Description of Knowledge of Mathematical Programs with TEX and XML

Balázs Vecsei
BUTE Mathematical Institute,
5th year student
1111 Budapest, Egry József u. 1, Hungary
vecseib@math.bme.hu
http://www.math.bme.hu/~vecseib

Abstract

There are many free mathematical programs and databases on the Internet, but use of these programs can be difficult. In this paper I would like to show an XML-based mathematical knowledge description language. In this language I will describe the ability of different kinds of mathematical programs. In my opinion, XML is a useful base for this language, because it is well standardized, and there are tools for handling it.

Résumé

Il y a une multitude de programmes et bases de données mathématiques sur Internet, mais leur utilisation peut être difficile. Dans cet article nous présentons un langage de description de connaissance mathématiques basé sur XML. Dans ce langage nous décrivons les fonctionnalités de divers programmes mathématiques. Nous pensons que XML est un fondement utile de ce langage, grâce à sa normalisation et à l'existence d'outils de traitement de celui-ci.

Introduction

I think we need a mechanism to describe the knowledge about free mathematical programs, because more than 50 mathematical programs are on the Internet, under the GPL or other free licenses.

The documentation of the free mathematical programs is often written in a TEX-based language, and we can download it, along with the programs' source.

I use this documentation to describe the programs. I would like to write a more widely applicable program recommendation system, connected with other projects in this area. This recommendation system needs a machine readable description of free mathematical programs. Some general categories:

- *activemath*: The goal of the project's research and development is a web-based interactive learning system (for mathematics). The project provides an architecture, basic knowledge representation, and techniques for new-generation on-line interactive mathematics documents (textbooks, courses, tutorials) and e-learning [9].
- *mathweb*: This project supplies an infrastructure for web-supported mathematics [6].
  - a software system that connects a wide range of mathematical services via a common mathematical software base.
  - a set of mathematical services that support all aspects of doing mathematics on the web.
  - an interlingua for mathematical communication (OMDoc)
  - a portal for potential users and a discusssion forum for developers
- *openmath*: OpenMath is an emerging standard for representing mathematical objects with their semantics, allowing them to be exchanged between computer programs, stored in databases, or published on the web [11].

GAP (Groups, Algorithms and Programming) and TPTP (Thousands of Problems for Theorem Provers) provide a selection mechanism to choose the most useful method or program.

TPTP system recommendations The TPTP Problem Library is a library of test problems for automated theorem proving systems [15]. It can convert between a lot of theorem provers:

Bliksem	CoDe	CLIN-S
Dedam	DFG	DIMACS
FINDER	ILF	KIF
leanTAP	3TAP	METEOR
MGTP	OSCAR	OTTER
PROTEIN	PTTP	Satchmo
SCOTT	SEM	SETHEO
SPRFN	THINKER	Waldmeister

This table shows we need system recommendations for theorem provers, and I think, not only in logic.

GAP method selection GAP is a GPL-licensed computer algebra system. Its main area is a discrete algebra, especially groups. GAP has a method selection algorithm, and can select the optimal methods from among the applicable methods to given problems. They assert that a specific method will be better than a general method [1].

Describing mathematical resources

We can use the program documentation to describe the mathematical resources provided. There are a lot of different `toc` (table of contents) structures, so I think it is helpful to convert the `toc` to an XML-based language, because it is standardized.

The knowledge description file has two part, the `metadata` and the `toc`. The metadata contains data about the program or resource. The toc contains the converted `.toc` file. I use Emacs Lisp programs to convert the `toc` files to XML.

Metadata I use the Dublin Core Metadata standard in our mathematical knowledge description (mkd) language. I made it by hand with Emacs psgml, because I think that is the simplest method [5].

```
<resource>
<metadata>
<title>gap</title>
<creator>gap group</creator>
<date>2002-12-04</date>
<type>software</type>
<description>
gap is a system for computational discrete
algebra ...
</description>
<format io="in">plain txt</format>
<format io="out">plain txt</format>
<source>http://www.gap-system.org</source>
<rights>GPL</rights>
</metadata>
```

TOC This part is automatically generated from `manual.toc`. Details from `manual.toc`, the main manual of gap:

```
\chapcontents {17}{Combinatorics}{141}
\seccontents{17.1}{Combinatorial Numbers}{143}
...
```

Details from description of gap:

```
<toc>
<from>doc/ref/manual.tex</from>
<chapter title= "Combinatorics"
  chap= "17" page= "141">
<title chap = "17" sec = "1" page = "143">
  Combinatorial Numbers
</title>
...
</chapter>
</toc>
</resource>
</mkd>
```

Document Type Definition The DTD file defines the syntax of a particular XML language.

Details from `mkd.dtd`:

```
<!ELEMENT mkd (metadata, resource+)>
<!ELEMENT resource (metadata, toc+)>
<!ELEMENT toc (from, chapter+, title+)>
<!ELEMENT chapter (title+)>
```

Generate the question

In this process I suppose the question is in MathML or OMDoc format. But these formats are similar to assembly code in programming. We can't write or read the MathML or OMDoc source directly, so we need authoring tools.

- *omdoc*: OpenMath is a standard for representing mathematical data in as unambiguous a way as possible. It can be used to exchange mathematical objects between software packages, or via email, or as a persistent data format in a database. It is tightly focused on representing semantic information and is not intended to be used directly for presentation, although tools exist to facilitate this [7]. QMath is a batch processor that produces an OMDoc file from a plain Unicode text document [12].
- *mathml*: MathML is an XML application for describing mathematical notation and capturing both its structure and content.

  The goal of MathML is to enable mathematics to be served, received, and processed on the World Wide Web [2].

  There are a number of programs which can convert LaTeX to MathML, or can save in MathML format:

  - OpenOffice
  - Amaya, the W3C browser
  - LaTeX2HTML, with MathML package
  - Maple
  - Mathematica
  - TeX4ht, a general TeX to SGML/XML translator
  - TtM, a TeX to MathML translator
  - WeM, an MathML editor that converts a subset of LaTeX to MathML.

Tools

In this section I enumerate some useful tools for handling XML files. All of them are free. The Emacs Lisp programs are at `www.emacswiki.org/cgi-bin/wiki.pl`.

- *emacs* and *Emacs Lisp* [8].

- *psgml* [13], an Emacs major mode for editing SGML and XML documents.
- *xml.el*, a full XML parser, parses a file and returns a list; it is now part of GNU Emacs.
  - `(car (xml-parse-file "gap.mkd"))`
  - `(xml-node-name node)`
  - `(xml-node-attributes node)`
  - `(xml-node-children node)`
- *dom.el*, generates a document object model tree from an XML data structure.
  - `(dom-make-document-from-xml (car (xml-parse-file "gap.mkd")))`
  - `(setq oms (car (dom-document-get-elements-by-tag-name doc1 'OMS)))`
  - `(mapcar 'dom-attr-value (xpath-attribute-axis oms))`
- *xpath.el*, *xpath-parser.el*, handles XPath expressions. They are based on an Emacs Lisp implementation of a semantic parser (similar to *yacc* or *bison*).
- *rxp*, XML parser and validator, Debian package, Windows.
- *xsv*, XML schema validator, Red Hat package, Windows.
- *xsltproc*, XSLT transformer, Debian package, Windows.

Summary

In this paper I have presented a possible description of mathematical programs, as a main part of a program recommendation system. The program recommendation procedure is based on the following steps.

1. describe the mathematical resources;
2. generate an OMDoc and/or MathML question;
3. select the relevant data from descriptions and questions;
4. select a program;
5. write or convert the code;
6. select a service where we can run our code, or download the program.

References

[1] The GAP Group. *GAP — Groups, Algorithms, and Programming, Version* 4.3, 2002. http://www.gap-system.org.

[2] Ron Ausbrooks, Stephen Buswell, and David Carlisle et al. Mathematical markup language (MathML) version 2.0. Technical report, W3C, 2003. http://www.w3.org/TR/2003/WD-MathML2-20030411/.

[3] Neil Bradley. *Az XML kézikönyv*. Szak Kiadó, 2000.

[4] Robert J. Chassell. *Programming in Emacs Lisp: An Introduction*. FSF, October 1997. http://www.gnu.org/manual/emacs-lisp-intro.

[5] Dublin Core Metadata Initiative. *Dublin Core Metadata*, 2003. http://dublincore.org.

[6] Andreas Franke, Michael Kohlhase, Paul Libbrecht, and Jurgen Zimmer. *Mathweb*, 2003. http://www.mathweb.org/.

[7] Michael Kohlhase. OMDoc: An open markup format for mathematical documents. Technical report, Computer Science, Carnegie Mellon University, February 2003. http://www.mathweb.org/omdoc/omdoc.ps.

[8] Bil Lewis, Dan LaLiberte, and Richard Stallman. *GNU Emacs Lisp Reference Manual*. FSF, May 1998. http://www.gnu.org/manual/elisp-manual-20-2.5/.

[9] Paul Libbrecht. *Activemath*. DFKI Saarbruecken, 2002. http://www.activemath.org/.

[10] Brett McLaughlin. *Java és XML*. O'Reilly, Kossuth, 2001.

[11] The OpenMath Society. *OpenMath*, 2002. http://www.openmath.org/.

[12] Alberto Gonzalez Palomo. *QMath OMDoc authoring tool*, 2003. http://personal.iddeo.es/ret009t0/qmath/.

[13] Lennart Staflin. *Editing SGML with Emacs and PSGML*, 1996. http://www.lysator.liu.se/~lenst/about_psgml/.

[14] Richard Stallman. *GNU Emacs Manual*. FSF, August 1998. http://www.gnu.org/manual/emacs-20.3/emacs.html.

[15] Geoff Sutcliffe and Christian Suttner. *TPTP Manual*, 2003. http://www.cs.miami.edu/~tptp.