The SymbolicData Project

Hans-Gert Gräbe and Andreas Nareike

Universität Leipzig, Germany
(graebe|nareike)@informatik.uni-leipzig.de

Abstract. We report about a complete redesign of the tools and data of the SymbolicData project according to Linked Data principles and RDF technologies that proved to be powerful within modern semantic web approaches. During that redesign the focus of the project changed from a merely data store towards the vision of a Computer Algebra Social Network that aims at technical support of the intercommunity communication between different CA subcommunities. The redesign was released as version 3 of SymbolicData in September 2013 and since then steadily evolved.

1 Introduction

The SymbolicData project grew up from the Special Session on Benchmarking at the 1998 ISSAC conference to continue the efforts started by the PoSSo [11] an FRISCO [5] projects. It aimed at building a reliable and sustainably available reference of Polynomial Systems data, to extend and update it, to collect meta information about the records, and also to develop tools to manage the data and to set up and run tests and benchmark computations on the data. A first prototype was developed during 1999–2002 by Olaf Bachmann and Hans-Gert Gräbe, that collected data from *Polynomial Systems Solving* and *Geometry Theorem Proving*.

Olaf Bachmann left the project for a new job at the end of 2000, and there was almost no advance during 2002–2005. In a second phase around 2006 the project matured again. Data were supplied by the CoCoA group (F. Cioffi), the Singular group (M. Dengel, M. Brickenstein, S. Steidel, M. Wenk), V. Levandovskyy (non commutative polynomial systems, G-Algebras) and R. Hemmecke (Test sets from Integer Programming). In 2005 the German Fachgruppe Computeralgebra launched the Web site `http://www.symbolicdata.org`. During the Special Semester on Groebner Bases in March 2006 we tried to join forces with the GB-Bibliography project (B. Buchberger, A. Zapletal) and the GB-Facilities project (V. Levandovskyy).

In 2009 we started to refactor the data along standard Semantic Web concepts based on the Resource Description Framework (RDF). In 2012 these efforts were supported by a 12 months grant *Benchmarking in Symbolic Computations and Web 3.0* for Andreas Nareike within the *Saxonian E-Science Initiative* [3]. Within that scope we completed a redesign of the data distinguishing more consequently between data (*resources* in the RDF terminology) and meta data (*knowledge bases* in the RDF terminology) and refactoring the meta data along Linked Data principles. The new SymbolicData data and tools were released as version 3 in September 2013.

2 The SymbolicData Infrastructure

Our resources (examples for testing, profiling and benchmarking software and algorithms from different areas of symbolic computation) are publicly available in XML markup, meta data in RDF notation both from a public git repo, hosted at `github.org`, and from an OntoWiki [10] based data store at `http://symbolicdata.org/Data`. Moreover, we offer a SPARQL endpoint [14] to explore the data by standard Linked Data methods.

The website operates on a standardized installation using an Apache web server to deliver the data, the Virtuoso RDF data store [16] as data backend, a SPARQL endpoint and (optionally) OntoWiki to explore, display and edit the data. This installation can easily be rolled out at a local site (tested under Linux Debian and Ubuntu 12.04 LTS; a more detailed description can be found in the SymbolicData wiki [13]) to support local testing, profiling and benchmarking.

The distribution offers also tools for integration with a local compute environment as, e.g., provided by Sagemath [12] – the Python based *SDEval package* [7] by Albert Heinle offers a JUnit like framework to set up, run, log, monitor and interrupt testing and benchmarking computations, and the *SDSage package* [9] by Andreas Nareike provides a showcase for SymbolicData integration with the Sagemath compute environment.

We follow a development process along the Integration-Manager-Workflow Model. This makes it easy to join forces with the SymbolicData team: Fork the repo to your github account, start development and send a pull request to the Integration Manager if you think you produced something worth to be integrated into the upstream master branch. Even if your contribution is not pulled to the upstream, people can use it, since they can pull it from your to their github repo. This allows even for agile common small feature development – a widely practised way to advance projects hosted at `github.com`. You are encouraged early to start a discussion about your plans and regularly report your progress on the SymbolicData mailing list.

3 SymbolicData Resources and Maintenance

Preparing SymbolicData version 3 we decided to strengthen the SymbolicData part that is *not* involved with Polynomial Systems Solving. These efforts led to a more consequent distinction between data (owned and maintained by different CA subcommunities) and meta data. Such a distinction is well supported by RDF design principles – the Resource Description Framework is about *description* of *resources*, represented by (globally unique) *resource identifiers* (URIs). We follow the Linked Data best practise to provide URIs in such a way, that they are accessible by the HTTP internet protocol and a valuable part of structured information about that resource is delivered upon HTTP request to that URI.

Currently the SymbolicData data collection contains resources from the areas of Polynomial Systems Solving (390 records, 633 configurations), Free Algebras (83 records), G-Algebras (8 records), GeoProofSchemes (297 records) and Test Sets from Integer Programming (28 records). These resources are stored in a flat XSchema based XML syntax using well established intracommunity syntaxes for the internal data.

Such a concept is not restricted to centrally managed resources, but can easily be extended to other data stores on the web that are operated by different CA subcommunities and offer a minimum of Linked Data facilities. There are draft versions of resource descriptions about Fano Polytopes (8630 records) and Birkhoff Polytopes (5399 records) hosted by Andreas Paffenholz and about Transitive Groups (3605 records) from the Database for Number Fields of Jürgen Klüners and Gunter Malle that point to external resources. This part of our project requires further solicitation.

It was one of the great visions of the SymbolicData Project to collect not only benchmark and testing data but also valuable background information about the records in the database as, e.g., information about papers, people, history, systems etc. concerned with the examples in our collection. Within the redesign we developed a general concept of the `sd:Annotation` RDF class to store background information in a unified way.

We use that concept in particular to relate bibliographical entries of type `sd:Reference` to different data records. The management of bibliographical references was completely redesigned with SymbolicData version 3 exploiting RDF and the established Dublin Core ontology [2] to represent bibliographical information in a way that is queryable by standard means and tools. On the other hand, we strongly reduced the part of information about bibliographical references kept inside SymbolicData since there are comprehensive bibliographical stores available on the web that provide all required information.

4 Towards a Computer Algebra Social Network

From the five stars to be assigned to a Linked Data project according to Tim Berners-Lee's classification [1] SymbolicData earned four stars so far (for offering data in interoperable RDF format on the web and providing a SPARQL querable triple store). For the fifth star one has to build up stable semantic relations to foreign knowledge bases and thus become part of the Linked Open Data Cloud [8].

Much of such interrelation, e.g., a list of interoperability references for people, software and bibliographical data with the Zentralblatt, is on the way. Moreover, we joined forces with the efforts of the board of the German Fachgruppe to store and provide information about people and groups working on CA topics at their new Wordpress driven web site [4]. We developed a first prototype to store this information in RDF format, to extract it by means of SPARQL queries and to view it on the web site using the Wordpress shortcode mechanism via a special Wordpress plugin. We apply the same technique to maintain information about upcoming conferences at this site.

The vision of a Computer Algebra Social Network goes far beyond that: Set up and run within the CA community a semantic aware Facebook like Social Network and contribute to it about all topics around Computer Algebra using tools that express your contributions in an RDF based syntax that the community agreed upon. This sounds quite visionary, but it is in no way utopic. We operate a first prototypical node of a tool that realizes the challenging concept of a *Distributed Semantic Social Network* [15], see our wiki for more information.

Due to page restrictions we cannot explain this and the new RDF based techniques in more detail and refer to the extended version [6] of this paper.

References

1. Berners-Lee, T.: 5 stars for Open Data. `http://5stardata.info/` [2014-03-05]
2. DCMI Metadata Terms. `http://dublincore.org/documents/dcmi-terms/` [2014-02-27]
3. Das eScience-Forschungsnetzwerk Sachsen. `http://www.escience-sachsen.de` [2014-02-19]
4. Website of the German Fachgruppe Computeralgebra. `http://www.fachgruppe-computeralgebra.de/` [2014-03-06]
5. FRISCO – A Framework for Integrated Symbolic/Numeric Computation. `http://www.nag.co.uk/projects/FRISCO.html` [2014-02-19]
6. Gräbe, H.-G., Nareike, A.: The SymbolicData Project – from Data Store to Computer Algebra Social Network. Computeralgebra-Rundbrief **54** (2014, to appear).
7. Heinle, A.: The SDEval framework. `http://symbolicdata.org/wiki/SDEval` [2014-02-28]
8. Linked Data. `http://en.wikipedia.org/wiki/Linked_data` [2014-03-05]
9. Nareike, A.: The SDSage package. `http://symbolicdata.org/wiki/PolynomialSystems.Sage` [2014-02-28]
10. OntoWiki: A tool providing support for agile, distributed knowledge engineering scenarios. `http://aksw.org/Projects/OntoWiki.html` [2014-02-19]
11. The PoSSo Project. `http://posso.dm.unipi.it/` [2014-02-19]
12. Sage – a free open-source mathematics software system. `http://www.sagemath.org` [2014-02-19]
13. The SymbolicData project wiki. `http://symbolicdata.org/wiki`
14. The SymbolicData SPARQL endpoint. `http://symbolicdata.org:8890/sparql`
15. Tramp, S. et al.: DSSN: towards a global Distributed Semantic Social Network. `http://aksw.org/Projects/DSSN.html` [2014-03-06]
16. Virtuoso Open-Source Edition. `http://virtuoso.openlinksw.com/` [2014-02-19]