INFORMATION SOCIETY TECHNOLOGIES
(IST)
PROGRAMME

Project IST-2001-33562 MoWGLI

Report n. D7.b
Technological Implementation Plan

Bernd Wegner

Project Acronym: MoWGLI
Project full title: Mathematics On the Web: Get it by Logic and Interfaces
Proposal/Contract no.: IST-2001-33562 MoWGLI

Contents

1 MoWGLI exploitation report

The following summarizes plans and initiatives for the exploitation of the achievements obtained during the MoWGLI-project. This will present a part of the potential applications of the tools and structures developed during the project. Further possibilities for exploiting the results for additional STI activities are under promising discussion wit the relevant parties like publishers and providers of mathematical information services.

1.1 The MoWGLI Final Prototype

The MoWGLI Final Prototype is a unified Web interface to manage the whole life cycle of a MoWGLI mathematical document. In particular the operations that can be performed on a mathematical document can be classified into:

1. Generation and indexing. Starting from a source document — either a LaTeX document of a Coq source file — one or more XML documents are automatically generated. Metadata are also automatically extracted or computed from the source files.

2. Administrative operations. For instance, a document can be updated or removed. An externally contributed document (i.e. a document that was not generated using the prototype) can be moved. Administrative operations for stylesheets and metadata are more numerous and complex.

3. Searching and rendering. Mathematical documents can be searched using the Moogle search engine and they can be rendered using the MoWGLI Final Prototype.

The prototype is still meant to be just a prototype that must be refined in a production environment. We consider now this part of the prototype mature enough for adoption for third part projects.

INRIA, the institute that develops the Coq proof assistant and that was also part of the project, has now a very positive impression and it is willing to adopt some of the tools developed in MoWGLI. In particular Moogle, the MoWGLI search engine, will be considered by them. INRIA will integrate Moogle in the Web interface and documentation of the Coq system. Moreover, CoqIDE (the graphical integrated development environment for Coq) will become a Moogle client. A user will be able to type a query inside CoqIDE and the query will be forwarded to Moogle. The XML document that describes the result of the query will be sent back to CoqIDE that will interpret it by showing the results to the user and loading the appropriate libraries in case the user is interested in the results.

This is a major achievement with respect to the current query facilities of CoqIDE: indeed Coq (and its GUIs) is now able to perform queries only on the libraries that have already been loaded. Since loading all the available libraries every time Coq is started is practically unfeasible (due to memory requirements and time constraints), the user is now obliged to restrict her attention of a predefined set of libraries, proving again results that are not in the chosen set and that she has no way to discover and locate. The practical consequences is that many efforts have been duplicated and there are several lemmas proved independently

several times by different people. In the future integration with Moogle is meant to avoid this phenomena, improving at once user productivity and the coherence of the Coq library.

Furthermore, INRIA already has a tool called coqdoc to automatically extract HTML documentation from the source files, following the literate programming approach. This is the tool currently used to document the standard library of the system. However, the tool does not allow the study of dependencies between proofs, lemmas, definitions and axioms, and it can show only the proof scripts, that are not meant for user consumption (and so they are hidden by default in the HTML pages). The current INRIA plan consists in integrating coqdoc and the INRIA pages with the MoWGLI rendering prototype by automatically generating hyperlinks to the corresponding MoWGLI pages. Thus the Coq website will be partially powered by the MoWGLI technology.

The Calculemus Network (founded in the 5th Framework Programme of the European Commission) extensively studied integration of deduction and computation systems, both from a technological point of view — how to achieve the system integration — and from the methodological point of view — what problems need this kind of integration to be solved. Several systems have been extended to allow easy communication with other mathematical agents, and "standard" interchange formats have been developed both in the Network and in the Monet IST project (5th Framework Programme). However, INRIA has not participated in these efforts and the Coq system provides no facility for communication with other tools. All the existing experiments of interaction, say, with computer algebra systems have been performed by unreliable communication at the textual level and ad-hoc implementations. INRIA is going to improve this situation by adopting the MoWGLI XML format for proof terms not only as an output format, but also as an input mode.

1.2 The GtkMathView Browsing Engine

The GtkMathView browsing engine provides two different backends: one which is meant to render MathML markup with support for interaction, and which is the one mainly used in the scope of the MoWGLI project, and a PostScript backend, which is capable of rendering MathML formulas to PostScript. The implementation of the PostScript backend, which was born more as an exercise rather than something needed, ended up being appealing to the John Wiley & Sons, Inc. publisher, which was considering the on-line publication of some scientific encyclopaedias. The publication process of this scientific material goes through several steps, among which is the re-encoding of the material in XML, with bits of TeX and LaTeX for the displayed mathematics. Because of the recent diffusion of MathML, the publisher adopted a parallel production line in which the bits of mathematics are encoded in MathML, with the goal of improving long term usability and exploitation of the on-line documents.

At that point the problem was finding a tool able to typeset mathematics encoded in MathML so that the result would be comparable to that achieved with TeX and LaTeX. GtkMathView seemed to be the most appealing candidate, and our collaboration in terms of support for the PostScript backend and consulting for MathML usage resulted in the publication, among others, of the following texts:

- Burger's Medicinal Chemistry and Drug Delivery (Abraham)

- Encyclopaedia of Catalysis (Horvath)
- Encyclopaedia of Smart Materials (Schwartz)
- Encyclopaedia of Software Engineering (Marciniak)
- Encyclopaedia of Polymer Science and Technology
- Handbook of Chemicals and Gases for the Semiconductor Industry (Misra)
- Occupational Toxicants and MAK Values (Deutsche Forschungsgemeinschaft)
- Stevens' Handbook of Experimental Psychology (Pashler)
- Textbook of Biochemistry (Devlin)
- Ullmann's Encyclopaedia of Industrial Chemistry (German branch of Wiley)

All the bits of displayed mathematics (formulas standing on their own) included in these texts have been rendered by the GtkMathView engine into PostScript, and finally converted into an image for direct inclusion inside the HTML page.

1.3 Hermes

The development of Hermes (a LaTeX into MathML conversion) finally has passed beyond the initial requirements of the MoWGLI project: it continues to grow towards a fully endowed semantic authoring and publishing tool guided by the feedback of the current actual users.

One of the "early adopters" of Hermes is Zentralblatt für MATH, a reference data base for mathematics, produced under the governance of FIZ Karlsruhe in Berlin. It had been decided to use Hermes for displaying answers to user queries from their reviews database, originally written in AMSTeX, in XML+MathML. The conversion of the about 90,000 records (abstracts) from Zentralblatt has been tested. The samples are covering all domains in mathematics typically handled by Zentralblatt. Along the way, the Hermes macro collection for AMSTeX has been developed. The conversion on the fly is working successfully, providing the user with an immediate convenient rendering of the formulas appearing in the reviews.

Another independent group at the Faculty of Informatics of Loránd Eötvös University, Budapest, Hungary, is already using Hermes to create a small number of fundamental books in computer science written originally in LaTeX in XML in a government funded project.

A group, led by Prof. Günter T̈ner, Duisburg University, Germany, has installed a project for validating and archiving TeX documents mathematics. The project is called TeXDocC. At a later stage Hermes should be used to move the emphasis on archiving XML documents and storing the TeX sources, instead of archiving the documents in the original TeX format. Main partner of this project is the State and University Library in Göttingen. They are involved in EMANI, the Electronic Mathematics Archiving Network Initiative with additional partners in China, USA and France. Through this cooperation Hermes will support the long-term preservation of digital documents in mathematics by enabling the storage of the documents in the MathML format.

Hermes will also be used, during 2005, after the end of the MoWGLI project, to convert some of Einstein's original papers into XML+MathML. This will be done in a collaboration with the Max Planck Institute for History of Science, Berlin, Germany. Living Reviews in Relativity and Hermes are in full collaboration now, aiming at complete conversion of the review articles originally in TEX source, into XML+MathML; the results are work in progress quality and available at `http://www.aei.mpg.de/hermes/livrev/`.

1.4 Extended OMDoc

DFKI has developed an extended OMDoc representation (structure and metadata model). It is ready for reuse by other groups and institutions. The further development is concentrating on the educational point of view. Exploitation by other groups will grow with the upcoming acceptance of MathML as a publishing standard in mathematics and the increased availability of MathML-coded documents through Hermes for example.