

# Content dictionary description: select symbols from Chapter 9 of the KLS dataset in the DRMF

Moritz Schubotz

Dept. of Computer and Information Science,  
University of Konstanz, Konstanz, Germany,  
`moritz.schubotz@uni-konstanz.de`

Howard S. Cohl

Applied and Computational Mathematics Division,  
National Institute of Standards and Technology,  
Gaithersburg, Maryland, U.S.A.,  
`howard.cohl@nist.gov`

## Abstract

The NIST Digital Repository of Mathematical Formulae (DRMF) uses parallel MATHML markup to express mathematical formulae. However currently, the content MATHML symbols do not point to a content dictionary. In this paper, we describe the naming convention and creation process for mathematical symbols in the DRMF. Moreover, we publish a first, experimental version of a content dictionary that contains only a few manually selected symbols from Chapter 9 of the KLS dataset.

## 1 Introduction

The NIST Digital Repository of Mathematical Formulae [Coh+14] is an online compendium of mathematical formulae data for orthogonal polynomials and special functions. The initial formula data was obtained from the NIST Digital Library of Mathematical Functions [Olv+17]. DLMF uses semantic macros [Mil16], which are converted by  $\LaTeX$ ML [Mil10] to parallel markup including content MATHML. We refer to the symbols generated by  $\LaTeX$ ML as implicit content dictionaries [Sch16]. However, the DRMF does not only contain formula-data from the DLMF. In 2015, we started to extend our library with sources that do not use the DLMF macro schema [Coh+15]. In this process, we introduced new symbols and associated  $\LaTeX$  macros. In this content dictionary, we release an initial version of the symbols introduced while processing Chapter 9 of the KLS dataset [KLS10]. Currently, we are extending our work on Mathematical Language Processing [Sch+16; Coh+17]. A content dictionary is beneficial for this task, since it specifies the semantics in a standard conforming way.

## 2 Symbol Naming Conventions

We use the same naming convention for  $\LaTeX$  macro names and the content dictionary symbols. The symbols follow simple naming conventions that we will discuss in this section. The goal for our naming convention is to have short, memorable and unambiguous names. Currently, the DRMF website uses 689 custom  $\LaTeX$  macros.

---

*Copyright © by the paper's authors. Copying permitted for private and academic purposes. Portions of this article were prepared by a US Government employee as part of his official duties. Official contributions of the NIST are not subject to copyright in the United States.*

In: H. Geuvers, M. England, O. Hasan, F. Rabe, O. Teschke (eds.): Proceedings of the OpenMath Workshop, Edinburgh, UK, 17-July-2017, published at <http://ceur-ws.org>

Table 1: Abbreviations used in the macros for Chapter 9 in the KLS dataset.

Abbreviation	Normal form
cts	continuous
norm	normalized

Their length varies between 1 and 26 characters and the median length is 8 characters. The name is derived from a short symbol description using the following conventions:

**concatenation** Individual elements are concatenated without spaces. Camel case is not used.

**capitalization** Individual’s names are capitalized. Mathematical identifiers such as  $q$  and standard macros such as `log`, `Log` are used as normal capitalization. Properties such as continuous, monic, generalized, normalized are written in lower case.

**abbreviations** Only a limited set of abbreviations is used. See Table 1.

### 3 New Symbols

Extensibility is one of the key goals of the DRMF project. Consequently, we developed a mechanism to update our custom macros or add new macros. Adding a macro follows this procedure:

1. Identify a symbol and its definition in the literature.
2. Describe the symbol with a short name but without ambiguities. This description can include properties, individuals names (cf. Section 2) as well as presentational features.
3. Derive the symbol name based on the description:
  - Remove whitespace.
  - Replace presentational features with their verbalization, e.g.,  $\$ \tilde{p} \$ \mapsto p \tilde{d} e$ .
  - Replace abbreviations.
  - Remove superfluous words like “polynomial”. (Our convention to drop the word “polynomial” for this dataset is because these are the most abundant objects.)
4. Add a definition or a longer description of the symbol.
5. Add references to the literature.
6. Update the DRMF  $\text{\LaTeX}$  style file locally and define calling sequences, that is, the different numbers of parameters which can be used for the same macro name. For example the macro `normWilsonWtilde` can be called with 5 or with only one parameters/arguments.
7. Update the macro definitions for the DRMF  $\text{\LaTeX}$ ML server.
8. Add a Wikipage in the namespace “Definition” with the name of the symbol that:
  - describes the calling sequences;
  - lists the definition;
  - shows the rendering; and
  - gives references to the literature.
9. Link the new wikipage on the symbol definition overview.

To rename a symbol, the DRMF style file has to be adjusted and to be uploaded again to the server. The definition page has to be moved manually. In contrast, all occurrences of the macro can be replaced using the search and replace MediaWiki plugin. This will trigger automated updates for the MathSearch index [Sch17].

## 4 Outlook and Future Works

Currently, the content dictionary is a proof of concept. For the future, we are planning to automate the creation of content dictionaries for other chapters and sources. Chapter 14 of KLS is already in preparation.

### References

- [Coh+14] H. S. Cohl, M. A. McClain, et al. “Digital Repository of Mathematical Formulae”. In: *Intelligent Computer Mathematics, Lecture Notes in Artificial Intelligence 8543*. Ed. by S. M. Watt, J. H. Davenport, A. P. Sexton, P. Sojka, J. Urban. Vol. 8543. LNCS. Springer, 2014, pp. 419–422. DOI: 10.1007/978-3-319-08434-3\_30.
- [Coh+15] H. S. Cohl, M. Schubotz, et al. “Growing the Digital Repository of Mathematical Formulae with Generic  $\text{\LaTeX}$  Sources”. In: *Intelligent Computer Mathematics, Lecture Notes in Artificial Intelligence 9150*. Ed. by M. Kerber, J. Carette, et al. Vol. 9150. LNCS. Springer, 2015, pp. 280–287. DOI: 10.1007/978-3-319-20615-8\_18.
- [Coh+17] H. S. Cohl, M. Schubotz, et al. “Semantic Preserving Bijective Mappings of Mathematical Formulae between Document Preparation Systems and Computer Algebra Systems”. In: *10th Conference on Intelligent Computer Mathematics*. 2017.
- [KLS10] R. Koekoek, P. A. Lesky, and R. F. Swarttouw. *Hypergeometric Orthogonal Polynomials and Their  $q$ -Analogues*. Ed. by T. H. Koornwinder. Springer Monographs in Mathematics. With a foreword by Tom H. Koornwinder. Berlin [u.a.]: Springer-Verlag, 2010, pp. xx+578. XIX, 578. DOI: 10.1007/978-3-642-05014-5.
- [Mil10] B. Miller. *LaTeXML: A  $\text{\LaTeX}$  to XML Converter*. 2010. URL: <http://dlmf.nist.gov/LaTeXML/>.
- [Mil16] B. Miller. “Drafting DLMF Content Dictionaries”. English. Talk presented at the OpenMath workshop of the 9th Conference on Intelligent Computer Mathematics CICM 2016. Bialystok, Poland, July 25, 2016. URL: <http://cicm-conference.org/2016/cicm.php?event=&menu=talks#02> (visited on 10/03/2016).
- [Olv+17] F. Olver, A. Olde Daalhuis, et al., eds. *NIST Digital Library of Mathematical Functions*. <http://dlmf.nist.gov/>, Release 1.0.14 of 2017-12-21. F.W.J. Olver, A.B. Olde Daalhuis, D.W. Lozier, B.I. Schneider, R.F. Boisvert, C.W. Clark, B.R. Miller and B.V. Saunders, eds. 2017. URL: <http://dlmf.nist.gov>.
- [Sch+16] M. Schubotz, A. Grigorev, et al. “Semantification of Identifiers in Mathematics for Better Math Information Retrieval”. In: *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR ’16. Pisa, Italy: ACM, 2016, pp. 135–144. DOI: 10.1145/2911451.2911503.
- [Sch16] M. Schubotz. “Implicit Content Dictionaries in the NIST Digital Repository of Mathematical Formulae”. English. Talk presented at the OpenMath workshop of the 9th Conference on Intelligent Computer Mathematics CICM 2016. Bialystok, Poland, July 25, 2016. URL: <http://cicm-conference.org/2016/cicm.php?event=&menu=talks#03> (visited on 10/03/2016).
- [Sch17] M. Schubotz. *Augmenting Mathematical Formulae for More Effective Querying & Efficient Presentation*. Epubli Verlag, Berlin, 2017. URL: <https://www.epubli.de/preview/publication/64471>.

## Listing of the Content Dictionary

```
1 <CD xmlns="http://www.openmath.org/OpenMathCD">
2   <CDComment>
3
4     This document is distributed in the hope that it will be useful,
5     but WITHOUT ANY WARRANTY; without even the implied warranty of
6     MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.
7
8     The copyright holder grants you permission to redistribute this
9     document freely as a verbatim copy. Furthermore, the copyright
10    holder permits you to develop any derived work from this document
11    provided that the following conditions are met.
12    a) The derived work acknowledges the fact that it is derived from
13    this document, and maintains a prominent reference in the
14    work to the original source.
15    b) The fact that the derived work is not the original OpenMath
16    document is stated prominently in the derived work. Moreover if
17    both this document and the derived work are Content Dictionaries
18    then the derived work must include a different CDName element,
19    chosen so that it cannot be confused with any works adopted by
20    the OpenMath Society. In particular, if there is a Content
21    Dictionary Group whose name is, for example, 'math' containing
22    Content Dictionaries named 'math1', 'math2' etc., then you should
23    not name a derived Content Dictionary 'mathN' where N is an integer.
24    However you are free to name it 'private_mathN' or some such. This
25    is because the names 'mathN' may be used by the OpenMath Society
26    for future extensions.
27    c) The derived work is distributed under terms that allow the
28    compilation of derived works, but keep paragraphs a) and b)
29    intact. The simplest way to do this is to distribute the derived
30    work under the OpenMath license, but this is not a requirement.
31    If you have questions about this license please contact the OpenMath
32    society at http://www.openmath.org.
33  </CDComment>
34
35  <CDName>drmf-cls9 </CDName>
36  <CDBase>https://cd.formulasearchengine.com/</CDBase>
37  <CDURL>https://cd.formulasearchengine.com/drmf-cls9.ocd</CDURL>
38  <CDReviewDate>2018-07-10</CDReviewDate>
39  <CDStatus>experimental</CDStatus>
40  <CDDate>2018-07-10</CDDate>
41  <CDVersion>0</CDVersion>
42  <CDRevision>0</CDRevision>
43
44  <Description>
45    This CD defines symbols from KLS Chapter 9.
46  </Description>
47
48  <CDDefinition>
49    <Name>normWilsonWtilde</Name>
50    <Role>application</Role>
51    <Description>
52      normalized Wilson polynomial  $W_{\tilde{W}}$ 
53    </Description>
54  </CDDefinition>
55  <CDDefinition>
56    <Name>monicWilson</Name>
57    <Role>application</Role>
58    <Description>
59      monic Wilson polynomial
60    </Description>
61  </CDDefinition>
62  <CDDefinition>
63    <Name>monicRacah</Name>
64    <Role>application</Role>
65    <Description>
66      monic Racah polynomial
67    </Description>
68  </CDDefinition>
69  <CDDefinition>
70    <Name>normctsHahnptilde</Name>
71    <Role>application</Role>
72    <Description>
73      normalized continuous Hahn polynomial  $p_{\tilde{p}}$ 
74    </Description>
75  </CDDefinition>
76 </CD>
```

Listing 1: The content dictionary