xref="m8.1.2.1.1">n</m:mi>
            <m:mo xml:id="m8.1.2.1.2.pmml" xref="m8.1.2.1.2">=</m:mo>
            <m:mn xml:id="m8.1.2.1.3.pmml" xref="m8.1.2.1.3">1</m:mn>
          </m:mrow>
          <m:mi mathvariant="normal" xml:id="m8.1.3.1.pmml" xref="m8.1.3.1">∞</m:mi>
        </m:msubsup>
        <m:mfrac xml:id="m8.1.4.pmml" xref="m8.1.4">
          <m:mrow xml:id="m8.1.4.2.pmml" xref="m8.1.4.2">
            <m:mi xml:id="m8.1.4.2.1.pmml" xref="m8.1.4.2.1">sin</m:mi>
            <m:mo xml:id="m8.1.4.2a.pmml" xref="m8.1.4.2"></m:mo>
            <m:mrow xml:id="m8.1.4.2b.pmml" xref="m8.1.4.2">
              <m:mo xml:id="m8.1.4.2c.pmml" xref="m8.1.4.2">(</m:mo>
              <m:mi xml:id="m8.1.4.2.3.pmml" xref="m8.1.4.2.3">n</m:mi>
              <m:mo xml:id="m8.1.4.2d.pmml" xref="m8.1.4.2">)</m:mo>
            </m:mrow>
          </m:mrow>
          <m:mi xml:id="m8.1.4.3.pmml" xref="m8.1.4.3">n</m:mi>
        </m:mfrac>
      </m:mrow>
    </m:math>
  </pquery>
  <cquery>
    <m:math>
      <m:apply xml:id="m8.1.5" xref="m8.1.5.pmml">
        <m:apply xml:id="m8.1.5.1" xref="m8.1.5.1.pmml">
          <m:csymbol cd="ambiguous" xml:id="m8.1.5.1.1">superscript</m:csymbol>
          <m:apply xml:id="m8.1.5.1.2">
            <m:csymbol cd="ambiguous" xml:id="m8.1.5.1.2.1">subscript</m:csymbol>
            <m:sum xml:id="m8.1.1" xref="m8.1.1.pmml"/>
            <m:apply xml:id="m8.1.2.1" xref="m8.1.2.1.pmml">
              <m:eq xml:id="m8.1.2.1.2" xref="m8.1.2.1.2.pmml"/>
              <m:ci xml:id="m8.1.2.1.1" xref="m8.1.2.1.1.pmml">n</m:ci>
              <m:cn type="integer" xml:id="m8.1.2.1.3" xref="m8.1.2.1.3.pmml">1</m:cn>
            </m:apply>
          </m:apply>
          <m:infinity xml:id="m8.1.3.1" xref="m8.1.3.1.pmml"/>
        </m:apply>
        <m:apply xml:id="m8.1.4" xref="m8.1.4.pmml">
          <m:divide xml:id="m8.1.4.1"/>
          <m:apply xml:id="m8.1.4.2" xref="m8.1.4.2.pmml">
            <m:sin xml:id="m8.1.4.2.1" xref="m8.1.4.2.1.pmml"/>
            <m:ci xml:id="m8.1.4.2.3" xref="m8.1.4.2.3.pmml">n</m:ci>
          </m:apply>
          <m:ci xml:id="m8.1.4.3" xref="m8.1.4.3.pmml">n</m:ci>
        </m:apply>
      </m:apply>
    </m:math>
  </cquery>
  <words>
    infinite series conditionally convergent
  </words>
</query>
<relevance>
```

$$\sum_{n=1}^{\infty} \frac{\sin(n)}{n}$$

Structured data with
complex XML representation
called MathML

The Goal of NTCIR Math

- Exploring methods for mathematical content access
- Two major goals of Math-2
  - Achieving reusable test collection
  - Establishing and supporting the Math IR community (mathematicians and IR&NLP researchers)

Task Outline: Dataset

O Given a document collection, retrieve relevant mathematical formulae for a given query.

O Dataset: Reuse and adapt NTCIR-10 dataset
- 100,000 papers from ArXiv.org
- 35,000,000 formulae
- Converted into XML+MathML by arXMLiv project
- Retrieval unit: minimal subsections of ArXiv documents

Task Outline: Assessment

○ Pooling and assessment

- 50 Topics
- Multiple assessment (two) for inter agreement check
- Pooling size: 100
- Include "Relevant, Partially relevant, Non relevant, Can not be assessed"

Task Outline: Topics
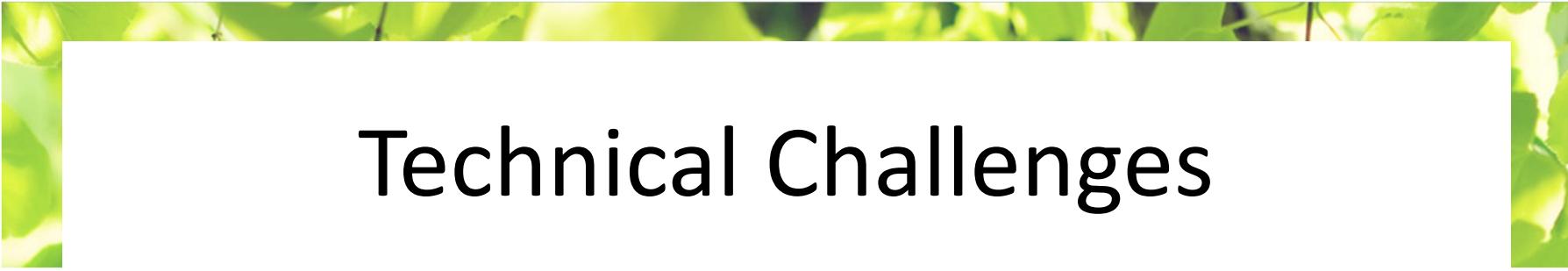
○ Topic development

🍃 Topic structure

- Topic ID
- Query (formula + key words)
- Description (short description of what a user is looking for)
- Narrative (precise description of the user situation and information need and relevancy criteria)

🍃 All topics should include multiple relevant documents

Task Outline: Runs

O Runs

  - Submit compulsory automatic runs using query only field

  - Encourage participants to submit manual runs (with manually generated queries)

  - Results will include supporting evidence (formulaID, sentenceID,etc.), optional

Technical Challenges

- Mathematical knowledge management

- Digital math library

- Large scale XML tree search (with variables)

- Semantic search based on deep language analysis

- Domain specific information retrieval

Contact Information

○ **E-mail**
- organizers ML: ntcadm-math@nii.ac.jp
- community ML: ntcir-math@nii.ac.jp

○ **Task Web page**
- http://ntcir-math.nii.ac.jp/

○ **Community Site**
- http://ntcir.mathweb.org/

○ **Task Organizers**
- Akiko Aizawa (National Institute of Informatics, Japan)
- Michael Kohlhase (Jacobs University Bremen)
- Iadh Ounis (University of Glasgow)

Tentative Schedule

September, 2, 2013   NTCIR-11 kick-off meeting

November, 2013   Call for participation

February, 2014   Initial dataset and example topics release

May, 2014   Topic release

June, 2014   Result submission deadline

December, 2014   NTCIR-11

PLEASE JOIN!