The SymbolicData Project – from Data Store to Computer Algebra Social Network
Talk given at the AKSW Colloquium

Hans-Gert Gräbe

Leipzig University, Germany
`http://bis.informatik.uni-leipzig.de/HansGertGraebe`

Leipzig, 2014-03-17

Aim and Scope

Vision:

- Develop concepts and tools for profiling, testing and benchmarking Computer Algebra Software (CAS) from different areas of Computer Algebra
- Collect and interlink relevant data and activities from different Computer Algebra subcommunities

SymbolicData is an

- inter-community project that has its roots in the activities of different Computer Algebra Communities and
- aims at interlinking these activities using modern Semantic Web concepts.

Tools and data are designed to be used both

- on a local site for special testing and profiling purposes
- and to manage a central repository at http://www.symbolicdata.org.

What does SymbolicData offer?

Data:

- Polynomial Systems Solving
- Geometry Theorem Proving
- Fano Polytopes (A. Paffenholz)
- Free Algebras
- G-Algebras
- Test Sets from Integer Programming

Draft:

- Birkhoff Polytopes (A. Paffenholz)
- Transitive Groups (J. Klüners, G. Malle)

What does SymbolicData offer?

Tools:

- SDEval Package (Albert Heinle)
  - Aim: Set up, run, log, monitor standardized Computations on SD data series in a reliable way
  - Technology: Python standalone on top of the OS
- SDSage Package (Andreas Nareike)
  - Aim: Call the new Polynomial Systems format from Sagemath
  - Technology: Sagemath Python Package

Referring to the talk by Andreas Nareike last year, I will not touch that topic in my talk this year.

What does SymbolicData offer?

Infrastructure:

- Github repositories (following the Integration Master Pattern)
- A project wiki at http://symbolicdata.org
- A mailing list
- Web access to the XML resources
- A centrally operated OntoWiki based RDF data store of meta informations Based on the Virtuoso RDF store
- Organized along Linked Data Principles
- Regular Dumps of RDF data in Turtle format
- A SPARQL endpoint to query the data
- Advise for easy local installation of tools and data based on Virtuoso and a local Apache Web server (OntoWiki optional)

Links

- `http://symbolicdata.org` – the SD Wiki
- `http://symbolicdata.org/XMLResources` – the SD XML Resources
- `http://symbolicdata.org/RDFData` – the SD RDF Data Turtle Files
- `http://symbolicdata.org/Data` – the SD OntoWiki view on RDF data
- `https://github.com/symbolicdata` – the SD Repository at github

Some History

ISSAC 1998: Special session on Benchmarking

1999-2002: Phase 1 – Olaf Bachmann, Hans-Gert Gräbe

- Focus: Polynomial Systems, tools and concepts
- Technology: XML-like special markup, elaborated Perl tools

2005-2007: Phase 2 – around the Groebner Special Year

- Focus: Geometry Theorem Proving, first interlinking projects with the GB bibliography and the GB facilities projects
- Technology: Switch to true XML concepts

2012-2014: Phase 3 – E-Science Saxonia supported project (Andreas Nareike, Hans-Gert Gräbe)

- Focus: Switch to Linked Data and Semantic Web concepts, XML resources, RDF meta data, data reorganization
- Release of version 3 in Sept. 2013

Linked Data Principles

- *Resources:* URI, HTTP Get access
  Access to worldwide distributed data in a unified way
- *Resource Descriptions:* Deliver a valuable piece of information in structured RDF format, that can be combined with other pieces of information from other sources into new RDF sentences.
- Run *RDF Triple Stores* as part of a worldwide distributed data storage infrastructure
- Run *SPARQL Endpoints* on RDF triple stores

SymbolicData Data Structure

Resources:

- SD provides own resources in an XML based format
  - Polynomial Systems, Geometry Theorem Proving, . . .
- Draft: SD addresses other resources at different stores
  - Polytopes, Transitive Groups
- Maintenance of resources requires special semantic knowledge, semantic aware tools and semantically educated people

Resource Descriptions:

- Precomputed fingerprints of the different resources in RDF format to navigate and search within the data
- Requires semantic knowledge to use the fingerprints in an appropriate way

SymbolicData Data Structure

An example in Turtle syntax:

```
<http://symbolicdata.org/Data/Ideal/Czapor-86c.Flat>
  a sd:Ideal ;
  rdfs:comment "Flat variant of Czapor-86c" ;
  sd:createdAt "1999-08-27" ;
  sd:createdBy sdp:Bachmann_O ;
  sd:hasDegreeList "3,3,3" ;
  sd:hasLengthsList "4,4,4" ;
  sd:relatedPolynomialSystem
    <http://symbolicdata.org/Data/IntPS/Czapor-86c> ;
  sd:hasVariables "x,y,z,a,b,c,d,e,f,g,h,k" .
```

rdfs:, sd: and sdp: are namespace prefixes.

SymbolicData Data Structure

Use SPARQL to search for examples with given degree and lengths lists.
Run the following query at http://symbolicdata.org:8890/sparql

```
PREFIX sd:  <http://symbolicdata.org/Data/Model#>
select ?a
from <http://symbolicdata.org/Data/PolynomialSystems/>
where {
  ?a a sd:Ideal .
  ?a sd:hasLengthsList "4,4,4" .
  ?a sd:hasDegreeList "3,3,3" .
}
```

SymbolicData Data Structure

Linked Data: Link directs to a valuable resource description as, e.g.,
`http://symbolicdata.org/Data/Ideal/Sym1_311.Homog`

```
<http://symbolicdata.org/Data/Ideal/Sym1_311.Homog>
   sd:createdAt "1999-06-04" ;
   sd:createdBy sdp:Bachmann_O ;
   sd:hasDegreeList "3,3,3" ;
   sd:hasLengthsList "4,4,4" ;
   sd:hasVariables "x,y,z,hv" ;
   sd:homogenize sdideal:Sym1_311 ;
   sd:homogenizedWith "hv" ;
   a sd:HomogeneousIdeal, sd:Ideal ;
   rdfs:comment "Homogenized version of Sym1_311" .
```

It is the homogenized version of another example.

SymbolicData Data Structure

Hence Resource Descriptions have to provide rules to compute derived examples from basic ones.

- Requires semantic aware tools to extract the derived examples from the basic ones.
- No strong restriction, since most of real applications work within semantic aware environments anyway.
- Realized for Polynomial Systems and the Sagemath system by the SDSage package of Andreas Nareike.

SymbolicData Data Structure

Background information: Use RDF to manage additional data, try to interlink that data with other sources along the Linked Data Principles.

- Annotations – a system of background information on different examples and series of examples
- Bibliography – bibliographical references system (to be aligned with ZBMath)
- Conferences – data base of upcoming conferences
- People – different people and groups (to be aligned with ZBMath)
- Systems – list of CA systems (aligned with swmath)

Towards a CA Social Network

Valuable background information is information the people care about.
Try to gather data only once, but in a form that it can be multiply reused.
Build views (web sites) that harvest information.
Care about the rules

- Maximizing reuse minimizes use.
- Make things as simple as possible, but not simpler.

Vision:

- People – enlarge the database, link it to the ZBMath people database. Used to display people from the CAFG Board within the Wordpress based CAFG site.

    `http://www.fachgruppe-computeralgebra.de/fachgruppenleitung/`

Towards a CA Social Network

Vision (continued):

- Groups – collect standard information about CA working groups. Used to display such information within the Wordpress based CAFG site

    `http://www.fachgruppe-computeralgebra.de/arbeitsgruppen/`

- Conferences – do not only send conference announcements around mailing lists, but store it in a commonly agreed format within a CA Social Network.
  A very first prototype is used to display such information within the Wordpress based CAFG site

    `http://www.fachgruppe-computeralgebra.de/tagungsankuendigungen/`

Vision (continued):

- The stakeholders understand, that this is a techno-social, and even more a social than a technical process that is best discussed on the Symbolicdata Mailing list.

- The German CA Social Network germ at

    `http://symbolicdata.org/wiki/CASN`

    matures thanks to common efforts and get companions all over the world.

Towards a German CA Social Network

Much (technical) similarities with

- the aim and scope of the DSSN project
  `http://aksw.org/Projects/DSSN` and the Xodx prototype
  `http://aksw.org/Projects/Xodx`
- the overall design of the AKSW website

> *So let's talk about synergies to dig.*
> *But be aware: The project of a German CA Social Network*
> *to mature is a techno-**social** project.*

German CA Social Network Bootstrap

What does exist yet?

- `http://symbolicdata.org/Data`,
  `http://symbolicdata.org:8890/sparql` – OntoWiki and
  SPARQL Endpoint of symbolicdata.org, the RDF data store serving
  the international community.
- `http://symbolicdata.org/CASN-OW`,
  `http://symbolicdata.org:8891/sparql`,
  `http://symbolicdata.org/CASN` – OntoWiki, SPARQL Endpoint
  and Xodx instance to serve the German CA Social Network. It
  contains
  - (for the moment a copy of) the SD Person Data – 274 records
  - CAFG-Intern extended Person Data – 74 records
  - German Working Groups – 24 records
  - Conference Announcements – 25 records
  - References to papers in the CA-Rundbrief – 58 records