Technical Word Proccessing and Euromath

Michiel Hazewinkel
C.W.I., Amsterdam
POBox 4079
NL-1009AB Amsterdam
The Netherlands

(DRAFT VERSION, 23.4.87)

1.The Euromath project

The present day working scientists' environment consist principally of documents, documents, to be edited documents, and more documents. Much of his work is thinking and doing things about these documents. Add to this some work involving computational and dataprocessing codes and an 95% accurate description of the non-organiztional part of the more theoretical scientist's work is achieved. This holds in particular for mathematicians, and for the time being Euromath is, as the name implies, a project on behalf of Europe's 10.000 or so professional mathematians. Later expansions and applications of the same ideas to other sciences are expected, and in fact such considerations played a role when the CEC (Commission of the European Community) decided to fund the project.

Briefly the Euromath project aims to provide Europe's mathematicians with an integrated interface comprising electronic mail and conferencing, database access and search facilities, and electronic creation and transmission of mathematical documents. Scanners and document readers (plus the then necessary recognition software) will probably be also present somewhere in the system. Also the modern working mathematicians environment should include electronic formula manipulation facilities and simulation possibilities (for, so to speak, experimental mathematics, also known as calculating examples[1]).

In the many meetings and discussions that preceded the final formulation of the project and in a number of preliminary studies it became clear that one major concern is the electronic creation, transmission and re-editability of mathematical documents. The sections below deal with this particular aspect of Euromath.

The present status of the project is roughly as follows. The total project as formulated by us calls for funding in the neighborhood of 2M ECU (a bit over in fact). At the present time funding for the first phase has been granted by the CODEST Committee of the CEC to the amount of 750K ECU. This first phase involves amoung other things the basic requirements, feasability, and functionality studies; setting up the basic communications facilities and evaluation of what is available in the way of technical word processing.

Euromath is an initiative of the EMC (European Mathematical Council), chaired by Prof. Sir Michael Atiyah. At the present time its steering committee consists of Prof.

F. Topsøe (Copenhagen) (chairman), Prof. C.J. Mulvey (Sussex), Prof. J. van Lint (Eindhoven), and Prof. M. Hazewinkel (Amsterdam).

In terms of networks and machines the total Euromath system will eventually look roughly as follows:

a number of backbones or nodes, typically one in each country; some ot these will have extra special functions in terms of administration of the system, database input aspects of the sytem,... ; roughly machines at the level of a VAX 780, running under UNIX.

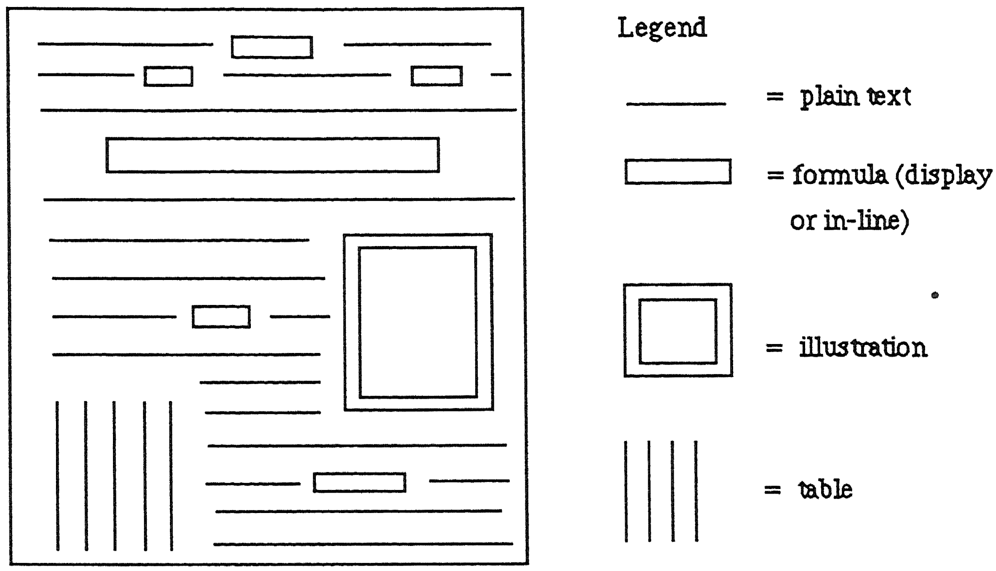
for each participating institute a machine at the level of a SUN 3 linked to the backbone machine of the country in question; typically the link will be via the local academic network.

a large number of micros and terminals hooked up to the various institute machines; here the present description of the project mentions Macintoshes, IBM PC compatibles, others (such as Amiga), in that order.

For a more complete but still non-technical description of Euromath, cf [2].

2.Mathematical documents

Euromath involves many things: distributed database problems, automatic updating facilities in a network, automatic siphoning off of relevant pieces of scientific papers and storage of these pieces in a (grey) literature database, integration aspects with respect to the various academic networks, various compatibility matters, ... However, basic for the whole idea of Euromath is no doubt providing mathematicians with the facilities to create mathematical documents electronically in a reasonably straightforward and easy manner, to transmit these documents through some form of electronic mail and to be able to re-edit a received document. Symbolically a mathematical document can be represented more or less as the illustration below.



The problems involved in illustrations and tables, i.e. basically layout problems, are not typical for the mathematical or scientists' world; indeed these problems tend in these worlds to be less crucial than in various other contexts. Also the various problems concerning footnotes, headers, footers, tables of contents, indexes, lists of references,... etc. are not our particular immediate concern. Certainly the problem bulking largest on the horizon of the electronically inclined scientist is how to deal with formulas, both displayed formulas and (simpler) formulas in a line of text as illustrated above.

As examples of the kind of formulas which occur as display formulas the following may perhaps serve:

$$U^{3\beta}_{m_1,n_1}(p_1,p_2)\delta_{n_1,n_2} + \int_0^\infty \frac{\delta p_3\, p_3^2}{3\pi^3} \sum_{\nu}\sum_{m}\sum_{\beta_2} (-1)^\mu \frac{U^{3,3}_{m_1,n_1}(p_1,p_2)}{\pi_3^2 - k^2}\, \zeta_{3m_1,n_1}\, \eta_\nu(p_3,p_2)$$

or the following diagram:

$$\begin{array}{ccccccc} 0 \longrightarrow & \mathrm{Hom}(U, \mathbf{C}^*) & \longrightarrow & \{\text{group of data } (H,\alpha)\} & \longrightarrow & \left\{\begin{array}{l}\text{group of hermitian} \\ H: V\times V \longrightarrow \mathbf{C} \\ \text{with } \mathrm{Im}(U\times U) \subset \mathbf{Z}\end{array}\right. \\ & \downarrow & & \downarrow & & \downarrow \\ 0 \longrightarrow & \mathrm{Pic}^0 X & \longrightarrow & \mathrm{Pic}\, X & \longrightarrow & \mathrm{Ker}\,[H^2(X,\mathbf{Z}) \rightarrow H^2(X, \end{array}$$

or perhaps a matrix like:

$$\begin{bmatrix} \frac{1}{2}e^{-q_1} & -p_1^2 & 0 \\ \frac{1}{\|\vec{x}_1\|} & \frac{1}{2}e^{-q_2} & -p_2^2 \\ 0 & \frac{1}{\|\vec{x}_2\|} & \frac{1}{2}e^{-q_3} \\ 0 & 0 & \frac{1}{\|\vec{x}_3\|} \end{bmatrix}$$

where it should be noted that the above is a rather simple matrix as such things go.

In-line formulas tend to be a good deal simpler and usually do not go much beyond expressions like the following ones: an integral with the simplest kinds of upper and lower bounds and a mildly complicated integrand like: $\int_0^1 \frac{f(x)dx}{(x-a)}$; or perhaps an expression involving a radical of something simple like: $\sqrt[3]{(x^2-b)}$; or, again a number of expressions involving over-arrows and overbars like: $\overline{1-\|\vec{x}\|_2^2}$, $\overline{\overrightarrow{AB}}$, $\overrightarrow{\overline{AB}}$, $\overline{\overline{(x-a)}}$; or an expression involving those ubiquitous operators 'inf', 'sup', and 'lim' like: $\inf_{x,y \notin A}$; or, finally, expressions involving roughly both level two subscripts and superscripts like the following ones $\gamma^{\underline{\rho}\ \underline{\sigma}}_{k_1\ k_2}$, $x^{p^2}_{i_1}$.

For curiosity's (and comparision's) sake: the first two of the formulae above were constructed using the formula mode of Word3.0; the last eight with the DA MacΣqn; the result of doing the first two with this DA is: $\int_0^1 \frac{f(x)\,dx}{(x-a)}$, $\sqrt[3]{(x^2-b)}$.

3. WYSIWYG and (S)(G)ML.

There are of course, globally speaking, basically two entirely different ways to produce, say, a displayed formula like the first one above. They go by the code names ML (= Mark-up Language) with the variants GML (= Generalized ML) and SGML (= Standard GML) and WYSIWYG (= What You See Is What You Get). For instance in TEX one types the following sequence: \int_0^{pi/2}\sin^2\theta\,d\theta to encode the formula $\int_0^{\pi/2} \sin^2\theta\, d\theta$. Using an ML one does not see the result on the screen, at least not directly as the formula is typed; it is difficult from such a string of symbols to see what the formula will look like, and in case there is a small mistake, it is very difficult to find it and correct it in many cases. Our experience at the CWI is that an ML system is fine for a mathematically inclined professional typist and for those scientists who really use it day in day out; it is awful for the casual user and our experience is that about half of our secretaries essentially never master it sufficiently. Even for those who can handle it fine it is needlessly complicated. ML's simply do not use the many (wordprocessing) facilities of a modern Micro like a Macintosh, which seems a waste. With a WYSIWYG processor like MacΣqn one types ⌘-i and an integral sign appears on the screen and the cursor goes to the lower bound position; type in anything and hit [illegible] and the lower bound appears on the screen and the cursor goes to the upper bound (so in the case at hand one types '0 [illegible]'); now type in 'π/2 [illegible]' and the cursor goes to the position for the integrand, etc.. A really good formula processor like this will also automatically make the integral sign larger if the expression behind it requires that and it will know about such typesetting conventions as that variables in formulas are set in italic (but names of standard functions like 'sin' are in roman), nominator and denumerator of a fraction are centered with respect to one another, etc., etc.

The point is that for Euromath we really need both ways of doing things simultaneously. For ease of input one wants things as much WYSIWYG as possible and one wants to be able to edit at this level; in any case an interactive system rather than a batch sytem like T_EX or TROFF/EQN. But for electronic transmission one needs an ASCII file in some SGML (which retains the structural information present in mathematical formulas).

This leads to the following desired set-up for the future Euromath TWP. It is an interactive (WYSIWYG) TWP which at the same time generates automatically the corresponding SGML file; say, a T_EX file. Thus it is so to speak a preproccessor for the relevant SGML. At the same time the TWP is able to read the SGML file and to produce from it not only a socalled preview, but an editable version of that; thus it also contains so to speak a reverse-pre-processor.

Incidentally, depending on the structure of the SGML involved it is theoretically quite possible that the WYSIWYG TWP is driven by essentially the same keyboard stroke combinations which would be needed to produce the SGML file.

Except for fine control of the final type-set result - if desired- the user of the TWP will never see the underlying SGML file. All this accords well with the various remarks in [3,5].

It is not clear to me, incidentally, whether the ML's T_EX and TROFF/EQN retain enough information for such a reverse pre-processor to work on; in particular I wonder about pre-super-scripts and pre-sub-scripts as e.g. in: ${}^{t}A^{m}_{kl}(x)$.

3. Feasibility

Certainly as a first step T_EX or TROFF/EQN would be acceptable as ML for the Euromath system. Both are widely implemented on many machines ranging from mainframes to micros. For instance T_EX is implemented on both the Mac and the IBM PC (and the version on the Mac which I have tried out a bit (Kellerman and Smith) is a joy to use (modulo the awful input sequences of course)). There remains the question whether preprocessors and reverse preprocessors as described in section 2 above exist and can exist. For the IBM PC T_EX implementation there appears to exist a satisfactory preprocessor called EasyTEX, but I have no first hand knowledge. A second implementation of T_EX by FTL can in any case read a MacWrite file complete with all the formatting information and produce the corresponding T_EX file.

A formula editor called Edimath for the Mac is a straight port from a corresponding ML editor on UNIX machines; the last is now part of a syntax directed editor called GRIF.

Finally at the CWI there is a prototype of a formula editor called INFORM (of the interactive (WYSIWYG) kind) which also produces the corresponding EQN file and which can read EQN files [6].

Thus all in all the outlook for such pre-processors and reverse-pre-processors as desired seems quite bright, though no doubt a substantial amount of work remains to be done.

One question which should not be ignored when specifying the communication ML is compatibility with such standards as have been developed and adopted by e.g. the AAP (American Association of Publishers). That is the ML files of the Euromath system should be downgradable to that SGML.

4. Customizability

There is one aspect concerning TWP's for scientists which I want to stress particularly at this point. It can be summed up by the word "customizability". For one thing despite the fact that the Harris phototypesetter at our institute offers some 1800 different symbols to play with we still manage to come up with distressing regularity with some signs which are not available. But things go much deeper than that. For instance the formula processor MacΣqn offers such constructs (templates) as 'Sum from ... to ...' and 'Product from ... to ... ', e.g. $\sum_{i=1}^{k} f(i)$, $\prod_{i,j,k} R_{ij}(k)$, but what about corresponding expressions with the S and P signs replaced by $\otimes$, or $\oplus$, or by intersection, ... The point is not that there are a number of well known kinds of formulae for which this particular editor makes no provision; the point is rather that no closed editor can possibly be satisfactory because scientists will always come up with a kind of formula which will be difficult to handle with any given editor-assuming it is possible at all. This consideration was the main reason why the creators of INFORM([6]) elected to make their editor grammar driven; it is a lot easier to change things in the grammar and to add things to it than to change things in the program itself.

5. Features and Desires.

Below I have made sort of a list of all the features one encounters in word processors such as are familiar to me from the various WP's for the Macintosh. I have grouped them into four categories in descending order of importance from the scientists point of view.

(i) Basic commands as under Edit and File, word wrap, rulers, font/style changes, import graphics.

(ii) formula processor, in-line formulas, flow past illustration or table, vertical and horizontal kerning, generate and read corresponding Ascii GML file, macro facilities (for defining keyboard equivalents), ad-hoc quick font and size changes, customizability, tables, multiple columns, multiple windows, nondestructive backspace, multiple super- and sub-scripts, horizontal and vertical movement of structural parts of formulas.

(iii) glossary, list of references, drawing tools, read-write text files, file conversion, outliner.

(iv) split windows, headers, footers, footnotes, time stamp, date stamp, auto-page-numbering, binding margins, guttering, mail merge, word count, document history, spelling checker, go-to-page, global font changes, on-line help, switcher compatible, background, decimal tabs, hyphenation, table of contents, index, cursor keys, zoom window, horizontal scrolling, soft hyphen, show/search non-printing characters, ...

A striking fact is that no WP that I know of even comes close to having the essentials for a scientist, viz the features listed under (i) and (ii). Many have a lot of the features listed under (iv) which must be regarded as pure frills as long as the features under (ii) have not been taken care of. As a matter of fact a number of features under (iv) strike me as frills in any case.

It appears that there is something of a gap in the (T)WP offerings; none seems to try to aim specifically at the large scientific market.

References

1.M.Hazewinkel, Experimental mathematics, Math. Modelling **6** (1985), 175-211.

2.F.Topsøe, EUROMATH: the integrated communications system for European mathematicians, Nieuw Archief voor Wiskunde (1986), 155-160.

3.M.Demazure a.o., Report of the EMC subcommittee on normalization (for mathematical electronic communication), Jan. 1986.

4.M.Hazewinkel a.o., Report of the EMC subcommittee on functionality, Jan. 1986.

5.M.Hazewinkel, S.Mullender, Euromath project proposal (intermediate discussion paper of April 1986).

6.S. van Egmond, F.C.Heeman, J.C. van Vliet, INFORM: an interactive syntax-directed formulae editor, Report of the CWI, Amsterdam, 1986.