

Krextor – An Extensible Framework for Contributing Content Math to the Web of Data

Christoph Lange

Computer Science, Jacobs University Bremen,
ch.lange@jacobs-university.de

1 Problem Statement: No Math on the Web of Data Yet

An increasing amount of scientific knowledge is being contributed to the emerging Web of [Linked] Data, where it is made available in a machine-comprehensible way and interlinked with other related datasets. This already powers distributed query answering engines and intelligent semantic mashups enriching web publications – however, it still largely lacks mathematical functionality. There are e-science datasets – with mathematical model descriptions opaque to machines. There are statistical datasets – without explicit descriptions of how values have been derived. There are databases of scientific publications – with information about who cited your paper, but not who is building on your mathematical *ideas*.

2 The Krextor XML→RDF Extraction Library

In contrast to the document-oriented, often XML-based content markup of MKM, the graph-based RDF data model is most widely used for representing knowledge on the Web of Data. Therefore, in order to contribute mathematical knowledge to the Web of Data, we have developed the Krextor [5] library, and, on top of that, extraction modules that translate the structural outlines of OpenMath and OMDoc content markup to RDF. Krextor is an XSLT library that aims at facilitating the repetitive task of implementing translations from several XML input languages. It does so by offering convenience templates and functions for frequently occurring patterns in XML→RDF translation, such as creating RDF resources for things represented by XML elements, generating (“minting”) Linked Data compliant URIs for these resources, and translating XML text nodes or attribute to properties of these resources. Krextor allows for flexible integration into applications by supporting multiple output serializations of the RDF extracted, including callbacks to Java application code – whereas traditional hard-coded XSLT implementations would rather translate from exactly one XML input language to exactly one RDF output serialization (e.g. RDF/XML). Besides OpenMath and OMDoc¹, we have developed extraction modules for special MKM applications, such as encoding semantic web ontologies in OMDoc, and external developers have adopted Krextor outside of MKM.

¹ cf. [1, chapter 3] for a detailed description of the target RDF vocabularies/ontologies that we have developed for capturing mathematical knowledge, or for mappings to existing RDF vocabularies that we reused, e.g. for metadata

3 Publishing the OpenMath CDs as Linked Data

In contrast to previous work [2], the current focus of Krextor development is on expanding the coverage of the OpenMath 2 CD language (and proposed extensions beyond that), in order to prepare the publication of the official CDs at openmath.org as Linked Open Data [3]. Krextor has been used with OpenMath CDs before, but specifically for maintaining the, then, experimental collection of “OpenMath/MathML 3 CDs” *inside* a closed semantic wiki [4]. The current focus is on making most out of the official OpenMath CDs *as they are*, which means: (i) Supporting the maintenance of links from concepts in the OpenMath CDs to semantically equivalent concepts in related datasets – such as the Digital Library of Mathematical Functions (DLMF) or the PlanetMath encyclopedia (cf. [3]) – next to the CDs but without touching the CDs themselves. (ii) A prerequisite for that: Giving stable identifiers to mathematical properties of symbols – even though the OpenMath 2 CD language does not provide such identifiers. (iii) Utilizing existing XSLT code for translating OpenMath objects into Content MathML and other machine-comprehensible representations, so that interested applications can retrieve them right from the same dataset.

4 Coverage of the System Demo

The demo will focus on (i) the OpenMath CD extraction module, but also on (ii) Krextor’s possibilities for implementing extraction modules for new MKM languages. Regarding (i), I will particularly explain how to create new links between the OpenMath CDs and external datasets, and how RDF- and/or OpenMath-aware client applications can utilize the OpenMath CD linked dataset. Regarding (ii), I am prepared for a “hacking session” with any visitors who are interested in extracting RDF from their XML-based MKM language, in order to contribute their mathematical knowledge collections to the Web of Data.

References

- [1] C. Lange. “Enabling Collaboration on Semiformal Mathematical Knowledge by Semantic Web Integration”. PhD thesis. Jacobs University Bremen, 2011. URL: <https://svn.kwarc.info/repos/swim/doc/phd/phd.pdf>.
- [2] C. Lange. “Krextor – An Extensible XML→RDF Extraction Framework”. In: *Scripting and Development for the Semantic Web*. 2009. URL: <http://ceur-ws.org/Vol-449/ShortPaper2.pdf>.
- [3] C. Lange. “Towards OpenMath Content Dictionaries as Linked Data”. In: *OpenMath Workshop*. 2010. arXiv:1006.4057v1 [cs.DL].
- [4] C. Lange and A. González Palomo. “Easily Editing and Browsing Complex OpenMath Markup with SWiM”. In: *MathUI Workshop*. 2008. URL: <http://www.activemath.org/workshops/MathUI/08/>.
- [5] C. Lange et al. *Krextor – The KWARC RDF Extractor*. URL: <http://kwarc.info/projects/krextor/> (visited on 03/15/2011).