The Euromath System – The Structured Editor for Mathematicians

Janka Chlebíková

Department of Informatics Education, Comenius University, Bratislava

Keywords: TeX, SGML, structured document

Abstract. *The Euromath system is the result of a project funded through the SCIENCE programme of the European Commission and administered through the European Mathematical Trust. Its aim is to create a homogeneous computer working environment for mathematicians based on a uniform data model and to stimulate interchange among them based on modern information technology.*

The core of the system is a powerful SGML structure editor Grif combining the advantages of a WYSIWYG approach and structure editing. SGML is rapidly becoming the standard for publishing and for full-text databases. The Euromath system is at the forefront in exploiting the benefits of SGML for scientific documents and also the typesetting qualities of the TeX system.

SGML versus TeX

Mathematicians form a community in which TeX is widely used. The reason for this is simple: twenty-year-old TeX is still the best typesetting system for their documents! Mathematicians also use the TeX markup language for the mathematical formulae in electronic mail. If the goal is just to print a good-looking scientific document, TeX is perfect enough, with the possible exception of its lack of WYSIWYG approach. On the other hand, it is hardly possible to write applications for general TeX documents. The reason for this is that TeX (especially plain TeX) uses *specified markup* for its documents. This means that the content of a document is mixed with the presentation markup and logical elements of the document structure cannot be automatically recognized. In documents using standard macro packages such as LaTeX or AMS-TeX styles, markup tags for each logical element of the document structure can be used. For example in 'article.cls' style, tags for elements like `title, author,` `classifications, keywords, abstract`, etc. are available. These documents use *procedural* markup. Each tag for a logical element corresponds to a macro, which provides a presentation of the element. An advantage of such an approach is that the presentation of the document is separated from its content. This is the very important point for several reasons:

— the author can concentrate on the document content and not on typesetting aspects of the document,
— several different presentations can be used for the given document,
— applications can be created for such markup documents.

For example, from documents using style 'article.cls' it should be possible to generate an input to a database system that contains only some logical elements, e.g. `title, author, keywords` and `abstract`. But an author is not constrained (during creation of his document in LaTeX) to follow strictly the document structure and to use markup tags for elements. One can prefer to use his own presentation of an abstract and then he does not use markup tags for it. Such a document, of course, cannot be handled automatically.

This problem arises from the fact that TeX (including all its macro packages) is a typesetting system primarily intended to produce good-looking printed documents. In spite of the fact that one of the main aims of the LaTeX3 project is to overcome this problem and to separate the document structure from its content clearly, this problem cannot be solved very satisfactorily. It is natural because the TeX language was not intended to play a role of generalized markup language. Standard Generalized Markup Language (SGML) was accepted to this purpose as an international standard, to facilitate information exchange and reusability of documents (in other, even yet unknown contexts) via a description language. It is more suitable to use SGML in the situation, when future usage of the document is not necessarily restricted to printing. SGML is not a typesetting system; it serves as a meta-language, allowing new markup languages to be defined in a standardized way. This means, markup rules (Document Type Definition – DTD) can be defined for any class of documents. At present, many publishers are changing their publication systems for journals and books to SGML (for instance, Elsevier and the American Physical Society) to enable reusability of their publications.

Several articles have been written in which SGML and TeX possibilities are compared from several points of view, see for example [1], [2], [4] and [3]. Papers try to explain why "SGML should be inserted in front of TeX". The main reason for this is that for SGML documents the applications as well as translators to other typesetting systems (for example TeX) can be written. Thus, the input information can be confined to one file (the instance file) and many applications can be used for generating required outputs.

This is the central idea on which the Euromath system is based. The Euromath system uses the SGML format for its documents to allow building applications for them.

The Euromath system

The Euromath system is the result of a project funded through the SCIENCE programme of the European Commission and administered through the European Mathematical Trust. Its aim is to create a homogeneous computer working environment for mathematicians based on a uniform data model and to stimulate interchange among them based on modern information technology.

The core of the system is a powerful SGML structure editor Grif combining the advantages of a WYSIWYG approach and structure editing. The Grif structure editor was designed by V. Quint (Grenoble) and further developed by Grif SA (St. Quentin).

The Euromath system affords the same basic editing functions as non-structural text editors: file and print functions, edit functions such as undo, redo, find-replace, operations with clipboard, etc. It is also possible to change some layout parameters of current presentation for elements, for example type and size of fonts. The system contains a spelling checker for English, French and other languages. The possibility to incorporate easily graphics of various standards (Bitmap, Idraw, Postscript, Tiff) is given as well. The Euromath system facilitates simple WYSIWYG creation and handling of complex tables and commutative diagrams.

The possibility to deal with the document structure adds a new dimension to the Euromath system.

1. *The structure and layout of the document are given, the author of the document has only to take care of its content.*

During editing a user can see how the document looks in a satisfying presentation, while at the same time the system saves (behind the scenes) the document in SGML. The Euromath system forces the user to follow strictly the logical structure of the document and allows editing according to the relevant DTD (Figure 1).

The Euromath system contains DTDs for those types of documents that are frequently used by mathematicians: 'article.dtd', 'letter.dtd', 'sheet.dtd', 'slide.dtd', 'html.dtd', etc. These DTD correspond to LaTeX styles except 'html.dtd', which enables www-pages to be produced in the Euromath system (Figure 2).

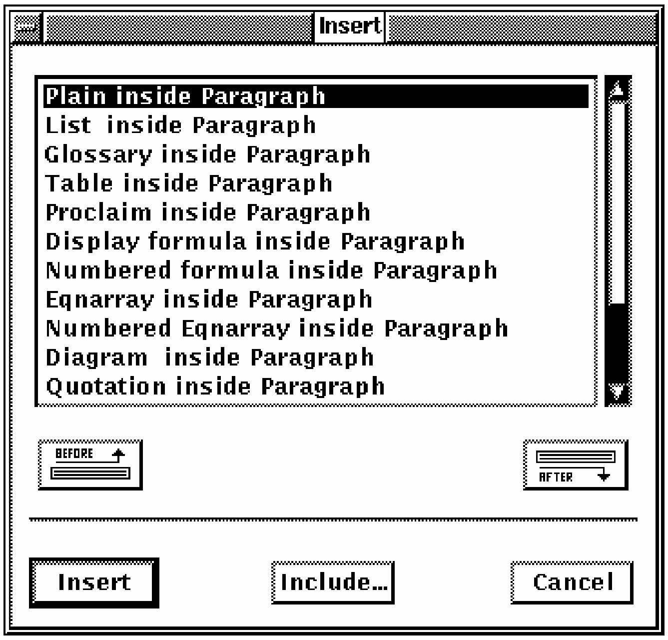

Figure 1 – Typical “Insert” window

The Euromath system was created as an open system, i.e. the possibility is given to the user to add new DTDs (with presentation) for arbitrary document classes. Thus, the Euromath system facilitates WYSIWYG editing of documents according to arbitrary publishers's DTD, the user's personal DTDs or for some standard DTD (if it will be accepted for scientific documents).

2. *For one document class (DTD) several different presentations can be defined.*

The document structure allows different kinds of presentation to be defined for its logical elements. For example, two presentations can differ from each other in the layout for some elements, such as `author` or `title` in the case of ‘article.dtd’. It is a substantial property of the system that one document (physically just one file) is edited and maintained for all its presentations.

For each DTD there exists the standard presentation ‘SGML’, in which the logical tree structure of the document is displayed together with tags for logical elements (which are usually hidden in other presentations). This format is similar to the one in which the document is saved.

The Euromath system enables the user to define his own presentation for any DTD. For instance, one can create a presentation for ‘letter.dtd’ which

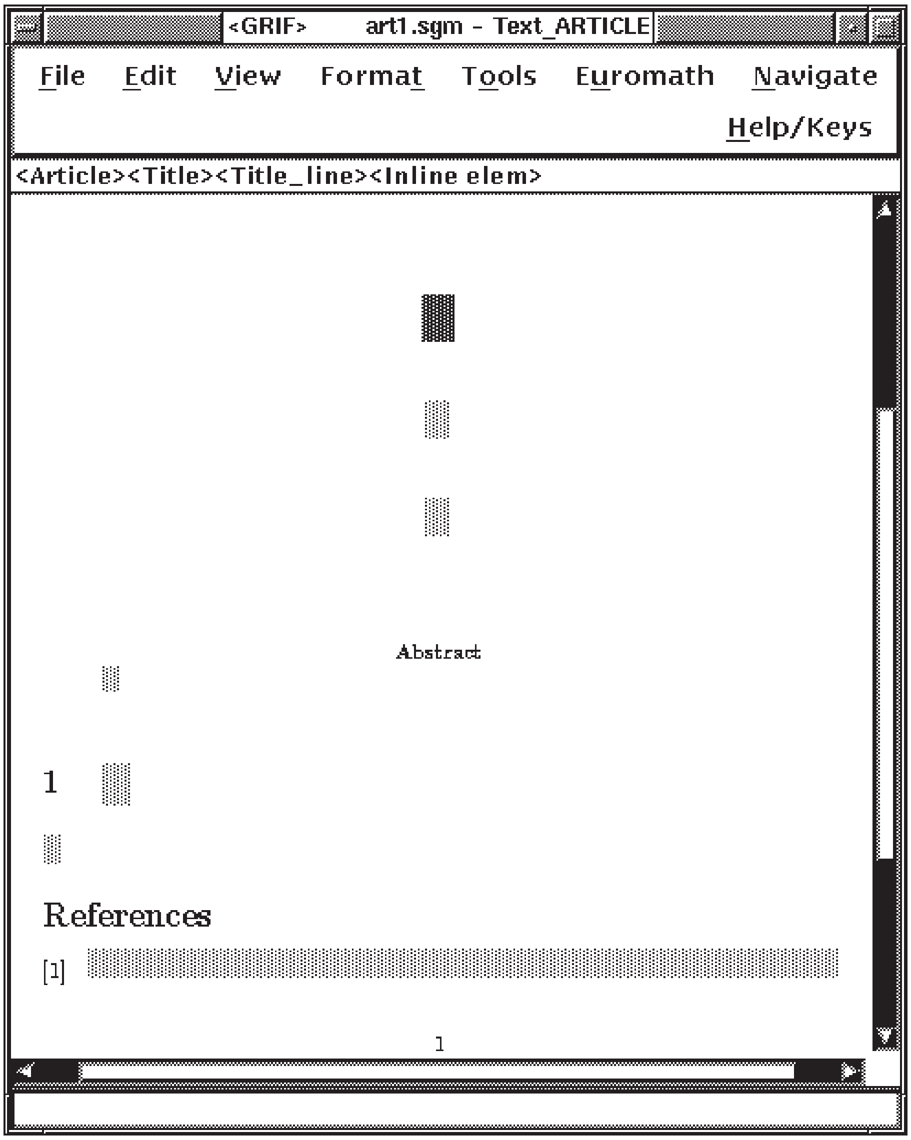

Figure 2 – The default template for the “article.dtd”

satisfies his personal requirements. Presentations are defined in P-language, which allows the layout of each logical element to be specified. Some features of the presentation (font size, indentation, etc.) can depend on fulfilment of certain conditions (for example, the first paragraph of a section, etc).

3. *The individual logical elements of the document can be displayed in several windows.*

For example in the case of 'article.dtd', the references or the table of contents can be displayed in separate windows and can be edited simultaneously (Figure 3).

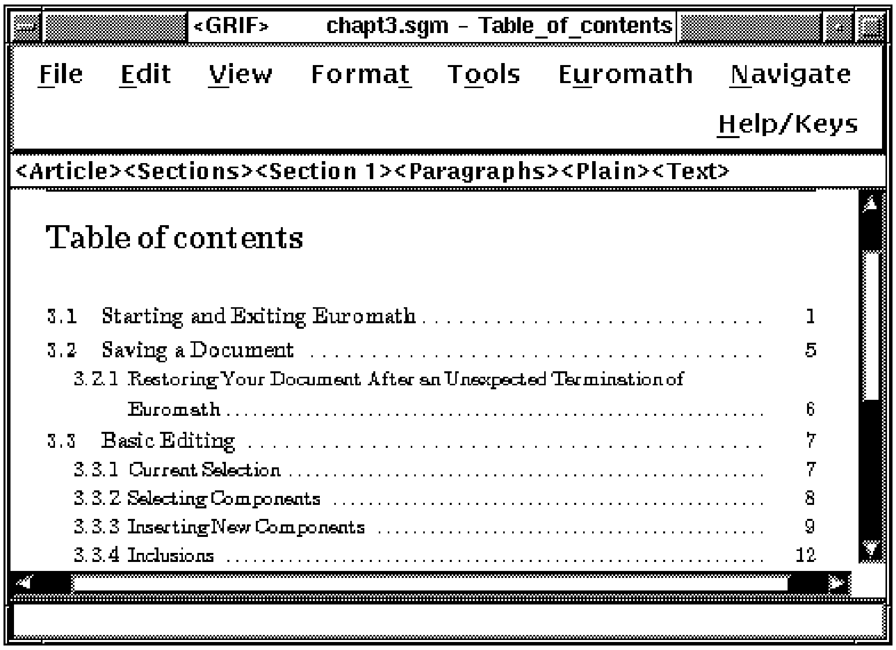

Figure 3 – "Table of contents" view

4. *Automatic housekeeping by the system.*

This feature is known also to users of LaTeX styles. The system takes care of updating numbering theorems, footnotes, cross-references and the creation of a table of contents according to the DTD currently used.

The system enables the user to create a cross-reference to specified logical elements (e.g. cross-reference to a section). These cross-reference nodes serve as hypertext nodes, just double clicking on them displays the referenced element. Cross-references can be made also between documents.

5. *Intelligent navigation through documents.*

The system enables fast movement through a document in several ways. As was mentioned before, cross-references in a document serve as hypertext nodes. It is possible to search for references to a particular logical element. The system also facilitates searching for logical elements of a specified type, e.g. tables and figures.

6. *The possibility to manipulate logical parts of the document.*

The system enables a user to handle whole logical parts of the document in several ways.

It is possible to copy whole logical parts under the condition that the logical structure of the document is preserved. It is also possible to change an element to an element of another type, if the latter one shares the same content model as the original element and it is at the same level in the document structure. For example, a numbered formula can be changed to a no-numbered formula.

The system offers the possibility to create some additional structure around an element, when the author decides to change its position in a tree-level structure, e.g. to redefine the selected mathematical expression to its square root. One can easily change attributes of logical elements, e.g. changing the item labels of a given list from numbers to stars (Figure 4).

All above-mentioned possibilities are case-sensitive and menu-guided, so the author is not supposed to be familiar with the document structure.

7. *Export to other formats.*

The Euromath system saves documents automatically in SGML format. The possibility to translate a document to other formats like LaTeX, plain ASCII text and HTML is given as well. As the system deals with a structured document, the translation to other formats is easy.

Before printing a document in its final form, one may find it more convenient to export it to LaTeX. This provides higher layout quality than printing directly from the system, which is not primarily focused to that purpose.

The user has the possibility to add an export to new formats for any DTD. Such new export is defined in T-language; for each logical element its export has to be described. The translation can depend on fulfilment of certain conditions similarly to in P-language.

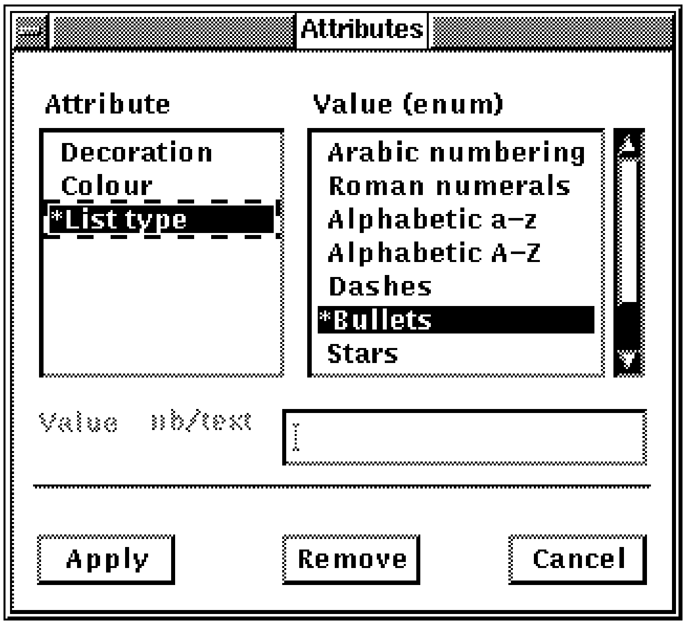

Figure 4 – “Attributes” window for list element

Euromath applications

The Euromath system is a general SGML editor with a programming interface that allows external actions to be attached to it. The Euromath applications extend the possibilities of the Euromath system as text editor and provide tools to help mathematicians in several ways.

Personal File System

The Personal File System (PFS) is an external program linked into the Euromath system and communicating with it (Figure 5).

The Personal File System connects the Euromath system with an electronic version of the database created by Zentralblatt für Mathematik (either on-line or from the CD-ROMs covering Zentralblatt reviews for the period 1931-1995). Searching through abstracts or bibliographic information about papers can be comfortably achieved using forms. The information found, that is automatically translated into SGML format, can be displayed and used by tools of the Euromath system (in particular, mathematical formulae are displayed in WYSIWYG mode).

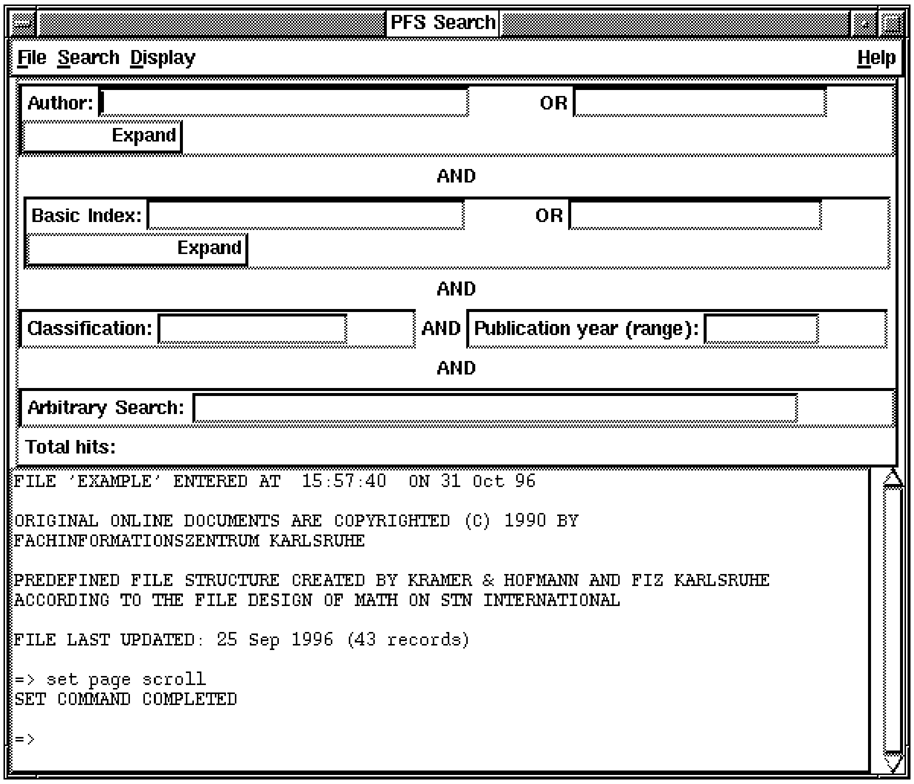

Figure 5 – "PFS search" window

A user can create his own local databases in the same format as that of Zentralblatt and use them in the same way.

Translation of TEX files

The Euromath system facilitates the translation of TEX files into SGML format. The best results are achieved from those TEX documents whose basic document structure can be easily recognized. For example, documents using LATEX style 'article.cls' are translated to 'article.dtd' style with very good results.

E-mail sender

An e-mail sender was implemented to enable a user to send electronic documents in SGML, plain ASCII text, LATEX or HTML format directly from the

Euromath system. When the document is sent in SGML format, its structure is preserved (including mathematical formulae) and hence it can be again handled by tools of the Euromath system. On the other hand, using plain ASCII text format any information about the document structure is lost, and the original structure cannot be determined from that format.

Interface to X.500 standard

The Euromath system provides a convenient interface to two directories based on the X.500 standard. One is the directory of European mathematicians and the other is the worldwide Paradise directory (which is intended to cover all higher education, research and development organizations in all countries).

The interface provides access to the directories via searching and browsing. It uses an SGML description of the X.500 data standard for representing the data. The directory is accessed via a user agent that understands SGML instructions and provides responses in SGML.

Retrieving documents across networks

With the Euromath system it is possible to retrieve files across a network. The combination of SGML and WYSIWYG makes the Euromath system an ideal editor for viewing remote SGML files. Formulae, tables and diagrams which are difficult to represent in plain ASCII text are shown immediately in the Euromath system; there is no need for any processing (as opposed to, for example, downloading TeX documents requires TeXing and previewing).

The Euromath system supports references to remote documents — just by double clicking on the reference in a document — and the remote document is opened on a screen.

The Euromath system has set up on an experimental basis a Euromath information server which is intended to contain documents of interest to Euromath users. These include Euromath News and Frequently Asked Questions with answers.

The Euromath system can be also used as a WWW browser for HTML documents.

Interface to computer algebra systems

Until now it was implemented only on an experimental level. The system will call on a proper mathematical program (e.g. Maple or MuPad) to evalute the selected numerical expression.

SGML versus TEX in Euromath System

The Euromath system is intended for mathematicians. They have become accustomed to TEX for publishing scientific papers. From this point of view, it is logical that the Euromath system tries to be related to TEX in several ways.

Similarity in styles

DTDs offered by the Euromath system ('article.dtd', 'letter.dtd', 'slide.dtd' and others) are based mainly on the similar LATEX styles, for example 'article.dtd' corresponds to 'article.cls'.

Interactive translator for mathematical formulae

The Euromath system offers two different ways of writing mathematical formulae in a document: in WYSIWYG mode, that chooses symbols from a palette following its logical structure, or as a LATEX string. Each formula can be switched at any moment from a LATEX string to WYSIWYG mode and vice versa. The translator between SGML (used in WYSIWYG mode) and LATEX format of mathematical formulae is based on the fact that both, LATEX and SGML, have sufficient tools to describe the logical structure of the mathematical formulae. Problems can appear in the case of mathematical formulae in which it is sometimes necessary to guess the structure that is implicitly given (but hidden) behind any LATEX formula.

Import to and export from TEX

The Euromath system allows TEX documents to be translated into DTDs used by the system and consequently to import documents into the editor. The best results are achieved from TEX documents in which the structure of the document is described using standard LATEX styles, for example 'article.cls'.

The translation in the opposite direction, from SGML to LATEX, is completely straightforward as the system deals with a structured document. (An appropriate LATEX style is chosen according to the DTD.)

Concluding remarks

SGML is rapidly becoming the standard for publishing and for full-text databases. The Euromath system is at the forefront in exploiting the benefits of SGML for scientific documents and also the typesetting qualities of the TEX system.

The latest version of the Euromath system is available for UNIX (X-windows systems) based on SUN platform. From the properties mentioned above the current version does not contain an interface to computer algebra systems. The new release will also provide the support for XML (Extensible Markup Language) and MathML (Mathematical Markup Language).

More information about the Euromath system can be found on the page `http:\\www.dcs.fmph.uniba.sk\~emt`.

Bibliography

[1] A. E. Dobrowolski. Typesetting SGML Documents Using TEX. *TUGboat*, Volume 12 (1991), No. 3 – Proceedings of the 1991 Annual Meeting, pp. 409–414.

[2] R. W. McGaffey. SGML versus/and TEX. *TUGboat*, Volume 12 (1991), No. 3 – Proceedings of the 1991 Annual Meeting, pp. 406–408.

[3] N. A. F. M. Poppelier, SGML and TEX in Scientific Publishing. *Proceedings of the TEX90 Conference*, pp. 105–109.

[4] C. G. van der Laan. SGML (, TEX and ...). *Proceedings of the TEX90 Conference*, pp. 90–104.