The SymbolicData Project
Towards a Computer Algebra Social Network
Talk given at the International Conference on Infrastructures and Cooperation in e-Science and e-Humanities

Hans-Gert Gräbe, Simon Johanning

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

Leipzig, 2014-06-05
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

I will not touch that topic in my talk.
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)
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 in Linz
- Focus: Geometry Theorem Proving, first interlinking projects with the GB bibliography and the GB facilities projects
- Technology: Switch to true XML concepts

- 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
  - URI = Unique Resource Identifier
  - 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.
  - RDF = Resource Description Framework
- Run **RDF Triple Stores** as part of a worldwide distributed data storage infrastructure
  - Triple: subject predicate object. as the basic RDF information unit.
- (Federated) Query Language **SPARQL**
- Run **SPARQL Endpoints** on RDF triple stores
RDF Basics

Main idea: Store pieces of information as triples.

- Subject and predicate have to be URIs, object can be an URI or a literal (i.e., plain or typed text)
- Different ASCII storage formats (RDF-XML, JSON, Turtle) and tools to parse and transform these formats
- We use the Redland RDF Libraries http://librdf.org/
- Representation as (directed) RDF graph: Subjects and objects as nodes (literals as annotated blank nodes), predicates as (labelled) edges.

Allows for navigation within the data: SPARQL Query Language

RDF allows to describe Resources, Concepts (i.e., meta information about Resources), Ontologies (i.e., meta information about Concepts) etc. in a uniform way.
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.
  It requires *semantic* knowledge both to compute fingerprints and to use them in an appropriate way.
SymbolicData Data Structure

Resource Descriptions (cont.)

- **Background information**: Use RDF to manage additional data, try to interlink that data with other sources along the Linked Data Principles.
  - Annotations – a notational system to associate background information to different examples and series of examples
- Bibliography – bibliographical references system (to be aligned with ZBMath)
- People – different people and groups (to be aligned with ZBMath)
- Systems – list of CA software (aligned with swmath)
Towards a CA Social Network

Valuable background information is information the people care about.

- Find out the places where such information is spread today. Usually it is *streamed*, not *stored*.
- Try to semantically annotate that information.
- Build views (web sites) that harvest information.

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
- Groups and Projects – collect standard information about CA working groups and their projects
  - Used to display such information within the Wordpress based CAFG site
Towards a CA Social Network

Vision (continued):

- 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 Wordpres based CAFG site
- 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 CASN germ at http://symbolicdata.org/wiki/CASN matures thanks to common efforts.
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