The GDML and EuKIM Projects: Short Report on the Initiative

Bruno Buchberger(✉)

Research Institute for Symbolic Computation (RISC),
Johannes Kepler University, Linz, Austria
bruno.buchberger@risc.jku.at
http://www.risc.jku.at/people/buchberg/

Abstract. We give a report on the EuKIM project, which was recently submitted to the EU Horizon 2020 program, INFRAIA-02-2017 (Integrating Activities for Starting Communities) topic, by a consortium of twelve European research groups. The project aims at building up a "Global Digital Math Library" (knowledge base) integrating and extending current efforts worldwide. A central part of the project is the design and implementation of a software system that organizes open and one-stop access to mathematical knowledge and to various tools for processing mathematical knowledge. Recent progress in automated reasoning is an important issue for achieving more sophisticated levels in this endeavor.

In 2014 the International Mathematical Union (IMU) called for creating a "Global Digital Mathematics Library" (GDML) (GDML 2016) – a comprehensive, extensible, machine-processable knowledge base of mathematics providing one-stop open access for researchers, teachers, students and all users of mathematics in science, technology, and industry.

Existing digital libraries provide web access and search of entire papers, but the fine-grain units of knowledge in mathematics are statements expressing concepts, theorems, conjectures, problems, algorithms, proofs and more. A mathematical knowledge base must, therefore, allow users to store, retrieve, verify, and even invent knowledge at the level of individual statements and theories (collections of statements) from which new statements can be derived.

In the past twenty years, major advances in automating mathematical reasoning by using algorithmic intelligence in interaction with artificial intelligence have been made with important contributions by European research groups. This was a strong motivation for the recent decision of twelve European academic institutes to form a consortium for establishing the proposed digital knowledge base. The name of this initiative is "European Knowledge Infrastructure for Mathematics" (EuKIM). Roughly, the EuKIM infrastructure should do for precise, human AND machine processable mathematical knowledge what Wikipedia has done for imprecise human-only processable general knowledge.

B. Buchberger—Speaker of the EuKIM Consortium and representative of the RISC Institute.

G.-M. Greuel et al. (Eds.): ICMS 2016, LNCS 9725, pp. 443–446, 2016.
DOI: 10.1007/978-3-319-42432-3_56

The first joint action of the EuKIM Consortium was the formulation and submission (March 2016) of a long-term EU-Proposal "EuKIM" in the frame of the EU-Program Horizon-2020-INFRAIA-02-2017. The EuKIM project has three aspects:

- joined research on the topics relevant for creating GDML
- building up and extending a (European and global) network of research institutes that want and are able to contribute to the goal of building up a GDML
- designing and developing a software system that can serve as the one-stop open-access interface to all existing mathematical knowledge bases and to all existing processing tools for mathematical knowledge and, at the same time, will provide new mathematical knowledge bases (with an emphasis on formal knowledge bases) and new processing tools for mathematics (with an emphasis on automated reasoning tools).

In this talk we focus on the software aspect of the EuKIM project. The following, very rough, diagram (Fig. 1) gives an overview on the software to be designed and implemented.

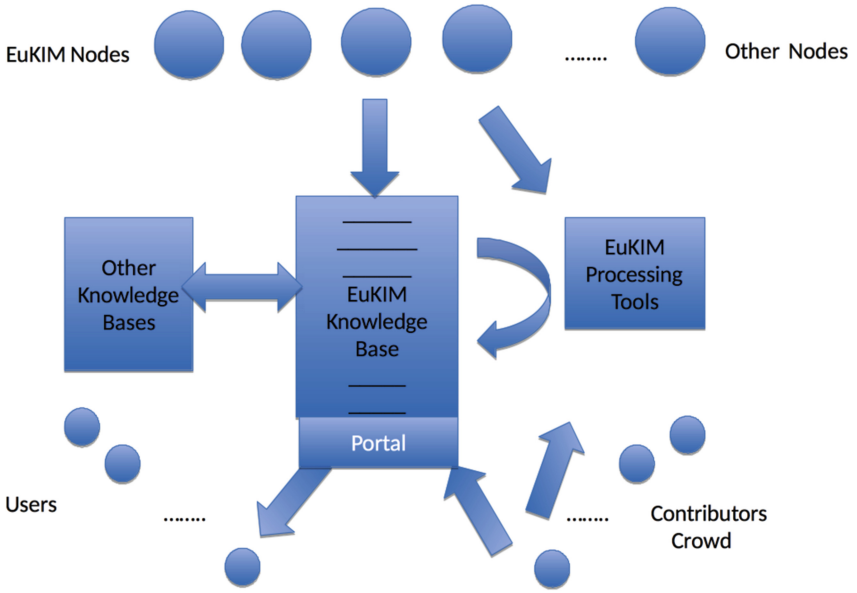

Fig. 1. Concept of the EuKIM infrastructure

We start from the side of the typical users (math researchers, teachers, students, people who apply math in science, engineering, etc.). The users should be able to input queries on mathematics in various levels of sophistication: On the first level, simple (natural language) keywords and combinations of keywords

can represent the input, like for ordinary search engines. On the second level, natural language sentences describing problems, questions, or conjectures in context should be presentable in a style similar to what is possible in very recent natural language question answering systems for everyday situations (e.g.: the new version of Siri from Apple presented at Disrupt NY, May 9–11, 2016). The third level (which is specific to mathematics) will include mathematical formulae in unambiguous formalization in some logic language (some version of predicate logic) as part of the input.

The EuKIM software should analyze the request and decide which existing web-accessible mathematical libraries and knowledge bases contain relevant information on the questions or problem and should give this information back in a good structure with a navigator through the items found. (This information may include links to papers whose download may request paying a fee.)

However, in addition, the system should direct the user also to all (open-access and commercial) tools for processing the knowledge found in relation to the request submitted. This will include all the recent automated reasoning tools that may help in finding logically equivalent, similar, or relevant mathematical knowledge, in deciding or semi-deciding the validity of formulae, in finding the logical dependence of formulae from each other, in using algorithms for processing mathematical data etc. Important examples of such tools are the existing mathematical algorithm libraries like Mathematica, Maple, etc., and the existing automated reasoning systems like HOL (Nipkow et al. 2015), Coq (Affeldt and Kobayashi 2008), Isabelle (Nipkow et al. 2015), Theorema (Buchberger et al. 2016), etc. The system will also deliver complete mathematical theories in formal representation as a starting point for original research, as far as such theories are available on the web.

In addition to existing tools for processing mathematical knowledge and existing mathematical theories, the EuKIM software will also add more and more mathematical theories and processing tools and will motivate research groups (and also users) to contribute to theories and tools under the structural guidance and tools-support of the system.

Over the years, the EuKIM software system should grow into an intelligent dialogue machine for mathematics: Not by re-invention of all the existing beautiful and powerful web-accessible mathematical paper collections, formulae collections, algorithm collections, but by integration of all these tools into an open-access (as much as possible) and one-stop portal and by the addition of new tools of high mathematical, i.e. automated reasoning, intelligence.

References

Affeldt, R., Kobayashi, N.: A Coq library for verification of concurrent programs. Electron. Notes Theor. Comput. Sci. **199**, 17–32 (2008). Proceedings of the Fourth International Workshop on Logical Frameworks and Meta-Languages (LFM 2004)

Buchberger, B., Jebelean, T., Kutsia, T., Maletzky, A., Windsteiger, W.: Theorema 2.0: computer-assisted natural-style mathematics. J. Formalized Reasoning **9**(1), 149–185 (2016). Special Issue: Twenty Years of the QED Manifesto

The Global Digital Mathematics Library Working Group. International Mathematical Knowledge Trust Charter (2016). http://mathontheweb.org/gdml/IMKT-Charter-final.pdf

Nipkow, T., Paulson, L.C., Wenzel, M.: Isabelle/HOL - A Proof Assistant for Higher-Order Logic. Springer, Heidelberg (2015)