The SymbolicData Project

Talk given at the Special Semester
on Groebner Bases Linz 2006

Hans-Gert Gräbe,
Dept. Computer Science, Univ. Leipzig, Germany
http://www.informatik.uni-leipzig.de/~graebe

March 8, 2006

Motivation

(1) For different purposes algorithms and implementations are tested on certified and reliable data.

(2) The development of tools and data for such tests is usually "orthogonal" to the main implementational efforts.

(3) In many cases tools and data could easily be reused - with slight modifications - across similar projects.

> *The SymbolicData Project is set out to coordinate such efforts within the Computer Algebra Community.*

(4) Commonly collected certified and reliable data can also be used to compare otherwise incomparable approaches, algorithms, and implementations.

(5) Benchmark suites and Challenges for symbolic computations are not as well established as in other areas of computer science.

Main Goals

(1) Collect data of examples from various areas of Computer Algebra

- *in a systematic and uniform way together with related background information;*
- *in a form that conveniently allows to extend, manipulate, and categorize the collected data;*
- *such that they can be extracted in a form readable by different Computer Algebra Software;*
- *such that interrelations of the collected data can be specified;*

(2) Share best practice experience how to run test computations on these data, i.e.,

- *to prepare data for input to different Computer Algebra Software;*
- *to set up, start, time, interrupt, and monitor the computations;*
- *to collect, analyse, and evaluate output data from these computations;*

(3) Run a net of Web sites that present data from the collection and results of test computations.

Some History

The SymbolicData project started at the ISSAC-98 special session on Benchmarking organized by H. Kredel. Mainly driven by Olaf Bachmann (Singular group, Kaiserslautern) and Hans-Gert Gräbe (Univ. Leipzig).

1999: joint forces with the symbolic computation groups of the University of Paris VI (J. C. Faugere, D. Lazard), of Ecole Polytechnique (J. Marchand, M. Giusti), and of the University of Saarbrücken (W. Decker).

1999: Incooperated into the benchmarking activities of the "Fachgruppe Computeralgebra" (Chair at those times: G.-M. Greuel).

Some History (2)

1999–2000: Main design decisions and implementations of the first prototype (O. Bachmann, H.-G. Gräbe) during two visits in Leipzig and Kaiserslautern.

> Flat XML-like syntax for data, Meta information stored in the same format, managed with elaborated Perl tools, set up a CVS repository.
>
> Data from Polynomial System Solving and Geometry Theorem Proving, Test computations at UMS Medicis, with main focus on Polynomial System Solving.
>
> Prototype was presented at the Meeting of the Fachgruppe Computeralgebra, Kaiserslautern, February 2000.

End 2000: O. Bachmann left the project (and science, and Kaiserslautern) for a new job.

Some History (3)

Main focus moved to Geometry Theorem Proving.

- M. Witte (Leipzig) digitized a great part of the 512 geometry theorems [Chou 88]
- Benchmark computations with the GeoProver package of H.-G. Gräbe.
- Talks at RWCA-02, ADG-02 and also in the CA-Rundbrief

Declining interest to really push the project during 2002-2005.

- Main focus 1 (2004/05): Move the format of data storage to a truly XML-based design and have the META information encoded as XSchema.
- Main focus 2 (2005/06): Incorporate new concepts of OWL ontology design.

Relaunch within the Special Semester on Gröbner Bases (Linz 2006).

Current State

Data: 350 Polynomial Systems, 297 Proof Schemes, 43 BIB items, 16 Testsets, 8 GAlgebra examples

Tools: Shift to collecting scripts as best practice examples

Community Tools: Mailing list with archive, CVS repository, domain www.SymbolicData.org, running on behalf of the CA-Fachgruppe on a server at the GI Bonn, release bundle of data for download from the Web site

Design: All data translated to true XML, first experiments with OWL (to be explained now in more detail)

Design of the SymbolicData Data Collection

Organized as a **Knowledge base** using the spirit of Web Ontology concepts as proposed in the W3C Recommendation for the **OWL Web Ontology Language**.

The data is divided into two parts, the XMLResources and the OWLResources.

XMLResources: Smallest indivisible units of information; XMLType described by a XSchema; stored locally to the SymbolicData distribution or (huge items) globally in locations all over the world

```
<XML XMLType="IntegerPolynomialSystem"
url="sdxml:INTPS/ZeroDim.example_61.xml"/>
```

OWLResources: Store information describing the XML-Resources and also relational information about them according to OWL design principles.

We use **Protege** (a Java based OWL tool with a big community) to develop and maintain a common ontology **Ontology.owl**.

For the moment, OWL classes are translated to XSchema and OWL individuals are stored as valid XML-files.

Individuals are identified according to their class and id.

```
<OWL xref="ZeroDim.example_7" class="INTPSAnnotation"/>
```