Authoring and Publishing Units and Quantities in Semantic Documents

Mihai Cîrlănaru, Deyan Ginev, Christoph Lange

Computer Science, Jacobs University Bremen, Germany
{m.cirlanaru,d.ginev,ch.lange}@jacobs-university.de

Abstract. This paper shows how an explicit representation of units and quantities can improve the experience of semantically published documents, and provides a first authoring method in this respect. To exemplify the potential and practical advantages of encoding explicit semantics regarding units w.r.t. user experience, we demonstrate a *unit system preference* service, which enables the user to choose the system of units for the displayed paper. By semantically publishing units, we obtain a basis for a wide range of applications and services such as *unknown unit lookup*, *unit and quantity semantic search* and *unit and quantity manipulation*. Enabling semantic publishing for units is also presented in the context of a large collection of legacy scientific documents (the arXMLiv corpus), where our approach allows to non-invasively enrich legacy publications.

1 Motivation

Units and quantities, although widely spread, lack a formal standard representation for semantic publishing. A multitude of problems [Usm] arise from the different flavors (country specific unit standards) and formats (abbreviations, special cases of occurrence) of units, making it hard for the untrained reader to fully understand the information provided. Semantic publishing solves most such problems by disambiguating the unit and quantity occurrences, which, further on, will enable a wide range of applications and services to interact with them.

A **unit** is *any determinate quantity, dimension, or magnitude adopted as a basis or standard of measurement for other quantities of the same kind and in terms of which their magnitude is calculated or expressed* [Oxf], but from the top-most level of perception, it simply provides information on a wide range of quantifiable aspects. Concrete examples for the great extent of units and quantities include cooking recipes, medical prescriptions, scientific papers and many other. Semantic publishing can provide the middle layer that would ensure a (automated) way of identifying and understanding these occurrences which can enable the evolution of useful technologies and services.

At the perception level, aside from quantifying properties and relations between objects, units bring the meaning of scale. Moreover, units have allowed scientists to better transmit and exchange knowledge among themselves.

In real life, the misinterpretation of units and their quantities has often caused accidents with harsh/expensive consequences. Consider losing a $125 million satellite [Mar] because of the differences between metric and imperial unit systems, or running out of fuel in mid-flight with an aircraft whose fuel sensors were faultily configured in displaying the units [Air]. Fields like medicine, commerce, civil engineering have also been marked by such types of errors and pitfalls [Usm]. This simply emphasizes the fact that units are frequently misinterpreted.

Providing semantics to units and their quantities for the publishing industry, either by supplying semantic authoring tools or by semantically enriching their occurrences in legacy documents, has high impact benefits. It will enable transparent exchange of scientific knowledge between different academic communities, typical of technical papers with high occurrence of units and quantities, and also enhance the reader's experience, via novel interactive services with day-to-day published material, e.g. cooking recipes or technical manuals.

In the following sections the preliminaries (section 2) and state of the art (section 4) for *units and quantities* are introduced in order to have a basis for the *unit and quantity interaction services* (section 6) presented in this paper. We outline immediate strategies (section 5) for extending the benefits of semantic units to legacy documents (section 7) and conclude with a summary of our mid-term outlook of future work (section 8).

2 Preliminaries

The core of semantic publishing resides in open and standardized markup languages used to encapsulate semantics. OpenMath and *Content* MathML are the most widely used semantic markup (also called "content markup") languages for mathematical expressions, which are ubiquitous in science and engineering.

2.1 OpenMath and Content MathML

OpenMath [Bus+04] and the semantically equivalent Content MathML [Aus+10] are standards for the representing the semantics of mathematical expressions [KR09] – as annotations to visual renderings, or for the purpose of communication between computational services. Our investigations focus on these two languages.[1]

Structurally, both OpenMath and MathML provide a valuable basis for machine processing of mathematical expressions; they are ideal markup languages for the purpose of semantic publishing of units and quantities. The expressivity of MathML, provided by its vocabulary having close to 100 XML elements for

[1] The prevalence of XML-based semantic markup languages for representing mathematical expressions – as opposed to RDF – has historical reasons but is also due to the complex n-ary and ordered structures of mathematical expressions, which are hard to break down into RDF triples. In both representations the vocabulary terms (here: functions, operators, sets, constants) are identified by URIs. We refer to [Lan11] for an in-depth treatment.

functions and operators for mathematics [KR09] and multiple *unit and quantity* representation possibilities [DN03], and the modularity and extensibility of OpenMath's vocabulary by way of modular ontologies ("Content Dictionaries", abbreviated as CDs), enable the development of applications and services (some of which are discussed in section 6.2) that build upon the semantic publishing of units and quantities.

2.2 The Semantic Publishing Pipeline

Semantic Publishing, as a process, consists of at least three components, namely *authoring*, *publishing* and *interaction*. Usually these processes imply three different groups of contributors – authors, publishers and readers. Incorporating the full publishing lifecycle into a single system, striving for integration and collaboration between the different participants, brings great benefits. In this paper, we take the benefits of the social web for well-established[2] and accepted and focus on the more novel semantic aspects of the publishing realm. To this extent, we develop our work in the context of the Planetary eMath3.0 system (see [Koh+11] for an introduction). Notably, the Planetary framework implements the architecture introduced in [Dav+10] for publishing its documents as XHTML+RDFa+MathML, enabling interactive semantic services.

In our work on units and quantities, we have concentrated on setting the necessary technological foundation, hence building on the languages introduced in section 2.1 to select and enhance the authoring and interaction aspects.

3 Semantic Units – Idea Outline

In order to understand how a semantic representation of units and quantities will integrate with the publishing flow of our framework of choice, one first needs to pinpoint what they comprise and how they are *represented*.

A computational *semantic entity* is an object with explicit *structure*, representable in a machine-understandable form, and denoting a corresponding real-world entity. The denotation is usually encoded via a machine-readable ontology. This definition is directly applicable to semantic units and quantities, which are exactly the machine-readable representations of their physical counterparts.

For the *representation* we choose OpenMath, since it encompasses units through modular ontologies, called Content Dictionaries (CDs) [Col09], which enable extensibility through the creation of new such ontologies that can add new symbols or simply through the extension of the existing unit ontologies/CDs.

[2] For mathematics, including the mathematical foundations of science and engineering, see, e.g., the PlanetMath free encyclopedia [Pla] and the Polymath wiki/blog-based collaboration effort [Bar10].

As a running example for this paper, we consider a semantic representation of the physical ***quantity*** 100 km/h; one possibility to represent it in OpenMath is[3]:

```
<OMA>
    <OMS cd="arith1" name="times"/>
    <OMI>100</OMI>
    <OMA>
       <OMS cd="arith1" name="divide"/>
       <OMA>
         <OMS cd="units_ops1" name="prefix"/>
         <OMS cd="units_siprefix1" name="kilo"/>
         <OMS cd="units_metric1" name="metre"/>
       </OMA>
       <OMS cd="units_time1" name="hour"/>
    </OMA>
</OMA>
```

Listing 1.1. OpenMath representation of 100 km/h

4 State of the Art

We review the relevant prior work involving units and quantities in the context of semantic publishing. Note that we do not cover the publishing dimension itself, as it is a stand-alone framework level, independent of the processed content.

4.1 Representation

The semantic publishing aspect of units in scientific documents has not yet accumulated a sizable body of prior work. Previous research has been mainly concerned with the standardization of unit and quantity representation which is far from complete (not covering every unit occurrence possibility) or sufficiently machine comprehensible. There is a number of units-related semantic web ontologies: The authors of the Measurement Units Ontology [BP09] review a number of ways of representing units in RDF. The SWEET ontology (Semantic Web Earth and Environmental Terminology [Swe; RP05]) is particularly remarkable for linking units to the fields of science where they occur. A general weakness of RDF/OWL unit ontologies is, however, that the computation of derived units (and thus unit conversion) cannot be described in a straightforward way (and is rarely done).

For OpenMath, a representation of units and quantities has been proposed (cf. [DN03]), and several CDs covering common units have been provided. The

[3] This is one out of several ways of representing units (cf. [DN03]). For a detailed description of the XML schema see section 3.1.2 of [Bus+04]

in-depth analysis of the prospective representations of units and their dimensions that [DN03] proposes (taking into account the pros and cons of each approach) allows for a broader view to the multitude of semantic publishing possibilities. The two most significant sets of OpenMath unit CDs have been developed by James Davenport and Jonathan Stratford [SD08] and Joseph Collins [Col09], respectively. The former are remarkable for their explicit representation of conversion rules (see also Section 4.3). The latter ones provide a standards-compliant implementation of SI[4] quantities and units, providing strong insight on the concepts of *quantity* and *unit* and on the prospects of capturing more of their semantics in the representation.

4.2 Authoring

In "pre-semantic" environments, such as LaTeX, there are first approximations of content-oriented macros that represent units. A prominent example is the LaTeX package *SIunits* [Hel] which covers the full range of base and derived units in the SI system, as well as SI prefixes, a range of widely accepted units external to SI and a couple of generic mechanisms for creating custom author-specified unit constructs. The package enables a large set of abbreviative commands, which are internally built up from the compositional application of atomic building blocks. In this sense, the authoring process via *SIunits* is *nearly semantic* on the interface level, but *entirely presentational* on the output side.

Still, all major semantic authoring systems (e.g. the semantic LaTeX extensions sTeX [Koh08], SALT [Gro+07], the Ontology Add-in for Microsoft Office Word [Fin+10], or the semantic content management system PAUX [PAU]) have so far neglected the specific use case of units. This can be partially explained by the lack of a widely agreed standard representation, as well as different primary development foci – mathematics for sTeX, rhetorical structures for SALT, life sciences terminology for the Word ontology add-in, and educational texts from areas unrelated to physics, such as law, for PAUX. Notably, sTeX could, in principle, support units already, as its wide coverage of the conceptual model of OpenMath and its generic mechanism for defining new symbols and concepts could easily be utilized for the specification of the relevant unit and quantity symbols. Section 5 presents how we have done that in a way that does not disrupt existing LaTeX authoring practices. While LaTeX is commonly used in mathematics, science, and engineering, our solution is unlikely to appeal to life scientists, where Microsoft Office Word is more widely used; however, we leave unit support for word processors to future work.

4.3 Interaction

Applications taking advantage of the semantic publishing of units and their quantities using OpenMath have also been experimented with by various authors, albeit the lack of authoring support. The unit conversion service [Str08; SD08]

[4] The International System of Units [Sib]

by Jonathan Stratford, which users can easily extend by uploading new Content Dictionaries (CDs) with new units and conversion rules, provides a good example of the power of semantically annotated units. Besides the implementation of such a service, Stratford's research also identifies the difficulties of unit conversion and the limitations of OpenMath's current state with regard to unit representation.

Stratford's conversion service is interactive in that users can enter quantities into a web form and upload definitions of new units. We have additionally made it interactively accessible from web documents that contain MathML formulas with OpenMath annotations, as created by the publishing pipeline explained in section 2.2 (cf. [GLR09]). This interaction with units in publications has, however, remained a proof of concept so far, as *producing* suitably annotated documents required manual authoring of quantity expressions in OpenMath XML markup – a barrier that we are trying to overcome with the work presented in this paper.

5 Semantic Authoring of Units and Quantities

We have revised the available methods and technologies and established that a semantic authoring support for units does not formally exist at present. Consequently, we set out to make the first steps towards extending one of the more prepared software solutions, namely sTeX, with a special authoring module for units, by building on the existing pre-semantic toolbox of the *SIunits* LaTeX package. sTeX [Koh08] is essentially a collection of LaTeX packages that offer semantic macros. sTeX can be translated into XML markup using LaTeXML [Mil] bindings, thus enabling easier subsequent processing – including semantic web publishing (cf. [Dav+10]). Our units extension follows a similar approach[5].

As described in section 4.2, *SIunits* provides an sTeX-like content authoring interface. For our running example, we are interested in authoring $\boxed{100\,\mathrm{km/h}}$ in order to create the content representation shown in Listing 1.1. There are many ways to author the representation in LaTeX, e.g. via `$\textrm{100\,km/h}$`. The *SIunits* package makes the process less ad-hoc by focusing on the content and factoring out the presentational quirks, in the form of package options. Hence, one would instead write the more semantic `\unit{100}{\kilo\metre\per\hour}`. It is interesting to observe that a completely different motivation than ours, namely to provide a convenient and centralized interface to control the *presentation* of the unit entities on a document level, leads to the essentially same result which we desire – a *semantics-oriented* authoring interface.

In our effort to leverage this functionality, we first created a LaTeXML binding for the *SIunits* package. It helped us to pinpoint the semantic map between the interface and the OpenMath representation and provided a non-invasive semantic enrichment for LaTeX documents based on the package. Next, we use the gained understanding in building a native sTeX module for units, roughly

[5] The SIunits bindings and sTeX extension will be released in the respective bundles (the arXMLiv binding library and the sTeX package on CTAN) with the authors' strong committment to free software licenses compatible with the originals.

based on the *SIunits* interface. Table 1 shows a small snippet comparing the different stages. One easily notices the abbreviative power of the sTEX approach, which hides the verbose and overly complex binding declaration under its hood, exposing the author to a controlled LaTeX vocabulary and facilitating reuse.

<table>
<tr><th>Language</th><th>Definition</th><th>Semantics</th></tr>
<tr><td>LaTeX</td><td><pre>\newcommand{\kilo}{\ensuremath{\mathrm{k}}}
\newcommand{\metre}{\ensuremath{\mathrm{m}}}</pre></td><td>✗</td></tr>
<tr><td>LaTeXML</td><td><pre>DefConstructor('\kilo{}','
<ltx:XMApp>
    <ltx:XMTok meaning="prefix" cd="units_ops1"/>
    <ltx:XMTok meaning="kilo" cd="units_siprefix1">
       k
    </ltx:XMTok>
    #1
</ltx:XMApp>');
DefConstructor('\metre','
<ltx:XMTok meaning="metre" cd="units_metric1">
    m
</ltx:XMTok>');</pre></td><td>✓</td></tr>
<tr><td>sTEX</td><td><pre>\symdef[name=kilo,cd=units_siprefix1]{kiloPX}{\mathrm{k}}
\symdef[name=metre,cd=units_metric1]{metre}{\mathrm{m}}
\symdef[name=prefix,cd=units_siprefix1]{prefixFN}{}
\symdef{kilo}[1]{\mixfixii{}{\kiloPX}{\prefixFN}{#1}{}}</pre></td><td>✓</td></tr>
</table>

Table 1. Definitions for \kilo\metre, typeset as 'km'

6 Interaction with Units and Quantities

Given the provisions for authoring support, we move to the added-value benefits one could reap from interacting with a published document. This section details relevant use cases and explains the prerequisites that are already available.

6.1 Unit (System) Preference Service

A concrete scenario for a prospective service that would take advantage of semantically published papers, based on the ideas from section 3, can be evolved on top of common published material like *cooking recipes*. These provide a good use case thanks to the high density of units and quantities they contain. Moreover, the physical quantities are restricted to a small subset (quantity/mass related units) including special types of *units* [21c] which are not formally defined and might prove to be misleading:

$$1\ \textbf{teaspoon (tsp)} \approx 5\ \textit{millilitres (mL)}$$
$$1\ \textbf{cup} \approx 250\ \textit{millilitres (mL)}$$

The idea of the *unit (system) preference* service is to allow the user/reader to choose a preferred system of units (e.g. imperial, metric) or simply preferred types of units (e.g. "minutes" instead of "hours", "kilogrammes" instead of "grammes") for the representation of physical quantities and then seamlessly adapt the document to these preferences. This can only be achieved at the end of the semantic publishing pipeline, since the process requires the technologies described in sections 2 and 4 for the *representation* and *authoring* parts. Once these prerequisites have been met, one can embed interactive scripts into the published document (here: XHTML with OpenMath-annotated MathML formulae), which invoke a web service for any computation. In our implementation, the JOBAD (Javascript API for OMDoc-based Active Documents) framework [GLR09] provides for client-server communication and manipulation of the document. Figure 1 visualizes the workflow.

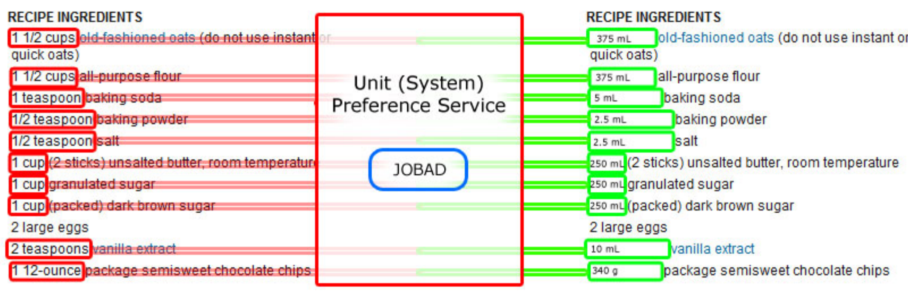

Fig. 1. Workflow for Chocolate Chip Cookies recipe [Crc]

6.2 Prospective Services based on Semantically Published Units

Having described in detail one service that enhances the user experience by publishing units semantically, we now list further potential services and applications that the same technology could enable:

- **Mapping Natural Sciences Concepts to their respective Units:** defining Content Dictionaries that would describe the connection of units to general natural sciences concepts like *force* (measured in Newtons: $N = \frac{kgm}{s^2}$ or any variant of the ratio) or *energy* (measured in Joules: $J = Nm = \frac{kgm^2}{s^2} = ...$) and plenty of other examples. The interconnection of concepts in sciences: $Energy = Force \times displacement$ can further enable scientific formula "spell checking" which might prove to be of great value to physicists, astronomers and many others.
- **Unknown Unit Lookup:** In theoretical scientific papers authors usually use abbreviations for concepts (e.g. N for *Newtons* – the unit for *force*) without mentioning anything about units/dimensions, which might turn out

to be difficult for the readers who would be interested to know, for example, the order of measurement (magnitude) for the unknown physical quantities and also a (small) description of the respective concept (e.g. Pa is the unit for *pressure*). Defining a generic way in which semantics can be added to such unknown symbols will enable showing/hiding units for expressions/formulas.
- **Unit and Quantity Semantic Search:** a library-level service that would allow searching for units by their type, name and magnitude and return the relevant results independently of the measuring standard of the occurrences in the paper (e.g. imperial or metric) and also independent of their form (N or $\frac{kgm}{s^2}$).[6]
- **Quantity and Unit's Magnitude Manipulation:** a document interaction service that is able to transform for example $100N \to 0.1kN$ or $0.1 \times 10^3 N$ or $0.1 \times 10^3 \frac{kgm^2}{s^2}$. This can be useful when it comes to simplifying representations and adapting them consistently to a certain type of magnitude (for example *all occurrences of force expressions should have their unit represented in kN*).

As detailed at the beginning of this paper, having a standard, uniform understanding of units and quantities can prevent hazards and even eliminate entire compatibility check processes in industry. The presented list of prospective enabling technologies shows only a few of the numerous opportunities of interacting with units and quantities in semantically published documents and serves as strong motivation for future research in this direction.

7 Enabling Semantic Units in Legacy Corpora

The ARXMLIV corpus is the ideal environment for the identification of units and quantities since it contains a collection of more than 600,000 scientific publications. It is based on Cornell University's ARXIV e-Print archive [Arx] originally typeset in LaTeX, converted to XML in order to achieve easy machine-readability, partial semantics recovery and clear separation of document modalities such as natural language and mathematical expressions [Sta+10]. Currently, the project has achieved a successful conversion rate of nearly 70% to a semantically enriched XHTML+MATHML representation, natively understandable by modern web browsers [Koh+08].

A proof-of-concept check, performed via the ARXMLIV build system (see [Sta+10]) revealed roughly 150 ARXIV articles using the *SIunits* package, with an outlook for close to tripling the number when considering sibling packages such as *units* and *SIunitx*. This gives our work on creating a semantic binding for *SIunits* an even stronger benefit, as we can directly and non-invasively enrich legacy publications, putting them one step further on the path to semantic publishing. An additional, mid-term benefit is the opportunity to build a linguistic *Gold Standard* for units; we created both legacy (to presentational MATHML) and semantic (to OPENMATH) bindings in order to provide a raw, presentational

[6] In contrast, state-of-the-art scientific publication search services, such as Springer's LaTeX search [Spr], do not support the semantics of units.

output and its annotated, semantic counterpart. Having both as a basis, unit spotters can then be developed using methods of Computational Linguistics and Machine Learning, further enriching the ARXMLIV corpus.

Such enhancements not only enable the interactive services of semantic publishing on legacy corpora, but also provide a tempting outlook to the development of an ecosystem of linguistic analysis modules, which can draw on the captured semantics of units and quantities, as originally envisioned by the LAMAPUN project [Gin+09].

8 Conclusions and Future Work

Units and quantities are sufficiently wide-spread and important to not be disregarded from the context of semantic documents. Unfortunately, by now, there have been only isolated approaches (see section 4) to exploit the semantic power of units. Also considering the wide range of existing unit types and representations, makes it almost impossible to identify and semantically enrich all of them, especially when we are talking about occurrence contexts as unrelated as cooking recipes, medical prescriptions, technical documents or scientific papers.

Through the separation of the semantic publishing process for units we emphasized the importance of three major components: *representation*, *authoring* and *interaction*, detailing technologies that can improve each of them. Moreover, by providing a cooking recipe interaction use case and also a series of further potential services and applications on top of semantically published units, we contribute means of better manipulation and interpretation of *units and quantities* to the Semantic Publishing Industry and to legacy corpora.

Acknowledgments: The authors would like to thank Michael Kohlhase for his extensive support and advice regarding the writing of this paper, the anonymous peer reviewers for their extensive helpful suggestions, and Anton Antonov for writing the LaTeXML bindings for the *SIunits* LaTeX package.

References

[21c] *Code of Federal Regulations – Food and Drugs*. 2004. URL: http://edocket.access.gpo.gov/cfr_2004/aprqtr/21cfr101.9.htm.

[Air] *Aviation Safety – Air Canada Accident Report*. URL: http://aviation-safety.net/database/record.php?id=19830723-0 (visited on 10/25/2010).

[Arx] *arxiv.org e-Print archive*. URL: http://www.arxiv.org.

[Aus+10] R. Ausbrooks et al. *Mathematical Markup Language (MathML) Version 3.0*. Recommendation. W3C, 2010. URL: http://www.w3.org/TR/MathML3.

[Bar10] M. J. Barany. "'[B]ut this is blog maths and we're free to make up conventions as we go along': Polymath1 and the modalities of 'massively collaborative mathematics'". In: *WikiSym*. Ed. by P. Ayers and F. Ortega. ACM Press. 2010.

[BP09] D. Berrueta and L. Polo. *Measurement Units Ontology*. Nov. 9, 2009. URL: http://forge.morfeo-project.org/wiki_en/index.php?title=Measurement_Units_Ontology&oldid=12301.

[Bus+04] S. Buswell et al. *Open Math Standard 2.0*. Tech. rep. OpenMath Society, 2004. URL: http://www.openmath.org/standard/om20.

[Col09] J. B. Collins. "OpenMath Content Dictionaries for SI Quantities and Units." In: *Calculemus/MKM*. Ed. by J. Carette et al. Vol. 5625. LNCS. Springer, 2009, pp. 247–262.

[Crc] *Cooking.com – "Giant Chocolate Chip Cookies"*. URL: http://www.cooking.com/recipes-and-more/recipes/Giant-Chocolate-Chip-Cookies-recipe-5112.aspx (visited on 03/05/2011).

[Dav+10] C. David et al. "Publishing Math Lecture Notes as Linked Data". In: *The Semantic Web: Research and Applications (Part II)*. Ed. by L. Aroyo et al. LNCS 6089. Springer, 2010, pp. 370–375. arXiv:1004.3390v1 [cs.DL].

[DN03] J. H. Davenport and W. A. Naylor. *Units and Dimensions in OpenMath*. 2003. URL: http://openmath.org/documents/Units.pdf.

[Fin+10] J. L. Fink et al. "Word add-in for ontology recognition: semantic enrichment of scientific literature". In: *BMC Bioinformatics* 11.103 (2010).

[Gin+09] D. Ginev et al. "An Architecture for Linguistic and Semantic Analysis on the arXMLiv Corpus". In: *Applications of Semantic Technologies (AST) Workshop at Informatik*. 2009. URL: http://www.kwarc.info/projects/lamapun/pubs/AST09_LaMaPUn+appendix.pdf.

[GLR09] J. Giceva, C. Lange, and F. Rabe. "Integrating Web Services into Active Mathematical Documents". In: *MKM/Calculemus Proceedings*. Ed. by J. Carette et al. LNAI 5625. Springer, 2009, pp. 279–293.

[Gro+07] T. Groza et al. "SALT – Semantically Annotated LaTeX for Scientific Publications". In: *The Semantic Web: Research and Applications*. Ed. by E. Franconi, M. Kifer, and W. May. LNCS 4519. Springer, 2007, pp. 518–532.

[Hel] M. Heldoorn. *The SIunits package: Consistent application of SI units*. URL: http://mirror.ctan.org/macros/latex/contrib/SIunits/SIunits.pdf (visited on 03/13/2011).

[Koh08] M. Kohlhase. "Using LaTeX as a Semantic Markup Format". In: *Mathematics in Computer Science* 2.2 (2008), pp. 279–304. URL: https://svn.kwarc.info/repos/stex/doc/mcs08/stex.pdf.

[Koh+08] M. Kohlhase et al. "MathWebSearch 0.4, A Semantic Search Engine for Mathematics". manuscript. 2008. URL: http://mathweb.org/projects/mws/pubs/mkm08.pdf.

[Koh+11] M. Kohlhase et al. "The Planetary System: Web 3.0 & Active Documents for STEM". In: accepted for publication at ICCS 2011 (Finalist at the Executable Papers Challenge). 2011. URL: https://svn.mathweb.org/repos/planetary/doc/epc11/paper.pdf.

[KR09] M. Kohlhase and F. Rabe. "Semantics of OpenMath and MathML3". In: *22nd OpenMath Workshop*. Ed. by J. H. Davenport. 2009. URL: http://kwarc.info/kohlhase/papers/om09-semantics.pdf.

[Lan11] C. Lange. "Ontologies and Languages for Representing Mathematical Knowledge on the Semantic Web". accepted by Semantic Web Journal. 2011. URL: http://www.semantic-web-journal.net/content/new-submission-ontologies-and-languages-representing-mathematical-knowledge-semantic-web.

[Mar] *CNN – "NASAs metric confusion caused Mars orbiter loss"*. 1999. URL: http://articles.cnn.com/1999-09-30/tech/9909_30_mars.metric_1_mars-orbiter-climate-orbiter-spacecraft-team?_s=PM:TECH (visited on 10/29/2010).

[Mil] B. Miller. *LaTeXML: A LaTeX to XML Converter*. URL: http://dlmf.nist.gov/LaTeXML/ (visited on 03/03/2011).

[Oxf] *Oxford English Dictionary. "unit" definition*. URL: http://dictionary.oed.com/entrance.dtl (visited on 10/29/2010).

[PAU] *PAUX Technologies*. URL: http://paux.de (visited on 10/10/2010).

[Pla] *PlanetMath*. URL: http://planetmath.org (visited on 01/06/2011).

[RP05] R. G. Raskin and M. J. Pan. "Knowledge representation in the semantic web for Earth environmental terminology (SWEET)". In: *Computers & Geosciences* 31 (2005), pp. 1119–1125.

[SD08] J. Stratford and J. H. Davenport. "Unit Knowledge Management". In: *Intelligent Computer Mathematics*. Ed. by S. Autexier et al. LNAI 5144. Springer, 2008, pp. 382–397.

[Sib] *The International System of Units (SI) 8. Edition*. Bureau International des Poids et Mesures, 2006. URL: http://www.bipm.org/utils/common/pdf/si_brochure_8_en.pdf.

[Spr] Springer, ed. *LaTeX Search*. URL: http://www.latexsearch.com (visited on 04/16/2011).

[Sta+10] H. Stamerjohanns et al. "Transforming large collections of scientific publications to XML". In: *Mathematics in Computer Science* 3.3 (2010): *Special Issue on Authoring, Digitalization and Management of Mathematical Knowledge*. Ed. by S. Autexier, P. Sojka, and M. Suzuki, pp. 299–307. URL: http://kwarc.info/kohlhase/papers/mcs10.pdf.

[Str08] J. Stratford. *Creating an extensible Unit Converter using OpenMath as the Representation of the Semantics of the Units*. Tech. rep. 2008-02. University of Bath, 2008. URL: http://www.cs.bath.ac.uk/pubdb/download.php?resID=290.

[Swe] *Semantic Web for Earth and Environmental Terminology (SWEET)*. NASA. URL: http://sweet.jpl.nasa.gov/ (visited on 08/22/2010).

[Usm] *US Metric Association "Unit Mixups" article*. URL: http://lamar.colostate.edu/~hillger/unit-mixups.html (visited on 10/25/2010).