The SymbolicData Project
Towards a Computer Algebra Social Network
Talk given at the CICM 2014 Work in Progress Section

Hans-Gert Gräbe, Andreas Nareike, Simon Johanning

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

Coimbra, 2014-07-07
Aim and Scope

Scope

- 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
- http://symbolicdata.org/wiki/SDEval

SDSage Package (Andreas Nareike)
- Aim: Call the new Polynomial Systems format from Sagemath
- Technology: Sagemath Python Package

Short demo on local data and sdsage.
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
RDF and Linked Data Data Principles

- **RDF = Resource Description Framework**
  - Main idea: Store pieces of information in a unified way as triples, use standard tools to manage these data.
- **Resources**: URI, HTTP 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.
- Run **RDF Triple Stores** as part of a worldwide distributed data storage infrastructure
- (Federated) Query Language **SPARQL**
- Run **SPARQL Endpoints** on RDF triple stores
SymbolicData Infrastructure

- Main repository http://github.com/symbolicdata and several clones (following the Integration Master Pattern)
- A project wiki at http://symbolicdata.org
- A mailing list
- Web access to the XML resources
- Two centrally operated Virtuoso based RDF data stores for meta informations ('Data' and 'casn')
- Organized along Linked Data Principles
- Regular dumps of RDF data in Turtle format
- Two SPARQL endpoints to query the data
- Advise for local installation of tools and data based on Virtuoso and a local Apache Web server
The SymbolicData Project

Gräbe, Nareike, Johanning

Aim and Scope
What does SymbolicData offer?
Some History
Basic Concepts
SymbolicData meets RDF
Towards a CA Social Network (CASN)
Links

SymbolicData Data Structures

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 Structures

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 (CASN)

How to turn a DDS\(^1\) into a vivid, well recognized Social Network with plenty of valuable background information?

Central observation: Valuable background information is information that 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 such information.

\(^{1}\)DDS = Dead Data Store
An RDF based Road Map to a CASN

How to reach such a goal with RDF based semantic technologies?

- Main idea: Turn passive users into active ones.
- Identify and shape appropriate ontologies.
- Collect RDF data of such types, link to other sources along the Linked Data Principles.
  A very first prototype is used to collect such information and to display it within the Wordpress 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/casn matures thanks to common efforts.
What is already done?

http://symbolicdata.org/casn/CAFG-Intern/

Basic information about People – 410 instances of RDF type `foaf:Person` (i.e., passive users) from different sources. Used in particular to display people from the CAFG Board within the Wordpres based CAFG site.
What is already done?

http://symbolicdata.org/casn/WorkingGroups/

Standard information about CA Working Groups – 17 Instances of RDF type foaf:Group and sd:WorkingGroup from the old CAFG site. Used to display that within the Wordpress based CAFG site.
What is already done?

http://symbolicdata.org/casn/SPP-Projekte/

Standard information about CA Projects – 60 instances of RDF type sd:Project, compiled from the list of projects within the SPP 1489 priority program.
What is already done?

http://symbolicdata.org/casn/UpcomingConferences/

Information about upcoming CA conferences – 60 instances of RDF type sd:Event, compiled from different sources. Used as input for the printed version of the CA Rundbrief.
What is already done?

http://symbolicdata.org/casn/Dissertationen/

Information about dissertations in CA – 28 instances of RDF type bibo:Thesis, compiled from the CA Rundbrief.
What is already done?

http://symbolicdata.org/casn/CAR-Beitraege/

Information about articles in the CA Rundbrief – 75 instances of RDF type sd:Reference to be displayed at the website of the German Fachgruppe.
What is already done?

http://symbolicdata.org/casn/News/

A first approach to Annotated News – 2 instances of RDF types `sioc:BlogPost` and `bibo:Document` related to blog posts on the website of the German Fachgruppe.

No picture – pure harvesting functionality to be used with SPARQL querying.
Aim and Scope

What does SymbolicData offer?

Some History

Basic Concepts

SymbolicData meets RDF

Towards a CA Social Network (CASN)

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