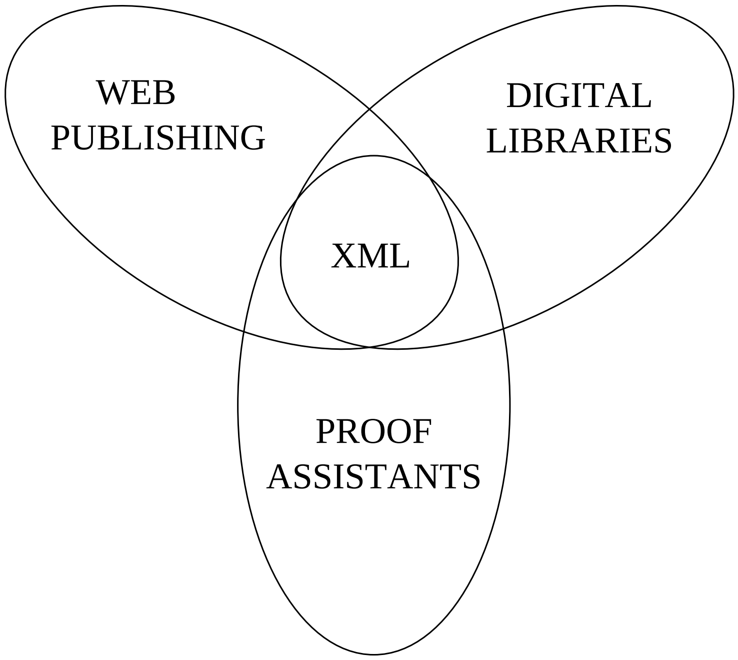
Formal Mathematics in MathML

Andrea Asperti, Luca Padovani, Claudio Sacerdoti Coen, Irene Schena
http://cs.unibo.it/helm/

Irene Schena
Phd-student in Computer Science
schena@cs.unibo.it

THE HELM PROJECT

The **Hypertextual Electronic Library of Mathematics** is meant to integrate the current tools for *automated reasoning* with the most recent technologies for the development of Web applications and electronic publishing, passing through the *Extensible Markup Language*.



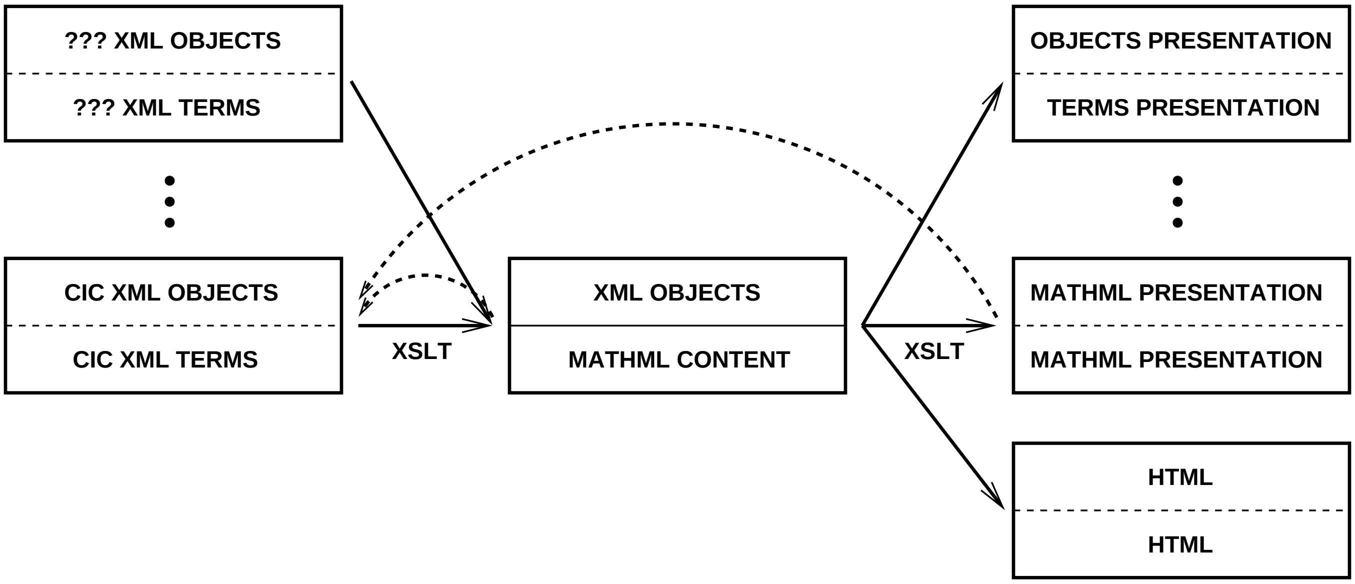
FORMAL MATHEMATICS

By **formalization** of mathematics we mean *expressing* mathematics, both statements and proofs, in a formal language syntactically defined by the rules of a grammar and a unambiguous semantics.

Formal Mathematics...

- is too *constricting* and too *verbose*
- lacks of any *intuition*
- lacks of any notation

THREE DIFFERENT LEVELS OF ENCODING



The backward arrows represent references from the content and presentation files to the corresponding XML files.

THE FORMAL LEVEL

A **Formal level** of encoding is necessary to:

- keep the formal semantics specific to the logical dialect
- verify the correctness of proofs
- allow complex searches
- proof-assistance and theorem-proving

Issues:

- content-oriented $\Rightarrow$ application-independent
- standard $\Rightarrow$ support for comunication and interaction
- extensible $\Rightarrow$ expressive

THE FORMAL LEVEL:

XML encoding

Why XML?

It is the flexible structured encoding which supplies the issues!

We have defined:

- a DTD which strictly reflects the abstract syntax of the CIC terms and objects
- a DTD which models a theory document, containing the formal objects, but not depending on the specific Logical Framework CIC
- a DTD specification for the annotations, providing the meaning of mathematical objects and terms in an informal but still structured language.

AN EXAMPLE OF XML ENCODING

```xml
<Definition name="proj1" params="0: A B">
<body>
 ...
</body>
<type>
 <PROD id="i18">
 <source>
  <APPLY id="i19">
   <MUTIND notype="0"
   uri="cic:/coq/INIT/Logic/Conjunction/and.ind"
   id="i20"/>
   <VAR reluri="0,A" id="i21"/>
   <VAR reluri="0,B" id="i22"/>
  </APPLY>
 </source>
 <target>
  <VAR reluri="0,A" id="i23"/>
 </target>
 </PROD>
</type>
</Definition>
```

THE INTERMEDIATE LEVEL

An **intermediate level**...

- gets rid of syntactical details
- improves modularity
- represents a common level

Issues:

- content-oriented
- standard
- extensible
- common

THE INTERMEDIATE LEVEL: MathML Content (1/2)

MathML Content Markup is...

- content-oriented (trivial!)
- a standard language
- conceived as an extensible language
- a common representation
- captures the *informal usage* of well-known operators (=)

The XSL Transformations is the standard mechanism to automatically generate the intermediate representation from a XML mathematical document.

THE INTERMEDIATE LEVEL: MathML Content (2/2)

MathML Content Markup:

- is designed to encode *classical* expressions, what about formal terms?
- cannot represent formal objects
- lacks of any associated standard semantics

We use:

- the *csymbol* extension mechanism to define new primitive formal operators (`arrow`)
- a XML level of encoding to describe formal objects (`Definition`)
- pointers to the associated low-level formal information (`helm:xref` attribute)

AN EXAMPLE OF MathML CONTENT

```
<Definition name="proj1"
 xmlns:m="http://www.w3.org/1998/Math/MathML">
  <Params>0: A B</Params>
  <body>
   ...
  </body>
  <type>
   <m:math
    xmlns:helm="http://www.cs.unibo.it/helm">
    <m:apply helm:xref="i18">
     <m:csymbol>arrow</m:csymbol>
     <m:apply helm:xref="i19">
      <m:and definitionURL="cic:/coq/INIT/Logic/
Conjunction/and.ind" helm:xref="i20"/>
      <m:ci helm:xref="i21">A</m:ci>
      <m:ci helm:xref="i22">B</m:ci>
     </m:apply>
     <m:ci helm:xref="i23">A</m:ci>
    </m:apply>
   </m:math>
  </type>
</Definition>
```

THE PRESENTATION LEVEL

At the **Presentation level** document, annotations and notations are combined together to specific target formatting languages.

Presentation is also a *vehicle of interaction* (hyper-links, cut & paste) provided that...

- the document format is flexible enough (PostScript, LaTeX, MathML)
- the application gives suitable support (*feedback*)

Current presentation of mechanized mathematics has:

- large proofs still unreadable
- in any case no deep intuition
- no notations

THE PRESENTATION LEVEL: MathML Presentation

The **MathML Presentation Markup** provides a standard on the web already implemented by several applications, and it is also a quite powerful rendering format.

A mathematical document is *transformed* in one or more steps and *combined* with other user-supplied documents:

- **Annotations** substitute/enrich fragments of the original document with more or less informal content:
  `(nat_ind ...)` $\Rightarrow$ By induction on $n\ldots$
- **Notations** are "standard" ways of rendering specific mathematical objects:
  `(le x y)` $\Rightarrow$ $\mathtt{x} \leq \mathtt{y}$

The XSL Transformations is the standard mechanism to automatically generate the associated notation from a XML mathematical document.

AN EXAMPLE OF MathML PRESENTATION

```
<m:math xmlns:m="http://www.w3.org/1998/Math/MathML"
 xmlns:helm="http://www.cs.unibo.it/helm">
 ...
  <m:mtext>DEFINITION proj1(O : A B) OF TYPE</m:mtext>
 ...
  <m:semantics xmlns:xlink="http://www.w3.org/1999/Xlink">
   <m:mrow helm:xref="i18">
    <m:mo stretchy="false">(</m:mo>
    <m:mrow helm:xref="i19">
     <m:mi helm:xref="i21">A</m:mi>
     <m:mo
      xlink:href="cic:/coq/INIT/Logic/Conjunction/and.ind"
      helm:xref="i20">
      <mchar name="wedge"></mchar>
     </m:mo>
     <m:mi helm:xref="i22">B</m:mi>
    </m:mrow>
    <m:mo color="Blue">&rarr;</m:mo>
    <m:mi helm:xref="i23">A</m:mi>
    <m:mo stretchy="false">)</m:mo>
   </m:mrow>
   <m:annotation-xml encoding="MathML">
    ...
   </m:annotation-xml>
  </m:semantics>
 ...
  <m:mtext>AS</m:mtext>
 ...
</m:math>
```

RENDERING

`http://cs.unibo.it/~lpadovan/mml-widget/`

GtkMathView is a GTK widget for rendering MathML (presentation) documents and it has been embedded into a first experimental browser for the Coq library.
Among the implemented features:

- hyper-linking definitions
- selection across presentational, intermediate and formal level
- annotations
- exporting to PostScript

FUTURE WORK AND CONCLUSIONS

By means of the standard **RDF**, we shall develop an *over*-structure of **metadata**:

- modeling the meta-information on mathematical documents
- modeling the meta-information on formal content
- accomplishing ad hoc and automatic searches on mathematical resources

Our work could be a first step towards a **content-centric** environment achieving interaction, sharing, communication and distribution.