Semantic Preserving Bijective Mappings of Mathematical Formulae between Semantic LaTeX and Computer Algebra Systems

Howard Cohl*, Moritz Schubotz[†], Abdou Youssef[‡], André Greiner-Petter[††], Jürgen Gerhard[‡‡], Bonita V. Saunders*, Marjorie A. McClain*, Joon Bang[§] and Kevin Chen[§]

*Applied and Computational Mathematics Division, NIST, Gaithersburg, Maryland, U.S.A.
[†]Dept. of Computer and Information Science, University of Konstanz, Konstanz, Germany
[‡]Dept. of Computer Science, George Washington University, Washington DC, U.S.A.
[††]Database Systems and Information Management Group, TU-Berlin, Germany
[‡‡]Maplesoft, Waterloo, Ontario, Canada, [§]Poolesville High School, Poolesville, MD, U.S.A.

JMM 2018: Mathematical Information in Digital Age of Science

January 12, 2018

Collecting information on special functions and OP

- 1964: Handbook of Mathematical **Functions** with Formulas, Graphs, and Mathematical Tables
  - Milton Abramowitz and Irene Stegun, editors.
  - 1064 pages (book)
  - definitions, approximations, identities, plots and numerical tables
- 2010: NIST Digital Library of Mathematical **Functions** (DLM**F**)
  - NIST Handbook of Mathematical Functions as successor of A&S
  - F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, editors.
  - 968 pages (printed version), MathML version (online)
  - Links, MathSearch, info boxes, metadata
  - Bruce Miller's LaTeXml package
- 2013: Digital Repository of Mathematical **Formulae** (DRM**F**)
  - Top down semantic information regarding individual formulae
  - Additional Sources
  - *Currently* MediaWiki/Wikitech (Wikimedia Foundation)
  - Community interaction and collaboration

Enhanced semantics of mathematical LaTeX expressions

- Special Functions and Orthogonal polynomials:
  - Trigonometric sine function
  - Euler gamma function
  - Jacobi polynomial
  - little q-Laguerre/Wall polynomial
- Rendered as:
  - $\sin z$, $\Gamma(z)$, $P_n^{(\alpha,\beta)}(x)$, and $p_n(x;a|q)$.
- LaTeX presentations given by
  - `\sin z`, `\Gamma(z)`, `P_n^{(\alpha,\beta)}(x)`, `p_n(x;a|q)`.
- Semantic LaTeX representations (Bruce Miller's macro set)
  - ```
    \sin@@{z}
    \EulerGamma@{z}
    \JacobipolyP{\alpha}{\beta}{n}@{x}
    \littleqLaguerre{n}@{x}{a}{q}
    ```
    ```

Bruce Miller's OPSF DLMF macro set

- Provides semantic content in formulas
- DLMF OPSF Macros via LaTeXML-server
  - 546 semantic DLMF/DRMF LaTeX OPSF macros
  - additional 180 semantic DRMF LaTeX macros
- Objects: `\sum,\int,\deriv{f}{x},\qderiv[n]{q}@{z}`
- Constants: `\expe,\iunit,\cpi,\EulerConstant`
- Special Functions and Orthogonal Polynomials

$\Gamma(z)$	`\EulerGamma@{z}`	`http://dlmf.nist.gov/5.30#E1`
$J_\nu(z)$	`\BesselJ{\nu}@{z}`	`http://dlmf.nist.gov/10.2#E2`
$Q^\mu_\nu(z)$	`\assLegendreQ[\mu]{\nu}@{z}`:	`http://dlmf.nist.gov/14.3#E7`
$P^{(\alpha,\beta)}_n(x)$	`\JacobipolyP{\alpha}{\beta}{n}@{x}`	`http://dlmf.nist.gov/18.3#T1.t1.`

Semantic macro classification

685 semantic DLMF/DRMF LaTeX macros (Bruce Miller's development)

- 395 macros for real and complex valued functions
- 185 macros for polynomials (orthogonal and whatnot)
- 29 macros for integer valued functions
- 40 macros for various semantic operators
- 14 macros for quantifiers, set operators and symbols
- 9 macros for sets of numbers
- 5 macros for constants
- 5 macros for linear algebra
- 3 macros for distribution theory

Information associated with macros

For each macro we store:

- The name of the macro
- Example calling sequence
- Example rendering of called semantic macro
- Concise Description
- Summary and description of calling options
- Link to url giving precise definition (either in the DLMF or to a definition page on the DRMF).

OpenMath Content Dictionaries

- Bruce Miller will publish content dictionary for DLMF macros
- Currently developing content dictionary for extra DRMF macros

Computer Algebra System (CAS) seeding and translation

One can effectively take advantage of semantic macro set to:

- seed the DRMF using CAS libraries
  - Use data produced in CAS languages to produce semantic LaTeX for inclusion in Digital Mathematics Libraries using LaTeXml
  - Two examples `eCF` and `CFSF` datasets
- Bijective translation to and from CAS source representations and semantic LaTeX
  - Provide CAS output source to user. Given semantic LaTeX source for a mathematical expression, assuming that the necessary semantic information is provided, then we can use tools such as Youssef's POM tagger:

    A. Youssef, Lecture Notes in Computer Science **10383**, Springer, 2017, pp. 243-257,

    and; Greiner-Petter's CAS translation tools.
  - Perform round-trip testing and verification of formulae. (powerful tool for digital mathematics libraries.)

Abdou Youssef's LaTeX POM tagger (first scan)

- Examines **terms** and groups them into **sub-expressions** when indicated, e.g., `\frac{1}{2}` is a sub-expression of numerator and denominator.
- A **term** (in the sense of Backus-Naur) is a pre-defined non-terminal expression and can represent LaTeX macros, environments, reserved symbols (e.g., line break command \\[0.2cm]) or numerical or alphanumerical expressions.
- **Sub-expressions** and terms get tagged due the first scan of the tagger, with two separate tag categories:
  1. definite tags (such as operation, function, exponent, etc.) that the tagger is certain of;
  2. tags which consist of alternative and tentative features which include alternative roles and meanings. (drawn from a knowledge base which has been collected for the tagger).

Greiner-Petter's tool to translate semantic LaTeX

- used mathematical language parser (MLP) as an interface for the first scan of the POM tagger to build syntax trees of mathematical expressions in LaTeX and provide CAS translations.
- Macro information is stored in CSV files.
- CAS translations for Greek letters, mathematical constants, and built-in commands for mathematical functions are stored separately in JSON files.
- Maple to semantic LaTeX translations use Maple syntax and various representations using their OpenMaple API.
- Round trip tests start from a valid semantic LaTeX expression or from a valid Maple expression. CAS translations from start representation to the other representation is called one cycle. Round trip tests reach a fixed point when the string representation is identical to its previous representations. We reached a fixed point in semantic LaTeX after one cycle and in Maple after $1\frac{1}{2}$ cycles in most cases we tried (useful in identifying syntax errors, since the CAS translation fails.)

Seeding for eCF continued fraction dataset

- Wolfram eCF dataset: **Wolfram** language source (**4,095** lines)
  Encoding Continued Fraction Knowledge in Computational Form in Wolfram|Alpha, Sloane funded *Wolfram Computational Knowledge of Continued Fractions Project* (Provided by Eric Weisstein, Wolfram)
  1365 eCF formulae in semantic LaTeX

```
ConditionalExpression[E^z==1+(2*z)/(2-z+z^2/(6*(1+Inactive[ContinuedFractionK]
[z^2/(4*(1+2*k)*(3+2*k)),1,{k,1,Infinity}]))),Element[z,Complexes]]
```

Our code produces the following semantic LaTeX

```
\expe^{z}=1+\frac{2z}{2-z+\frac{z^{2}}{6\left(1+\CFK{k}{1}{\infty}@@
{\frac{z^{2}} {4\left(1+2k\right)\left(3+2k\right)}} {1}\right)}}
```

This produces the following rendered formula

$$e^z = 1 + \frac{2z}{2 - z + \frac{z^2}{6\left(1 + \mathop{\mathrm{K}}_{k=1}^{\infty}\left(\frac{\frac{z^2}{4(1+2k)(3+2k)}}{1}\right)\right)}}.$$

Seeding for CFSF continued fraction datasets

- Maple CFSF dataset: **Maplesoft** source (**2,646** lines)
  Maple Continued Fractions for Special Functions Library (Provided by Annie Cuyt, Franky Backeljauw, Stefan Becuwe, University of Antwerp, Belgium) connected with the book *Handbook of Continued Fractions for Special Functions* (2008)
  - 252 CFSF formulae from 55 files located in 10 subdirectories.
  - Produced semantic LaTeX file with 10 sections and 55 subsections.

```
create( 'contfrac',}
  label = "EF.exp.sfrac.01",}
  booklabel = "11.1.2",}
  dlmflabel = "4.9.3",}
  front = 1,}
  begin = [[2*z, 2-z], [z^2/6, 1]],}
  general = [[(1/(4*(2*m-3)*(2*m-1)))*z^2,1]],}
  function = exp,}
  lhs = exp(z),}
  category = "S-fraction"}
):
```

Seeding for CFSF continued fraction datasets (cont.)

For the example create statement we generate the semantic LaTeX:

```
\expe^{z}=1+\frac{2z}{2-z}\subplus\frac{\frac{z^{2}}{6}}{1}\subplu
\CFK{m}{3}{\infty}@@{\frac{1}{4\left(2m-3\right)\left(2m-1\right)}
```

which when rendered by pdflatex produces the formula

$$e^z = 1 + \frac{2z}{2-z} + \frac{\frac{z^2}{6}}{1} + \mathop{\mathrm{K}}_{m=3}^{\infty}\left(\frac{\frac{1}{4(2m-3)(2m-1)}z^2}{1}.\right)$$

which is an alternate representation of

$$e^z = 1 + \frac{2z}{2 - z + \frac{z^2}{6\left(1+\mathop{\mathrm{K}}_{k=1}^{\infty}\left(\frac{\frac{z^2}{4(1+2k)(3+2k)}}{1}\right)\right)}}.$$

Extraction of 9897 DLMF semantic LaTeX formulae

- Identified all relevant (36) LaTeX chapter files.
- Extracted all `DLMF` formulae on September 16th 2016
  (prior to major update to macros naming convention)
- Formulas contained within: {`equation`, `math`, `equationmix`, `equationgroup`, `aligned`, `align`, `displaymath`} environments and mapped to a single line in a text file.
- Extracted `\label` and `\constraint` formula metadata
- Split $\pm$ and $\mp$ formula doubles, e.g.,

$$\tan u \pm \tan v = \frac{\sin(u \pm v)}{\cos u \cos v}$$

- Replicated formulae with multiple equal signs, e.g.,

$$\mathsf{P}_2(x) = P_2(x) = \frac{3x^2 - 1}{2}$$

Refinement of DLMF formulae

- Reduced to 9897→4165 formulas. Removed formulas containing
  - no semantic information
    ```
    z = (\tfrac{3}{2}\zeta)^{2/3} \label{eq:AI.RL.AIZ}
    ```
  - integrals,
    ```
    \int \sinh@@{x} \diff{x} = \cosh@@{x}
    ```
  - sums,
    ```
    \sum_{\ell=0}^n \LaguerrepolyL[\alpha]{\ell}@{x}
          = \LaguerrepolyL[\alpha+1]{n}@{x}
    ```
  - products,
    ```
    \JordanJ{k}@{n} = n^k \prod_{p \divides n} (1 - p^{-k})
    ```
  - asymptotic expressions,
    ```
    \expe^{\iunit x} = \BigO@{1}
    ```
  - expressions containing ellipsis,
    ```
    \EulerGamma@{\tfrac{1}{3}}
           = 2.67893\;85347\;07747\;63365\;\dots,
    ```
  - matrices, case statements, and lattices.

Evaluation techniques

- Translate semantic LaTeX formula.
- Identify left-hand side (LHS) and right-hand side (RHS)
- Simplify
- Compute and compare LHS, RHS, and verify whether LHS-RHS=0
- Discovered sign error in DLMF formula (14.5.14)

$$\mathsf{Q}_{\nu}^{-1/2}(\cos\theta) = -\left(\frac{\pi}{2\sin\theta}\right)^{1/2}\frac{\cos((\nu+\frac{1}{2})\theta)}{\nu+\frac{1}{2}}$$

Maple testing of DLMF formulae - August 2016

- Translated each test case in Maple
- Forward translation was able to translate 2232 formulae (approx. 53.6%)
- Verified 477 of these
- Pre-conversion improved effectiveness of simplify
- Pre-conversion increased verified to 662 and
  1570 test cases were translated but not verified
- Used conversions to exponential and hypergeometric form
- Remaining 1993 test cases were not translated because
  - contain DLMF/DRMF macros without known translation to Maple (987 cases) such as q-hypergeometric function
  - or an error appeared during the translation or verification process
- Increased verification by restricting to reduced sectors in the complex plane. Sometimes due to inaccuracies produced by CAS assumptions: e.g., $\sqrt{-1} := i, \quad \sqrt{1/z} \neq 1/\sqrt{z}$.

Applications: automated CAS testing for DLMF formulae

- **Symbolic** verification/simplification tests for formulae.
  - difference (LHS-RHS) and ratio (LHS/RHS) tests:

$$f(x) - g(x) = 0, \qquad \frac{f(x)}{g(x)} = 1.$$

  - Simplification (often fails).
  - Compare Taylor series coefficients (identifty variables).
- **Numerical** verification
  - Search for DLMF formulae errors using difference (LHS-RHS) and ratio (LHS/RHS) tests. Ratio breaks at roots of f, g. Use increasing precision. Use a distribution of points in the complex plane. Need ranges of variables. DLMF in principle provides this metadata.
- **Provide** CAS representations to users for usage, testing, examination.

Semantic enhancement via CAS representation

- Were able to enhance the semantics of 74 Wronskian $\mathcal{W}(f,g) = f'g - fg'$ relations for special function by rewriting the Wronskian macro to include the variable that is differentiated with respect to.
- 186 further cases in which the DLMF uses primes for derivatives.
- Build easy-to-use semantic macros which encapsulates missing information and presents the math as we wish
- integrals, sums, products, limits,
- ...
- This is an ongoing process to maximize content in mathematical expression written in LaTeX which contains sufficient semantic information to translate between computer algebra systems. (Why?) Because mathematicians (like us) like to write in LaTeX,
  ...and we like to know that our mathematics is correct and error free.