Math Information Retrieval Happening

Conferences on Intelligent Computer Mathematics, 2012

July 8th, 2012

1 Introduction

This document contains example challenges for the Mathematics Information Retrieval Happening, July 8th, 2012 hosted at CICM 2012, Bremen, Germany.

The challenges were preselected manually by three independent referees. Note that the examples below are not intended to serve as an exhaustive benchmark for the participating systems, neither do they come together with an automated testing framework. The participants are expected to co-create search scenarios together with the judges and explore the practical challenges of Math IR in a relaxed setting.

Notation: The single special notation used that differs from classic TEX/LATEX markup is the use of **?** to designate a "query variable", i.e. a placeholder that can be substituted with arbitrary subformulas by the search engines. We write `?x` and output x for a query variable "x".

2 Evaluation Tasks

This section contains the official MIR2012 challenges, each of which has a designated article expected to be retrieved from the MIR2012 sandbox.

2.1 Formula Search (Automated)

Challenge 2.1.1. Recollect a historical formula, such as:

TEX	`\sqrt{2} = 1 + \frac{1}{3} + \frac{1}{3\dot 4} ?- \frac{1}{3\dot 4\dot 34}}`
Math	$\sqrt{2} = 1 + \frac{1}{3} + \frac{1}{34} - \frac{1}{3\dot{4}34}$
Example:	http://arxmliv.kwarc.info/files/1010/1010.4331/1010.4331.xhtml
Sandbox:	f005795.xhtml

But was the last operator a plus(+) or a minus(-) sign ?

Judge: Dr. Patrick Ion

Challenge 2.1.2. Retrieve instances matching:

TEX	`B_{p+n} = B_n + B_{n+1} \bmod p \ \text{for all}\ n=0,1,2,\dots`
Math	$B_{p+n} = B_n + B_{n+1} \pmod p$ for all $n = 0, 1, 2, \dots$
Example:	http://arxmliv.kwarc.info/files/1008/1008.1573/1008.1573.xhtml
Sandbox:	f005794.xhtml

Judge: Dr. Patrick Ion

Challenge 2.1.3. Find examples of the use of the below metric:

TEX	`S(g) = \frac{s(g)-s_{\text{min}}}{s_{\text{max}}-s_{\text{min}}}`
Math	$S(g) = \frac{s(g)-s_{\min}}{s_{\max}-s_{\min}}$
Example:	http://arxmliv.kwarc.info/files/1203/1203.5158/1203.5158.xhtml
Sandbox:	f005796.xhtml

Judge: Dr. Patrick Ion

Challenge 2.1.4. Find Cardy's formula:

TEX	`\frac{3\Gamma(2/3)}{\Gamma(1/3)}\eta^{1/3}\,\rule{0pt}{10pt}_2 F_1(1/3,2/3,4/3;\eta)`
Math	$\frac{3\Gamma(2/3)}{\Gamma(1/3)}\eta^{1/3}\,{}_2F_1(1/3,2/3,4/3;\eta)$
Example:	http://arxmliv.kwarc.info/files/0909/0909.4499/0909.4499.xhtml
Sandbox:	f005692.xhtml

Judge: Dr. Daniel Meyer

Challenge 2.1.5. Retrieve instances matching:

TEX	`a?x^2+b?y^2$`
Math	$ax^2 + by^2$
Example:	http://arxmliv.kwarc.info/files/0812/0812.0067/0812.0067.xhtml
Sandbox:	f004977.xhtml

Similarly for $cx^2 + dy^2$, i.e. `c?x^2+d?y^2`

Notes: This is complicated for two reasons.

- The actual variables are x_1 and x_2, not x and y (as it happens, a etc. are the same).
- We actually have $ax_1^2 + bx_2^2 + \epsilon_1 x_1 x_2$, with the possibilities of ϵ_1 being either zero or non-zero (and $cx_1^2 + dx_2^2 + \epsilon_2 x_1 x_2$ similarly).

Judge: Dr. James Davenport

Challenge 2.1.6. Retrieve instances matching:

TEX	`\frac{e^2+3}42^{?l\choose 2}?n^?l`
Math	$\frac{e^2+3}{4}2^{\binom{l}{2}}n^l$
Example:	http://arxmliv.kwarc.info/files/0801/0801.2554/0801.2554.xhtml
Sandbox:	f004150.xhtml

Notes: The subtlety is that n, l are α-convertible, also called "query variables", but e is not, as it is a constant.

Judge: Dr. James Davenport

Challenge 2.1.7. Retrieve instances matching:

TEX	`?P\in \sum_{i=1}^r\Z ?{P_i}`
Math	$P \in \sum_{i=1}^{r} \mathbf{Z} P_i$
Example:	http://arxmliv.kwarc.info/files/0712/0712.3704/0712.3704.xhtml
Sandbox:	f004102.xhtml

Notes: The subtlety is that P and P_i are *independently* α-convertible, i.e. they are distinct "query variables"

Judge: Dr. James Davenport

2.2 Full-Text Search (Automated)

Challenge 2.2.1. Handle the following textual queries:

- Where can I find the formula for free cumulants in terms of the symmetric group?
- Aren't there some newer special polynomials involved?

Example:	http://arxmliv.kwarc.info/files/1010/1010.4331/1010.4331.xhtml
Sandbox:	f005795.xhtml

- Also, Kerov polynomials and zonal polynomials

Example:	http://arxmliv.kwarc.info/files/1005/1005.0316/1005.0316.xhtml
Sandbox:	f005793.xhtml

Judge: Dr. Patrick Ion

2.3 Open Information Retrieval (Semi-Automated)

Challenge 2.3.1. Retrieve instances matching:

TEX	`f_1(x_1,\ldots,x_n)<0\land f_2(x_1,\ldots,x_n)<0`
Math	$f_1(x_1,\ldots,x_n)<0 \wedge f_2(x_1,\ldots,x_n)<0$
Example:	http://arxmliv.kwarc.info/files/0801/0801.0586/0801.0586.xhtml
Sandbox:	f004115.xhtml

or conceivably: $f_1(x_1,\ldots,x_n)<0 \wedge f_2(x_1,\ldots,x_n) \wedge \cdots \wedge f_m(x_1,\ldots,x_n)<0$.

Notes: This is complicated for several reasons. The text talks about "$f_1\sigma_1 0,\ldots,f_m\sigma_m 0$", so one has to

- realise that "," is "$\wedge$";
- infer "$f_1(x_1,\ldots,x_n)$" from "f_1" and the earlier $f_i \in K[x_1,\ldots,x_n]$;
- infer "$f_1<0$" from "$f_1\sigma_1 0$" and the earlier $\sigma \in \{<,=,>\}^m$ (where $\sigma=(\sigma_1,\ldots,\sigma_m)$ is wholly implicit).

In fact, this is a remarkably hard problem, and a related question would be "what mathematically sensible queries *will* retrieve the opening paragraph of this paper?"

Judge: Dr. James Davenport

3 Open-ended Challenges

Find below challenges that the judges found interesting, but for which no evaluation article is known in the MIR2012 sandbox.

3.1 Formula Search (Automated)

Challenge 3.1.1. Discover that:

TEX	`z' = b + z(x-a)`
Math	$z' = b + z(x - a)$
Example:	http://arxmliv.kwarc.info/files/1008/1008.1573/1008.1573.xhtml
Sandbox:	f005794.xhtml

is part of the Roessler system

Judge: Dr. Patrick Ion

3.2 Full-Text Search (Automated)

Challenge 3.2.1. Is the scaling limit of critical percolation conformally invariant?

Judge: Dr. Daniel Meyer

Challenge 3.2.2. Triangular Cauchy-Riemann equations

Judge: Dr. Daniel Meyer

Challenge 3.2.3. I would really like from a math search engine to have simple topological questions answered. Here is an example:

Is a compact Hausdorff space metrizable?

Notes: if and only if it is second countable. (This is Urysohn's theorem).

Judge: Dr. Daniel Meyer

Challenge 3.2.4. Give Conformal map from disk to regular hexagon.

Notes: this can be given (more or less explcit) by the so-called Schwarz-Chritoffel map. Would be curious to know what the engines yields. This can of course be varied in many different ways.

Judge: Dr. Daniel Meyer

3.3 Open Information Retrieval (Semi-Automated)

Challenge 3.3.1. Is the complement of any $S^2 \subset \mathbf{R}^3$ simply connected?

Notes: This is false, as the famous Alexander horned sphere shows. This is a very old result, would be curious to know what the search engines yield. To make the query **really** challenging one could ask if the complement of a quasisphere $S \subset \mathbf{R}^3$ is simply connected. Answer (no) is the same, but I doubt that this is understood.

Judge: Dr. Daniel Meyer

Challenge 3.3.2. Retrieve instances matching the diagram at the top of [Rob08, p. 6].

Notes: The challenge is to get this, but not every commuting diagram (at least, commuting square) in the world!

Judge: Dr. James Davenport

References

[Rob08] L. Robbiano. On Border Basis and Gröbner Basis Schemes. http://arxmliv.kwarc.info/files/0802/0802.2793/0802.2793.xhtml, 2008.