

Workshop on SymbolicData Design

Leipzig, 13.–14. December 2012

<http://symbolicdata.org/wiki/Events.2012-12-13>

Within the E-Science Benchmarking Project we invited for a workshop and hackathon to discuss and promote different aspects of the SymbolicData Project. The workshop took place at HTWK – Hochschule für Technik, Wirtschaft und Kultur Leipzig. We had two days of intense discussions about the goals, philosophy, subprojects, links etc. of the SymbolicData Project [1].

First we discussed the current state of the project. *Hans-Gert Gräbe* (Uni Leipzig) explained in detail the work done so far towards a redesign of the Data collection according to Linked Open Data standards. Within this refactoring process we distinguish more clearly between *Data* (called *XMLResources*) and *Metadata* (called *RDFResources*; interlinking of metadata is nowadays best supported by the RDF based Semantic Web Stack [2]). Such a distinction allows to express more clearly another point: Data and its semantic meaning are managed *within* different Computer Algebra Communities, Metadata are required for *Cross Community Communication* purposes. The main future focus of SymbolicData will be on the needs of such a Cross Community Communication between different Computer Algebra Communities.

Albert Heinle (RWTH Aachen, now U Waterloo) presented the *SDEval* framework. It grew up from the profiling and testing needs of the Free Algebra community [3], but is generic enough to serve as a best practice how to organize automated set up, run, evaluation and comparison of dedicated computational tasks on a large amount of data. The framework is written in *python*, heavily uses UNIX process management facilities to flexibly define and set up computational environments with dedicated characteristics, and can be reused for a wide range of computational tasks with different CA software. SDEval continues the SymbolicData efforts to establish standards how to set up environments for testing and benchmarking of CA software on a larger collection of given data.

Satya Samal (Uni Bonn) presented the PoCaB Project [4], explained in detail structural approaches within the PoCaB Databases and how data are generated within the PoCaB framework. PoCaB mainly addresses topics around categorization of differential equation systems in mass action and non-mass action kinetics in chemical systems coming from a biological background. PoCaB is interlinked with different communities within CA (Polynomial Systems Solving and the Polymake communities) and also beyond. In particular, it heavily exploits biological databases (BioModel Database, KEGG Database) [5] that come with their own language SMBL and experiences how to express semantical aspects in a computer readable way.

Johannes Waldmann (HTWK Leipzig) gave a talk about Benchmarks and Competitions in Theoretical Computer Science presenting best practices of three TCS Communities: Termination, SAT and SMT. He explained the Termination Problems Data Base [6] and their way of benchmarking: They regularly organize Termination Competitions on previously agreed data from different problem categories at a central site. This competition accompanies the annual large conference in the field. Waldmann emphasized that most communities have their own (intracommunity) infrastructure – workshops, mailing lists, wiki (to adjust a „common story“) – and domain specific

- input syntax and semantics specification,
- standards for what is an acceptable proof trace,
- methods for selecting competition problems, and
- algorithms for scoring results,

that should be reused as much as possible by efforts like SymbolicData. Waldmann is involved with the StarExec Project [7] that „has the goal to provide a domain-agnostic execution platform (software and hardware) for running competitions in computational logics and developed some meta-model of competitions that covers standards for benchmarks, tools and results“.

At the meeting we decided about the future main road of the SymbolicData Project. First, the SymbolicData Project will be refocussed to address needs and efforts of *communities* within Symbolic Computation to profile, test and benchmark implementations on larger sets of data.

There is a commonly complained misrecognition of such efforts because they are not in the focus of reputational processes of the respective communities and are in rare cases acknowledged properly. Such questions arise in other experimentally based sciences, too.

SymbolicData (v.1 and v.2) had its origin within the Polynomial Systems Community, so such a refocussing has to be processed also as a reorganization of data for SymbolicData v.3. This work is on the way. A list of communities with benchmarking activities addressed by SymbolicData will be maintained on the SymbolicData website.

For the future there should be a better interlinking between (intracommunity) sources, resources and communication structures within such a community and SymbolicData. This will be carefully studied on a number of use cases in cooperation with the SPP 1489.

SDEval as a python based generic benchmarking compute framework represents best practice to run dedicated computational tasks on a large amount of given data. This code is available from the SymbolicData Public Repository.

In the near future we focus on consolidating SymbolicData and releasing a stable v.3. As a first step we moved to git and operate a public repository at github [8]. There is a Sparql endpoint [9] for SymbolicData that serves the latest RDFData. In the second half of July there will be another workshop in Leipzig to resume current progress.

Links:

- [1] The SymbolicData Project. <http://symbolicdata.org>.
- [2] The Semantic Web Stack. http://en.wikipedia.org/wiki/Semantic_Web_Stack.
- [3] The SD Free Algebra Subproject. <http://symbolicdata.org/wiki/FreeAlgebra>.
- [4] The PoCaB Project – Platform of Chemical and Biological Analysis Using Computer Algebra Methods. <http://pocab.cg.cs.uni-bonn.de>.
- [5] Bio Model Databases. <http://www.ebi.ac.uk/biomodels-main>.
- [6] TPBD – the Termination Problems Data Base. <http://termination-portal.org/wiki/TPDB>.
- [7] The StarExec Project. <http://www.starexec.org/starexec/public/about.jsp>.
- [8] For details, see <http://symbolicdata.org/wiki/Using.Git>.
- [9] The Sparql endpoint operates at <http://symbolicdata.ontowiki.net>.

Hans-Gert Gräbe (Leipzig)¹

¹to be published in „Computeralgebra-Rundbrief“ 52, März 2013