Topics for the NTCIR-10 Math Task
Full-Text Search Queries

Michael Kohlhase (Editor)
Jacobs University
`http://kwarc.info/kohlhase`

March 10, 2013

Abstract

This document presents the formats of the challenge queries and result for the Math Information Retrieval Subtask in the NTCIR-10 Math Task.

Contents

Chapter 1

Introduction

This document presents the challenge queries for the Math Information Retrieval (MIR) Subtask in the NTCIR Math Task. Participants have received the NTCIR-MIR dataset which contains 100 000 XHTML full texts of articles from the arXiv. Formulae are marked up as MathML (presentation markup with annotated content markup and LaTeX source).

1.1 Subtasks

The MIR Subtask in NTCIR-10 has three challenges; queries are given in chapters 2-4 below, the format is

Formula Search (Automated) Participating IR systems obtain a list of queries consisting of formulae (possibly) with wildcards (query variables) and return for every query an ordered list of XPointer identifiers of formulae claimed to match the query, plus possible supporting evidence (e.g. a substitution for query variables).

Full-Text Search (Automated) This is like formula search above, only that IR results are ordered lists of "hits" (i.e. XPointer references into the documents with a highlighted result fragments plus supporting evidence[1]).   EdN:1

Open Information Retrieval (Semi-Automated) In contrast to the first two challenges, where the systems are run in batch-mode (i.e. without human intervention), in this one mathematicians will challenge the (human) participants to find specific information in a document corpus via human-readable descriptions (natural language text), which are translated by the participants to their IR systems.

In all three cases results will be judged on precision, recall, results ranking and (if present) score.

1.2 Query Formats

Queries for the NTCIR-10 MIR Subtask come in two forms: a human-readable description in chapters 2-4 below, and corresponding machine-readable XML files (see `NTCIR-Math-formula-search.xml`, `NTCIR-Math-fulltext-search.xml`, and `NTCIR-Math-open-mir.xml`. The general form of queries is given in Figure 1.1 (see appendix A.1 for a RelaxNG schema): each query has an identifier ⟪typ-num⟫, type ⟪type⟫ (one of "Formula Search Query", "Full Text Query", or "Open MIR") and a title ⟪title⟫ (this should be irrelevant to the query). The query itself consists of a formula (for formula and full-text search; given as LaTeX source, presentation-, and content MathML) and a word list (only for full-text search). For Open MIR queries, the query only consists of a natural language text in the narrative element.

---

[1]EDNOTE: Really, or only the identifiers?

```
<topic xmlns="http://ntcir−math.nii.ac.jp/">
    <num>NTCIR10−《typ-num》</num>
    <type>《type》</type>
    <title>《title》</title>
    <query>
        <narrative>《description》</narrative>
        <TeXquery>《LATEX expressions》</TeXquery>
        <pquery>《Presentation MathML Formulae》</pquery>
        <cquery>《Content MathML Formulae》</cquery>
        <words>《word list》</words>
    </query>
    <relevance>《text》</relevance>
</topic>
```

Figure 1.1: Machine-Readable form of Queries

1.3 Reporting Results

Participants report the results of up to four "runs" of their search engine with the given queries
(see Section 1.2) over the dataset supplied. A "run" is your system outcome for a given approach.
In general, you can imagine a run as the outcome of a specific variant of your system testing a
specific hypothesis.

Results of these runs are reported by e-mailing a results file to `ntcir10adm-math@nii.ac.jp`.
Please name your file as 《group-id》.《ext》 and decorate every result with a 《runtag》 of the form
《group-id》_《run-id》, where 《group-id》 is the group identifier you have chosen upon NTCIR-10
registration and 《run-id》 can be chosen freely. Results can be reported in two forms depending
whether they have justifications.

1.3.1 Results with Justifications

Results with justifications should be reported in an XML file structured as in Figure 1.2 (see
Appendix A.2 for a RelaxNG schema). For each run there is a run element whose runtag attribute
specifies the run identifier as discussed above. For each query there is a result element, whose
for attribute identifies the query. It contains a list of hits identified by a URI reference 《URIref》
is a pointer into the data set of the form 《file URI》#《IDRef》, where 《file URI》 is the file in the
dataset and 《IDRef》 the identifier of the subformula that constitutes the hit. For instance `file:`
`///NTICR-sandbox/108/f042851.xhtml#id54300` is the identifier of the integral subformula in
equation (1) in the document `NTICR-sandbox/108/f042851.xhtml`. Note that URIs without
fragment identifier are also admissible, but arguably less precise, and the omission of the fragment
identifier will be penalized. A hit can be (optionally) further specified: For each wildcard (query
variable with name 《name》) in the query the instance of the hit can be given in a qvar element,
where the 《URIref》 points to the sub-formula that instantiates the query variable. Note that the
justifications will not be judged per se, but will make life of the evaluators easier.

1.3.2 Simple Results

Simple results to formula and full-text queries can be reported as lists of five-tuples (one line per
hit)

NTCIR10−《typ-num》 1 《URIref》 《rank》 《score》 《runtag》

where 《typ-num》 and 《URIref》, 《score》, and 《runtag》 are as above. 《rank》 gives the rank of the
"hit" in the answers to a particular query.

```
<results xmlns="http://ntcir−math.nii.ac.jp/">
  <run runtag="⟪runtag⟫">
    <result for="NTCIR10−⟪typ-num⟫">
      <hit id="⟪id⟫" xref="⟪URIref⟫" score="⟪score⟫">
        <qvar name="⟪name⟫" subst="⟪URIref⟫"/>
        <justification>⟪text⟫</justification>
      </hit>
    <result>
  </run>
</results>
```

Figure 1.2: Returning Justified MIR Results

1.3.3 Results for Open MIR Queries

As the results of Open Information Retrieval Queries are produced by first generating a query (either manually or with machine support) and then running this query, the results must also report the queries (there may be multiple). Therefore, the results must be reported in an XML file that has the form in Figure 1.3, where the elements are as in Figures 1.1 and 1.2. Note that as above the query components are optional.

```
<results xmlns="http://ntcir−math.nii.ac.jp/">
  <result for="NTCIR10−⟪typ-num⟫">
    <query>
      <TeXquery>⟪LaTeX expressions⟫</TeXquery>
      <pquery>⟪Presentation MathML Formulae⟫</pquery>
      <cquery>⟪Content MathML Formulae⟫</cquery>
      <words>⟪word list⟫</words>
    </query>
    <hit id="⟪id⟫" xref="⟪URIref⟫" score="⟪score⟫">
      <qvar name="⟪name⟫" subst="⟪URIref⟫"/>
      <justification>⟪text⟫</justification>
    </hit>
  <result>
</results>
```

Figure 1.3: Returning Justified MIR Results

4

Chapter 2

Formula Search (Automated)

NTCIR10-FS-1: Multiple integral computation, axis commutativity formula (Formula Search)

Query $\int_0^\infty \mathrm{d}x \int_x^\infty F(x,y)\mathrm{d}y = \int_0^\infty \mathrm{d}y \int_0^y F(x,y)\mathrm{d}x$

Relevance A query to find if such a formula.

NTCIR10-FS-2: Fourier Transform examples (Formula Search)

Query $X(?i\omega)$

Relevance A simple query to find Fourier Transform examples (in the typical notation). Note that j may be used instead of i.

NTCIR10-FS-3: Fermat Theorem and related (Formula Search)

Query $x^n + y^n = z^n$

Relevance This will find documents regarding Fermat's Theorem and related exercises.

NTCIR10-FS-4: Integral value (Formula Search)

Query $\int_{-\infty}^\infty e^{-?x^2}\, d?x$

Relevance Find the value or an approximation for the given integral.

NTCIR10-FS-5: Derivative approximation (Formula Search)

Query $\frac{f(?x+?h)-f(?x)}{?h}$

Relevance Determine if this is a valid derivative approximation.

NTCIR10-FS-6: Historical approximation formula (Formula Search)

Query $\sqrt{2} = 1 + \frac{1}{3} + ?x? - ?y$

Relevance   Determine the historical approximation of $\sqrt{2}$.

NTCIR10-FS-7: Variable matching (Formula Search)

Query $sin(?x)/?x$

Relevance   Find an occurrence of the famous expression.

NTCIR10-FS-8: Quadratic equation (Formula Search)

Query $?a?x^2 + ?b?x + ?c$

Relevance   Find a quadratic equation.

NTCIR10-FS-9: Multiple variable matching (Formula Search)

Query $\frac{e^{?x} + ?y}{?z}$

Relevance   The subtlety is that x, y, z can be alpha-renamed or fully substituted, while e is a constant.

NTCIR10-FS-10:   (Formula Search)

Query $?f^{?}n(?z)?f^{(?k)}(?a?z) \neq ?c$

NTCIR10-FS-11:   (Formula Search)

Query $\int_{?g \neq 0} |\nabla ?f|^q d?x \leq ?c \int_{?g \neq 0} |\nabla(?f + ?g)|^? q d?x$

Relevance   here f is actually a function, $f^{(k)}$ is the k-th derivative of f, z is the variable and a and c are constants.

NTCIR10-FS-12:   (Formula Search)

Query $?q?n|?a?n - ?a| \sim_{?n \to +\infty} ?q?n|\frac{?p?n}{?q?n} - ?a|$

NTCIR10-FS-13: Nevanlinna counting function (Formula Search)

Query $N_{?k)}(?r, \frac{1}{?f - ?a})$

NTCIR10-FS-14: Notation for second derivative on time (Formula Search)

Query $\ddot{?u}(?x, ?t) = ?u''(?x, ?t)$ $?x \in \mathbb{R}$

Relevance   Here we really have to check that the two occurrences of $?x$ are the same.

NTCIR10-FS-15: Weierstrass' $\wp$ Function (Formula Search)

Query $\wp(?z; \Lambda)$

NTCIR10-FS-16: Weierstrass' $\wp$ Function (Formula Search)

Query $\wp(?z; \omega_1, \omega_2)$

NTCIR10-FS-17: Complex Integral (Formula Search)

Query $\int_{-\infty}^{\infty} (-?a?x^2) e^{-2\pi x} dx$

NTCIR10-FS-18: Landau Sets (Formula Search)

Query $O(?n \log ?n)$

NTCIR10-FS-19: Integral (Formula Search)

Query $Rf(L) = \int_L f(\mathbf{x}) |d\mathbf{x}|$

NTCIR10-FS-20: Index of a Subgroup (Formula Search)

Query $|G : H| = \frac{|G|}{|H|}$

NTCIR10-FS-21: Cohomology Group of a Surface (Formula Search)

Query $H^n(X) = Z^n(X)/B^n(X)$

NTCIR10-FS-22: Homology Group (Formula Search)

Query $?A_n = \frac{1}{\pi} \int_{-\pi}^{\pi} ?F(x) \cos(nx) dx$

Chapter 3

Full Text Search (Automated)

NTCIR10-FT-1: Looking for points where the derivative of $\wp$ vanishes (Full Text Search)

Query $\wp$

Words   Points derivative vanishes

Relevance   The hits of this query should give information about zeroes of the derivative of Weiterstrass $\wp$ function.

NTCIR10-FT-2: Excluding Keywords from Formula Search (Full Text Search)

Query $\int_{?b}^{?a} ? f^2(x) dx$

Words   NOT(Parseval)

Relevance   Parseval's theorem should not be in the set of hits

NTCIR10-FT-3: LaTeX Pseudocode (Full Text Search)

Narrative   Find all articles with equations or inequalities that aren't marked up with LaTeX

Words   =, \geq, \leq

Relevance   but only return examples like "let N= 7477= 61*127" (not in dollar signs).

NTCIR10-FT-4: Divergence (Full Text Search)

Query $\prod_{N=1}^{\infty}(1 + Z/N)\ Z \in C$

Words   diverges diverge

Relevance   The "open problem" at `http://planetmath.org/?op=getmsg&id=23757` could use a hint.

NTCIR10-FT-5: Radius of Convergence (Full Text Search)

Query $\sum \frac{n!x^n}{n^n}$

Words radius of convergence

Relevance The open problem at `http://planetmath.org/?op=getmsg&id=23642` could use a hint.

NTCIR10-FT-6: Conditional Convergence (Full Text Search)

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

Words infinite series conditionally convergent

Relevance The open problem at `http://planetmath.org/?op=getmsg&id=23572` could use a hint.

NTCIR10-FT-7: Roots of a Polynomial (Full Text Search)

Query $8x^3 + 4x^2 - 4x - 1$

Words root

Relevance The open problem at `http://planetmath.org/?op=getmsg;id=22097` could use a hint. Can we match this to $x^3 + ax^2 + bx + c = 0$?

NTCIR10-FT-8: Polynomial Modulo (Full Text Search)

Query $?y^2 = ?x^3 + ?a?x + ?b \ p$

Words mod modulo

NTCIR10-FT-9: p-adic diophantine Equation (Full Text Search)

Query p

Words -adic diophantine equation

NTCIR10-FT-10: Ramsey Number (Full Text Search)

Query $r_k(C_4)$

Words estimated multicolor Ramsey number

NTCIR10-FT-11: Differential Equation (Full Text Search)

Query $?x'(?t) + \sum^? N_{?j=1} B_{?j}(?t)?x(?t - \tau_? j(?t)) = F(?t)$

Words   conditions boundedness

NTCIR10-FT-12: Solutions to Differential Equation (Full Text Search)

Query $\frac{\partial ?u}{\partial ?t} - \triangle ?u + \frac{\langle D^2 ?u D ?u, D ?u \rangle}{1+|D?u|^2} = 0$

Words   uniqueness of solutions

NTCIR10-FT-13: Stability (Full Text Search)

Query $?x_{?k+1} = \frac{A_1}{?x_{?k}^{?p_1}} + \frac{A_2}{?x_{?k-1}^{?p_2}} + \cdots + \frac{A_n}{?x_{?k-?n+1}^{?p_n}}$

Words   stability

NTCIR10-FT-14: Convergence (Full Text Search)

Query $\sum ?p_n ?a_n$

Words   convergence

NTCIR10-FT-15: Differential Equation (Full Text Search)

Query $d?X_t = b(?t, ?X_t)dt + \sigma(t, ?X_t)d?W_t$

Words   solution

Chapter 4

Open Information Retrieval for Mathematics

NTCIR10-OMIR-1: Isomorphic subspace (Open MIR)

Narrative   Let $p \in (1, \infty)$ and let q be conjugate to p. Is there a subspace of $\ell_1(\ell_p)$ isomorphic to ℓ_q?

Relevance   to come

NTCIR10-OMIR-2: Differential Operator (Open MIR)

Narrative   Let Γ be a discrete amenable group. If $\pi \colon \Gamma \to B(\mathcal{H})$ is a unitary representation of Γ on a separable Hilbert space $\mathcal{H}$, is the von Neumann algebra $\pi(\Gamma)''$ necessarily injective?

Relevance   to come

NTCIR10-OMIR-3: Differential Operator (Open MIR)

Narrative   What I know about differential equations with the following differential operator $-\mathrm{div}(a(x)|\nabla u|^{p-2}\nabla u) + h(x)|u|^{r-2}$?

Relevance   to come

NTCIR10-OMIR-4: Bounded Solutions of Differential Equation (Open MIR)

Narrative   When does the second-order vector differential equation $y'' + Cy' + Ay + F(y) = p(t)$ with C, A constant $n \times n$-matrices, $F : R^n \to R^n$ continuous and $p : R \to R^n$ continuous and bounded, have bounded solutions?

Relevance   to come

NTCIR10-OMIR-5: Looking for Proofs (Open MIR)

Narrative   Find occurrences of reductio ad absurdum proofs.

NTCIR10-OMIR-6: Theorem Search (Open MIR)

Narrative   Find an article related to the Four Color Theorem.

NTCIR10-OMIR-7: Associative Operators in Quantum Mechanics (Open MIR)

Narrative   Find all associative Operators in Quantum Mechanics

NTCIR10-OMIR-8: Non-group Commutative Monoids (Open MIR)

Narrative   Find commutative monoids that are not groups.

NTCIR10-OMIR-9: Dependencies (Open MIR)

Narrative   Find all results that depend the perfect graph theorem of László Lovász.

NTCIR10-OMIR-10: Inductive Hypotheses (Open MIR)

Narrative   Find proofs, where an inductive hypothesis is a product of sums.

NTCIR10-OMIR-11: Similar Research (Open MIR)

Narrative   Find all professors that do similar research to Herbert Jaeger (`http://minds.jacobs-university.de/herbert`)

NTCIR10-OMIR-12: Equivalence (Open MIR)

Narrative   Are you sure that "Sequential compactness is equivalent to compactness when X is a metric space."?

Relevance   Question asked at `http://planetmath.org/SequentiallyCompact.html`. An actual answer would require some natural language understanding, searchable access to source material.

NTCIR10-OMIR-13: Analogy (Open MIR)

Narrative   What is to $\coprod$ as Σ is to $\sum$?

Relevance   Question asked at `http://planetmath.org/SequentiallyCompact.html`. An actual answer would require some natural language understanding, searchable access to source material.

NTCIR10-OMIR-14: Finding Sum Representation (Open MIR)

Narrative   Let $S = \sum_j exp^{-aj^2}$. Is there a formula for this sum?

Relevance   Question asked at `http://planetmath.org/?op=getmsg;amp;id=24582`.

NTCIR10-OMIR-15: Finding Zeros of Polynomials (Open MIR)

Narrative   How can one solve polynomial of the form $ax^3 + bx + c$ equals zero?

Relevance   Question asked at `http://planetmath.org/?op=getmsg;amp;id=24463`.

NTCIR10-OMIR-16: Finding Roods of Polynomials (Open MIR)

Narrative   How can we find the roots of the polynomial $n^5 - 7n^4 + 17n^3 - 18n^2 + 7n - 1$?

Relevance   Question asked at `http://planetmath.org/?op=getmsg;amp;id=24330`. The open problem may be sad to know that we can't do that in closed form. Can we convert the question into plain text that would match "the general solution of the fifth degree equation"?

NTCIR10-OMIR-17: Finding Reformulations (Open MIR)

Narrative   Let A be a finite ring. Then any element a of A is either a zero divisor or invertible. Moreover the inverse of a is a power of a.

Relevance   This result is advanced at `http://planetmath.org/?op=getmsg;amp;id=24124`. A respondent points out that this is known, but it is "formulated otherwise". Any chance of recovering the equivalent statement?

NTCIR10-OMIR-18: Calculating Derivatives (Open MIR)

Narrative   If f is the real-valued function $f : R^3 \to R$ with $f(x, y, z) = x^2 + yz$, Calculate df.

Relevance   The question is posed here `http://planetmath.org/?op=getmsg;amp;id=24022`. Can we recognize that the key point is to "calculate df"?

NTCIR10-OMIR-19: Intersections of closed Sets (Open MIR)

Narrative   Let X_n be a decreasing sequence of nonempty closed sets in a Banach space such that their diameter tends to 0. Is their intersection nonempty?

Relevance   The question is posed here `http://planetmath.org/?op=getmsg;amp;id=23877`. Finding the general result for

Appendix A

Appendix

We provide the RelaxNG schemata for the XML query and result formats for convenience, they can be found `https://svn.mathweb.org/repos/NTCIR-Math/topics/lib`.

A.1 RelaxNG Schema for NTCIR Queries

A RelaxNG for NTCIR Topics
$Id: NTCIR - topic.rnc\,32\,2012 - 12 - 03\,08 : 09 : 34Z\,kohlhase$
$HeadURL : https : //svn.mathweb.org/repos/NTCIR - Math/topics/lib/NTCIR - topic.rnc$
(c) 2012 Michael Kohlhase, released under the GNU Public License (GPL)

namespace mws = "http://search.mathweb.org/ns"
namespace m = "http://www.w3.org/1998/Math/MathML"
default namespace ntcir = "http://ntcir−math.nii.ac.jp/"

topics = **element** topics {topic∗, **attribute** xml:id {text}}

num = **element** num {text}
type = **element** type {text}
title = **element** title {inline.model}
narrative = **element** narrative {inline.model}
note = **element** note {inline.model}
relevance = **element** relevance {inline.model}
reference = **element** reference {xsd:anyURI}

query = **element** query {((mquery |(TeXquery? & pquery? & cquery?))? & words? & narrative?}
TeXquery = **element** TeXquery {text}
mquery = **element** mquery {math}
pquery = **element** pquery {math}
cquery = **element** cquery {math}
words = **element** words {inline.model}

qvar = **element** mws:qvar {**attribute** name {text}}

math = **element** m:math {grammar {**include** "mathml3/mathml3.rnc" {**start**=MathExpression}
        PresentationExpression |= parent qvar
        ContExp |= parent qvar}}
            | grammar {

```
            include "LaTeXML/LaTeXML−common.rnc"
            include "LaTeXML/LaTeXML−math.rnc"
            start=Math}

inline.model = (text| inline.class)
inline.class &= grammar {
            include "LaTeXML/LaTeXML−common.rnc"
            include "LaTeXML/LaTeXML−math.rnc"
            include "LaTeXML/LaTeXML−inline.rnc"
            start=Inline.class
            Inline.model = (text| Inline.class)
            Flow.model = (text| Inline.class)}

ntcir−topic = num & type & title? & narrative? & relevance? & reference?

private = element private {examplehit∗ & contributor∗ & note∗}
examplehit = element examplehit {attribute href {xsd:anyURI}}
contributor = element contributor {inline.model}

topic = element topic {ntcir−topic & query & private? }
start = topics
```

A.2 RelaxNG Schema for NTCIR Results

```
# A RelaxNG for NTCIR Results
# Id : NTCIR − results.rnc392013 − 01 − 1409 : 32 : 36Zkohlhase
# HeadURL : https : //svn.mathweb.org/repos/NTCIR − Math/topics/lib/NTCIR − results.rnc
# (c) 2012 Michael Kohlhase, released under the GNU Public License (GPL)

default namespace ntcir = "http://ntcir−math.nii.ac.jp/"

results = element results {run∗}
run = element run {result∗, attribute runtag {text} }
result = element result {hit∗, attribute for {text}}
hit = element hit {qvar∗ & justification? &
                              attribute id {text} &
                              attribute xref {xsd:anyURI} &
                              attribute score {text}?}
qvar = element qvar {attribute name {text} &
                                attribute subst {xsd:anyURI}}
justification = element justification {text}

start = results
```