Journal as Active Math-Agents: Outline of a Project with a Mathematics Publisher

Bruno Buchberger

As in any other scientific field, the quality control (by peer reviewing), archiving and distribution of knowledge in mathematics is organized by journals. Currently, mathematical journals represent huge chunks of knowledge that sit passively on a shelf and wait for a human user to read or query them.

In contrast, what we want to achieve in this project is the transformation of the knowledge contained in mathematical journals into formal mathematical knowledge bases that behave as active reasoning agents.

After 50 years of intensive world-wide research in automated reasoning and symbolic computation, we have the technology that should enable us to construct reasoning tools and re-organize the knowledge contained in any mathematical journal in such a way that questions of the following type can be answered automatically or, at least, with significant computer-support:

- **Originality**: Is a given mathematical result already *contained* in the journal, i.e. can it be derived from the knowledge in the journal by *easy* (automated) reasoning?

- **Correctness**: Is a result in a paper correct, i.e. can it be formally derived from the available knowledge (using the proof *script* given in the paper)?

- **Importance**: Is the result important, i.e. how, where, and how often can it be used in the context of the available knowledge?

- **Completeness**: Are all *easy* consequences of a given result drawn in the available knowledge context?

- **Similarity**: Search for structurally similar results in all areas of the available knowledge! Generate a pattern.

- **Complete Knowledge**: Draw all consequences of the available knowledge using the currently available reasoners.

- **Knowledge reduction**: Given (increasingly sophisticated) automated reasoners, from time to time, the amount of mathematical knowledge to be stored can be reduced to the propositions whose proof is not *easy* (w.r.t. to the available automated reasoning power) whereas all other knowledge can be reproduced quickly, by the available automated reasoners, on demand.

- **Minimal Prerequisites**: Analyze existing mathematical propositions for minimal prerequisites necessary in their proof.

- **Trace History**: Generate the successive versions of a given result (notion, proposition, problem, method) in the knowledge base over time.

- **Re-organize knowledge**: Take (a part of) the given knowledge and re-structure it under a particular perspective (*view*).

- **Problem generation**: What questions can be asked and what problems can be formulated on the bases of the available knowledge.

It should be clear that, having the methodology to answer these questions will generate dramatic benefits for all parties involved in the generation and usage of mathematical knowledge through journals:

- **Pre-publication phase**: For authors, in the process of research on a certain topic and manuscript preparation, authors will be able to search for relevant knowledge, draw consequences of the available knowledge, build on a consistent body of definitions and propositions, and minimize the amount of explicit information for formulating their final results. Also, they will be able to pre-determine the quality of their results before submitting the manuscript to peer reviewing.

- **Publication phase**: Since some of the above questions (like originality, correctness, importance, completeness, quality of literature references) are also part of the quality control (peer reviewing) process, computer-support for the above questions will also have a drastic influence on the quality, speed, efficiency, reliability, and completeness of the refereeing process. Also, the feedback to authors from the reviewing process can be much more specific and constructive.

- **Post-publication phase**: For the reader, a formalized journal together with its reasoning tools will be kind of a living agent for discussing scientific questions in the specific area. It will mainly be an instrument for answering queries in a profound way. Ultimately, it will also be an agent that will generate questions, problems, and challenges.

The objective of this project is to turn the data base of a major mathematical journal into a formalized mathematical knowledge base and to turn the journal into an active reasoning agent by implementing and using current computer-supported reasoning technology. As a concrete case study, we will use the Journal of Symbolic Computation of Elsevier, the journal which was founded 1985 by the speaker. This journal is mathematical in content and, at the same time, contains part of the research that forms the basis of the computer-supported reasoning technology to be applied. Thus, this is a kind of self-application of the results published in the journal for automating itself.

If this experiment is successful, the technology can be applied to any other mathematical journal. In addition the knowledge bases from each journal will be interoperable and could contribute to a master mathematical knowledge base, linking logical statements from many journals together. This could drive mathematics research forward at a much greater rate than we currently experience.

It is clear that along the way of working on this project, as a side-effect, insight will be gained of how the future process of publishing in the area of mathematics will have to change in order to produce mathematical knowledge as an active reasoning agent at the moment it is generated. This will also result in a different style of writing, submitting, refereeing, archiving, and distributing mathematical papers.

In the talk we will discuss more details of the project and the Theorema reasoning technology on which it will be based as well as the chances of success of this project.