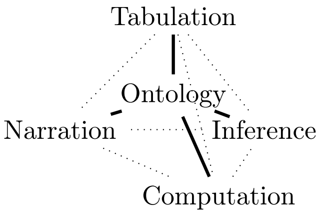
Evolving MathHub for Tetrapodal Math (Working Title)

Tom Wiesing

The STEM sciences – Science, Technology, Engineering and Mathematics – are moving to bigger and bigger developments, ultimately facing problems such as the *one-brain-barrier* [4]. To overcome these limitations, we must make use of computer support.

In the following I will use mathematics as a working example. Of the STEM sciences it is one with the most well-understood structure. I expect that the observations and results will generalize to all STEM sciences.

We observe the following aspects of mathematical practices: (i) taking existing mathematical knowledge and using it to make *inferences*; (ii) guiding a reader by writing *narrating* texts about a specific topic; (iii) generating a large set of information – such as return values of a particularly interesting function – and *tabulating* it inside some form of database; and (iv) making *computations*. We see these four as forming central aspects of mathematics. [1] arranges these in a terapodal strucuture and posits that all four must be integrated by a mathematical ontology for effective machine support.



To support each of these practices, different kinds of computer-provided mathematical services are required. Services commonly work on objects ranging from low-level mathematical objects, such as *numbers*, via complex objects, such as *specific mathematical groups*, to objects on a higher meta-level, such as *mathematical computations*. Any particular service might treat specific objects as blackboxes – knowing that it is an object but not knowing any of its' internal structure – or as a white box – being fully aware of the semantics of the objects and being able to infer all its' properties. We call services treating objects as the former *shallow mathematical services* and those as the latter *deep mathematical services*.

Deep services also have to take context into account. For example, in certain cases an integer might only be an integer, but in other cases the same integer might represent the value of a specific property of an object. This leads to the orthogonal distinction between *local services* – those taking only the current document into account – and *global services* – those taking an entire ontology of background knowledge or corpus of documents into account.

Consider on one hand a service that takes as input a simple mathematical term within a document, and then navigates to the document defining this term. This service is global (because it takes the corpus as a whole into account) and shallow (because it does not rely on the internal structure of the object). Consider on the other hand a service that converts all physical quantity expressions (e.g. $10\frac{\text{km}}{\text{h}}$) from the imperial system to the metric system. This service is local (because it only works on a particular expression) and deep (because for each quantity expression it takes into account its' internal structure consisting of value and units).

Services that are shallow and local are easy to implement – requiring only application of standard software engineering practices to quickly be usable on a researchers machine – whereas deep and global services require more infrastrucure. To be able to support these more complex services, a highly semantic, highly structured ontology is needed to bind the different mathematical aspects together.

The classification of aspects of 'doing mathematics' and – based on that – the services sketched above give

In: C. Kaliszyk, E. Brady, J. Davenport, W.M. Farmer, A. Kohlhase, M. Kohlhase, D. Müller, K. Pąk, and C. Sacerdoti Coen (eds.): Joint Proceedings of the FMM and LML Workshops, Doctoral Program and Work in Progress at the Conference on Intelligent Computer Mathematics 2019 co-located with the 12th Conference on Intelligent Computer Mathematics (CICM 2019), Prague, Czech Republic, July 8–12, 2019, published at http://ceur-ws.org

a plausible, albeit theoretical, model for designing a system providing computer support for mathematics. To validate this model, I want to build a concrete scalable implementation. This leads to the following research questions:

RQ1 Can an information system provide infrastructure to jointly support all four tetrapodal aspects of mathematics?

RQ2 Is it possible to support community-based workflows for authoring, maintaining, curating, and visualization of tetrapodal mathematics?

RQ3 What kind of ontology is needed for our model to work in practice?

A positive answer to question **RQ1** – backed by an implementation – will demonstrate the feasibility of our model. The aim of Question **RQ2** is to guide any implementation towards practical applicability. In particular, it is insufficient to build a system that is merely applicable to a specific subset of documents; instead we want to aim to make it broadly accessible to the community of mathematics as a whole. We expect these particular workflows to cause a need in an implementation to support both deep and global services.

Unlike the above, the answer to Question **RQ3** is not yet clear. At first glance, one would expect any such ontology to be fully formal, defining every part of mathematics in great detail. However, it might be possible to re-write mathematical services to not need a complete formalization of every object in question. As an alternative a flexi-formalized ontology – defining only core concepts and providing human-readable descriptions for the rest – might be sufficient[1].

Another possible requirement on the ontology might be to ensure that it is not system specific. In particular, it is possible for it to remain findable and accessible by other researchers, to further system interoperability? These desired properties are also known as FAIR [9].

I want to base my system on MathHub, a portal for active mathematical documents, an archive for flexiformal mathematics, and a frontend to mathematical databases initially conceived by Michael Kohlhase, Florian Rabe and others in [7]. MathHub roughly consists of three parts: *a*) Several *archives* of mathematical knowledge, each consisting of a set of thematically related OMDOC [5] documents; *b*) an MMT backend [8] capable of providing semantic services on these archives; and *c*) a Drupal [2] frontend presenting an interface to users.

MathHub was initially implemented by Mihnea Iancu and others during the scope of his PhD [3]. It provided services for several of the above-defined aspects of mathematics. However, there are two major problems with this implementation *a*) it does not implement a tetrapodal system as a whole; and *b*) several technological problems necessitate a re-design of parts of the system.

I want to base my work to answer the research questions on MathHub by developing the system further. On a technical level, I want to overcome the two limitations above, and then focus on the following workflow-oriented work packages that concretly instantiate the research questions:

WP1 Which tetrapodal *Active Document Services* can be developed and efficiently integrated into MathHub that make the system more user-friendly?

WP2 Which processes are efficient at keeping MathHub Content up-to-date and simultaneously scale to large amounts of data?

WP3 What kind of integrable *search functionality* is desirable for MathHub users?

WP4 How can MathHub improve *interoperability* with external systems and services?

WP1 addresses **RQ1**, **WP2** and **WP3** focus on **RQ2** and **WP4** focuses on the ontology in **RQ3**.

References

[1] Jacques Carette et al. "Big Math and the One-Brain Barrier – A Position Paper and Architecture Proposal". submitted to Mathematical Intelligencer. 2019. URL: `https://arxiv.org/abs/1904.10405`.

[2] *Drupal.org – Community plumbing.* URL: `http://drupal.org` (visited on 02/14/2015).

[1] There are already some ideas how such an ontology might look like in detail. We refer the interested reader to [6].

[3] Mihnea Iancu. “Towards Flexiformal Mathematics”. PhD thesis. Bremen, Germany: Jacobs University, 2017. URL: https://opus.jacobs-university.de/frontdoor/index/index/docId/721.

[4] Michael Kohlhase. “Mathematical Knowledge Management: Transcending the One-Brain-Barrier with Theory Graphs”. In: *EMS Newsletter* (June 2014), pp. 22–27. URL: https://kwarc.info/people/mkohlhase/papers/ems13.pdf.

[5] Michael Kohlhase. *OMDoc – An open markup format for mathematical documents [Version 1.2]*. LNAI 4180. Springer Verlag, Aug. 2006. URL: http://omdoc.org/pubs/omdoc1.2.pdf.

[6] Michael Kohlhase. “The Flexiformalist Manifesto”. In: *International Workshop on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2012)*. Ed. by Andrei Voronkov et al. Timisoara, Romania: IEEE Press, 2013, pp. 30–36. ISBN: 978-1-4673-5026-6. URL: http://kwarc.info/kohlhase/papers/synasc13.pdf.

[7] Christoph Lange et al. “The Planetary System: Executable Science, Technology, Engineering and Math Papers”. In: *The Semantic Web: Research and Applications (Part II)*. 8th Extended Semantic Web Conference (ESWC) (Hersonissos, Crete, Greece, May 29–June 2, 2011). Ed. by Grigoris Antoniou et al. LNCS 6644. Heidelberg: Springer Verlag, 2011, pp. 471–475. ISBN: 978-3-642-21033-4. arXiv: 1103.1482 [cs.DL].

[8] Florian Rabe and Michael Kohlhase. “A Scalable Module System”. In: *Information & Computation* 0.230 (2013), pp. 1–54. URL: http://kwarc.info/frabe/Research/mmt.pdf.

[9] Mark D. Wilkinson et al. “The FAIR Guiding Principles for scientific data management and stewardship”. In: *Scientific Data* 3 (2016). URL: https://doi.org/10.1038/sdata.2016.18.