Aligning Mathematical Concepts Across Libraries

Florian Rabe

for the working group of the October 2023 Dagstuhl seminar on Automated Mathematics

Motivation Irrespective of the knowledge aspect (such as deduction, computation, data, or documentation), mathematical libraries share several abstract characteristics. Most importantly, they can be seen as a set of fragments each carrying a global identifier and describing a mathematical concept. Here we use *concept* as a generic term to subsume any named type, object, operation, theorem, or similar. For example, in a logical library, these are (mostly) definitions and theorems; in a computer algebra library, they are (mostly) type and function definitions; in a concrete data set, they are the entries of the data set, often rows in a table; and in a document library, they are the articles explaining a concept.

The work group recognized the *alignment problem* [KKMR16, MGK+17] as a major roadblock to interoperability across mathematical software systems: The mathematical concepts introduced by the various libraries overlap substantially, but there is no systematic connection between the same mathematical concept described in different libraries. Concretely, we call a pair of identifiers, typically but not necessarily from different libraries, an *alignment* if both are descriptions of the same concept. Then the alignment problem can be formulated as the challenge of (a) collecting alignments for existing large mathematical libraries and (b) leveraging these alignments for knowledge interchange.

The group identified three levels at which the problem can be attacked: (i) At the *identifier level*, the alignments are just pairs of identifiers without a machine-checkable guarantee that they correspond to each other in any way. (ii) At the *expression level*, an alignment (c, d) additionally carries information how terms with head c can be translated to terms with head d. This translation can be quite complex and involve, e.g., changing (which may require computation), adding (which may require inference), omitting, or reordering arguments. (iii) At the *semantic level*, the alignments are additionally verified for correctness. This can be done, e.g., by translating theorems about c along the alignment and proving the translated theorems in the system of d. Critically, from (i) to (iii), task (a) becomes harder while task (b) becomes easier. But even identifier level alignments are useful, e.g., to cross-reference across libraries or to search for the same query in multiple libraries in parallel.

Results The group surveyed the available technologies and alignment collection efforts and concluded that, while semantic alignments must be the ultimate goal, only for identifier level alignments is a major community-driven collection effort feasible at this point.

The group compiled the following existing collections of identifier alignments and concept lists:

- the Math Subject Classification (MSC)
- the nLab page titles (`https://ncatlab.org/nlab/`)
- the SMGloM concept and translation library [GIJ+16]
- the concept list for the undergraduate math curriculum and the alignments into Lean Mathlib (`https://github.com/leanprover-community/mathlib4/blob/master/docs/undergrad.yaml`)
- the concept translation library maintained by Hosgood (`https://thosgood.com/maths-dictionary/`)
- the manual alignments collected for theorem prover libraries in [MRLR17]
- the concept list used by SageMath to align computer algebra systems integrated with SageMath (`https://doc.sagemath.org/html/en/reference/categories/index.html`)
- the MathGloss alignments for undergraduate math education [HdP23] (`https://mathgloss.github.io/MathGloss/`)
- the relevant subset of the Wikidata ontology, which includes various alignments to informal libraries
- the nNexus alignments across informal libraries [GC14]
- the semantic alignments between HOL systems found by machine learning in [GK19]

The work identified a set of several hundred concepts that can be used as a seed for an alignment library and started using the above resources to compile alignments for it. These resources are collected at `https://github.com/UniFormal/alignments/`.

In order to scalably collect, maintain, and leverage alignment sets in the future, the group makes two recommendations:

- All developers of math libraries should add a feature to their tool that allows tagging definitions with the aligned identifier in a central concept list. The build process of the library should generate the list of alignments between those central concepts and the tagged identifiers in the system's library. This list should be published alongside the library.

- Wikidata is suggested as the central concept list. This is motivated by the observation that Wikidata is a neutral library (in the sense of not being biased towards any research system or community) and the most likely to be scalably maintained and broadly used in the long term.

References

[GC14] D. Ginev and J. Corneli. Nnexus reloaded. In S. Watt, J. Davenport, A. Sexton, P. Sojka, and J. Urban, editors, *Intelligent Computer Mathematics*, pages 423–426. Springer, 2014.

[GIJ+16] D. Ginev, M. Iancu, C. Jucovshi, A. Kohlhase, M. Kohlhase, A. Oripov, J. Schefter, W. Sperber, O. Teschke, and T. Wiesing. The smglom project and system: Towards a terminology and ontology for mathematics. In G. Greuel, T. Koch, P. Paule, and A. Sommese, editors, *Mathematical Software - ICMS*, volume 9725, pages 451–457. Springer, 2016.

[GK19] T. Gauthier and C. Kaliszyk. Aligning concepts across proof assistant libraries. *Journal of Symbolic Computation*, 90:89–123, 2019.

[HdP23] L. Horowitz and V. de Paiva. Mathgloss: Linked undergraduate math concepts, 2023. `https://arxiv.org/abs/2311.12649`.

[KKMR16] C. Kaliszyk, M. Kohlhase, D. Müller, and F. Rabe. A Standard for Aligning Mathematical Concepts. In A. Kohlhase, M. Kohlhase, P. Libbrecht, B. Miller, F. Tompa, A. Naummowicz, W. Neuper, P. Quaresma, and M. Suda, editors, *Work in Progress at CICM 2016*, pages 229–244. CEUR-WS.org, 2016.

[MGK+17] D. Müller, T. Gauthier, C. Kaliszyk, M. Kohlhase, and F. Rabe. Classification of Alignments between Concepts of Formal Mathematical Systems. In H. Geuvers, M. England, O. Hasan, F. Rabe, and O. Teschke, editors, *Intelligent Computer Mathematics*, pages 83–98. Springer, 2017.

[MRLR17] D. Müller, C. Rothgang, Y. Liu, and F. Rabe. Alignment-based Translations Across Formal Systems Using Interface Theories. In C. Dubois and B. Woltzenlogel Paleo, editors, *Proof eXchange for Theorem Proving*, pages 77–93. Open Publishing Association, 2017.