

International Conference on T_EX,
XML and Digital Typography

TUG 2004 — *T_EX in the Olympics!*

The 25th Annual Meeting and Conference
of the T_EX Users Group

Hosted by the Greek T_EX Friends (*εφτ*)
and the Democritus University of Thrace

Xanthi, Greece
Aug. 30–Sept. 3, 2004

<http://obelix.ee.duth.gr/tug2004>



★ Preliminary Programme ★



The Conference at a glance...

Time	Aug. 30	Aug. 31	Sept. 1	Sept. 2	Sept. 3
9:00 AM	B. Papadopoulos: A welcome on behalf of DUTH	M. Banouni <i>et al.</i> : Dynamic Arabic fonts	W. Bzyl & T. Przechlewski: Migrating to XML	D. Kastrup: The bigfoot bundle	Discussion: The future of T _E X
9:30 AM	B. Beeton: A welcome on behalf of TUG	K. Píška: Type1 from METAFONT	T. Przechlewski: Managing T _E X with XML	J. Likos: Monotonic to polytonic Greek	
10:00 AM	A. Syropoulos: A welcome on behalf of GTF	L.N. Gonçalves: FEATPOST	L. Padovani: Editing of MathML	D. Filippou: Hyphenation of Greek	
10:30 AM	Coffee break				
10:45 AM	C. Loverdos & A. Syropoulos: Digital typography	G. Williams: Fonts – Beyond glyphs	J. Holeček & P. Sojka: Animations in pdfT _E X	P. Rosell-González: The mayan package	Concluding remarks
11:15 AM			M. Eddabini <i>et al.</i> : Arabic math	M. Athale & R. Athale: A Marathi-English dictionary	
12:00 AM	Swim, lunch and nap break				
6:00 PM	Y. Haralambous & G. Bella: OpenType and Ω	J.-M. Hufflen: MIBibT _E X	Tutorial: Advanced L ^A T _E X	J. Arias <i>et al.</i> : Basque: A case study	
6:30 PM	J. Plaice & P. Swoboda: Moving Ω to C++	J. Laurens: iT _E XMac		P. Szabó: Hungarian babel	
7:00 PM	J.-H. Cho & H. Okumura: CJK languages with Ω	F.-R. Shaefer: SaferT _E X	Tutorial: Fonts with FontForge	K.R. Beesley: The Deseret alphabet	

Detailed Conference Schedule

All presentations, tutorials and round-table discussions will take place in the main auditorium of the New Kimeria Campus of the Democritus University of Thrace, 1.5 km East of the town of Xanthi. The lectures and the presentations will begin every day at 9:00 AM.

A swim, lunch and nap break will allow the conference attendees to relax from noon until 6:00 PM. For those who will not spend their afternoon at nearby beaches, a late lunch will be served at the University Cafeteria at 2:15 PM. The evening sessions will last from 6:00 PM until about 8:00 PM. A late dinner will be served every evening at the University Cafeteria at 8:30 PM.

The Conference banquet will be held on Wednesday, Sept. 1, at 9:00 PM (more details will be given later). A round-table discussion on the future of \TeX has been planned for the morning of Friday, Sept. 3, just before the Conference closure.

Day 1

Monday, August 30, 2004

Opening Session

9:00 AM

A welcome message on behalf of the Democritus University of Thrace

Basil P. Papadopoulos
Democritus University of Thrace, Xanthi, Greece

9:30 AM

A welcome message on behalf of the \TeX Users Group

Barbara Beeton
 \TeX Users Group, Portland, OR, USA

10:00 AM

A welcome message on behalf of the Greek \TeX Friends

Apostolos Syropoulos
Greek \TeX Friends, Xanthi, Greece

(Coffee break: 10:30 AM)

Plenary lecture**10:45 AM****Digital Typography in the New Millennium: Flexible Documents by a Flexible Engine**Christos KK Loverdos⁽¹⁾ and Apostolos Syropoulos⁽²⁾⁽¹⁾University of Athens, Greece; ⁽²⁾Greek T_EX Friends Group, Xanthi, Greece

The T_EX family of electronic typesetters is the primary typesetting tools for the preparation of demanding documents, and have been in use for many years. However, our era is characterized, among others, by Unicode, XML and the introduction of interactive documents. In addition, the Open Source movement, which is breaking new ground in the areas of project support and development, enables masses of programmers to work simultaneously. As a direct consequence, it is reasonable to demand the incorporation of certain facilities to a highly modular implementation of a T_EX-like system. Facilities such as the ability to extend the engine using common scripting languages (e.g., Perl, Python, Ruby, etc.) will help in reaching a greater level of overall architectural modularity. Obviously, in order to achieve such a goal, it is mandatory to attract a greater programming audience and leverage the Open Source programming community. We argue that the successful T_EX-successor should be built around a microkernel/exokernel architecture. Thus, services such as client-side scripting, font selection and use, output routines and the design and implementation of formats can be programmed as extension modules. In order to leverage the huge amount of existing code, and keep document source compatibility, the existing programming interface is demonstrated to be just another service/module.

(Lunch break: 12:00 AM)

* * *

Session 1 – Omega**6:00 PM****OpenType and Ω : Past, Present and Future**

Yannis Haralambous and Gábor Bella

École Nationale Supérieure des Télécommunications de Bretagne, Brest, France

(Abstract not available.)

6:30 PM**Moving Ω to a C++-based Platform**

John Plaice and Paul Swoboda

The University of New South Wales, Sydney, Australia

(Abstract not available.)

7:00 PM**Typesetting CJK Languages with Ω** Jin-Hwan Cho⁽¹⁾ and Haruhiko Okumura⁽²⁾⁽¹⁾Korea Institute for Advanced Study, Korea; ⁽²⁾Matsusaka University, Japan

This paper describes how to typeset Chinese, Japanese, and Korean (CJK) languages with Omega, a 16-bit extension of Donald Knuth's T_EX. In principle, Omega has no difficulty in typesetting those East Asian languages because of its internal representation using 16-bit Unicode. However, it has not been widely used in practice because of the difficulties in adapting it to CJK typesetting rules and fonts, which we will discuss in the paper.

* * *

Day 2

Tuesday, August 31, 2004

Session 2 – Fonts and Graphics**9:00 AM****Dynamic Arabic Mathematical Fonts**Mostafa Banouni, Mohamed Elyaakoubi and Azzeddine Lazrek
University Cadi Ayyad, Marrakech, Morocco

This paper describes a font family designed to meet the requirements of typesetting mathematical documents in an Arabic *presentation*. Thus, not only is the text written in an Arabic alphabet-based script, but specific symbols are used and mathematical expressions also spread out from right to left. Actually, this font family consists of two components: an Arabic mathematical font and a dynamic font. The construction of this font family is a first step of a project aiming at providing a complete and homogeneous Arabic font family, in the OpenType format, respecting Arabic calligraphy rules.

9:30 PM**Creating Type 1 Fonts from METAFONT Sources: Comparison of Tools, Techniques and Results**

Karel Píška

Institute of Physics, Academy of Sciences, Prague, Czech Republic

This paper summarizes experiences in converting METAFONT fonts to PostScript fonts with T_EXtrace and mftrace, based on programs of autotracing bitmaps (AutoTrace and po-trace), and with systems using analytic conversion (MetaFog and MetaType1, using META- Ω T output or META Ω T itself). A development process is demonstrated with public Indic fonts (Devanagari, Malayalam). Examples from the Computer Modern fonts have been also included to illustrate common problems of conversion. Features, advantages and disadvantages of various techniques are discussed. Postprocessing•corrections, optimization and

(auto)hinting or even preprocessing may be necessary, before even a primary contour approximation is achieved. To do fully automatic conversion of a perfect METAFONT glyph definition into perfect Type 1 outline curves is very difficult at best, perhaps impossible.

10:00 AM

FEATPOST and a Review of 3D METAPOST Packages

L. N. Gonçalves

Universidade Nova de Lisboa, Portugal

META O T is able to produce figures that look almost like ray-traced raster images but that remain vector-based. A small review of three-dimensional perspective implementations with META O T is presented. Special emphasis is given to the abilities of the author's implementation: FEAT O T.

(Coffee break: 10:30 AM)

Invited lecture

10:45 AM

Beyond Glyphs, Advanced Typographic Features of Fonts

George Williams

Programmer and creator of FontForge (formerly PfaEdit), Santa Barbara, CA, USA

The author compares the two common standards for describing advanced typographic features, namely GX and OpenType. He discusses when it is possible to convert from one format to the other. And finally he presents some simple examples of creating these features with FontForge.

(Lunch break: 12:00 AM)

* * *

Session 3 – Software and Tools

6:00 PM

MIBibT_EX: Beyond L^AT_EX

Jean-Michel Huffle

University of Franche-Comté, Besançon, France

This article sums up our experience with MIBibT_EX, our multilingual implementation of BibT_EX, and points out some possible improvements for better co-operation between L^AT_EX and MIBibT_EX. Also, MIBibT_EX may be used to generate bibliographies written according to other formalisms, especially formalisms related to XML, and we give some ways to ease that.

6:30 PM**iTeXMac: An Integrated T_EX Environment for MacOS X**

Jérôme Laurens

Université de Bourgogne, France

The article will be devoted to the presentation of iTeXMac (<http://itexmac.sourceforge.net>) an integrated freeware to produce high quality documents using TeX. Some part of the discussion concerning the T_EX data and meta data storage can be of interest to the whole *NIX audience.

7:00 PM**SaferT_EX— Optimized Code Appearance for Automated Typesetters**

Frank-Rene Schaefer

Cologne, Germany

While T_EX as a typesetting engine provides high quality type setting features, its usability suffers due to its macro-like command language. Many tools were developed over the years simplifying the T_EX interface, such as ConT_EXt, pdfT_EX, L^AT_EX, and NTS. The system introduced in this paper, however, is radical in its way to target optimized code appearance.

The primary goal of SaferT_EX is to make typesetting code as close as possible to human readable text, as we are accustomed to over the last few centuries. Using indentation, empty lines and a few triggers allows one to express interruption, scope, listed items, etc. A minimalized frame of ‘axioms’ shall span a space of possible typesetting commands. Letters such as ‘_’ and ‘\$’ do not have to be backslashed any more. Transitions from one type of text to another are automatically detected, to the effect that environments do not have to be bracketed explicitly.

The following paper introduces the programming language SaferT_EX as user interface to the T_EX typesetting engine. It is shown how the development of a redundancy reduced language at the same time increases the beauty of code appearance.

* * *

Day 3

Wednesday, September 1, 2004

Session 4 – Electronic Documents**9:00 AM****Migrating to XML: The Case of the *GUST Bulletin* Archive**

Włodzimierz Bzyl and Tomasz Przechlewski

University of Gdańsk, Poland

The dominant role played by the Web in information exchange in modern times has motivated publishers to make printed documents widely available on the Internet. It is now

common that many publications are available on the Web only, or before they are printed on paper. Articles published in the GUST bulletin are available on the Web in PostScript and PDF. Unfortunately, these formats decrease document accessibility, searching and indexing by Web search engines. For broad accessibility to automated services, it is better to use XML as the format of such data. However, one issue with XML is that it is difficult to maintain the high quality presentation of \TeX documents. This is caused by incompatibilities between browsers and incomplete or immature implementations of W3C Consortium standards.

We are optimistic that these issues will disappear in the near future, and believe that XML will become pervasive in the online environment. However, in our context, a key to the adoption of XML is the degree to which it can be integrated with existing \TeX technologies.

In this paper we examine one strategy for storing GUST articles in a custom XML format and publishing them with both \TeX and XSLT/FO. Also, the problems of converting the existing \TeX files to XML and the possibility of using \TeX 4ht—an authoring system for producing hypertext—are discussed.

9:30 AM

Managing \TeX Resources with XML Topic Maps

Tomasz Przechlewski

Katedra Informatyki Ekonomicznej, Sopot, Poland

For many years the Polish \TeX Users Group newsletter has been published online on the GUST web site. The repository now contains valuable information on \TeX , METAFONT, electronic documents, computer graphics and related subjects. However, access to the content is very poor: it is available as PS/PDF files with only a simple HTML page facilitating navigation. There is no integration with information resources from other sections of the site, nor with the resources from other LUG or CTAN sites.

Topic maps were initially developed for efficient preparation of indices, glossaries and thesauruses for electronic documents repositories, and are now codified as both the ISO standard (ISO/IEC 13250) and the XTM 1.0 standard. Their applications extend to the domain of electronic publishing. Topic maps and the similar RDF standard are considered to be the backbone of corporate knowledge management systems and/or the Semantic Web.

The paper contains an introduction to the Topic Maps standard and discusses selected problems of Topic Map construction. Finally the application of Topic Maps as an interface to the repository of \TeX related resources is presented, as well as the successes and challenges encountered in the implementation.

10:00 AM

Interactive Editing of MathML Markup Using \TeX Syntax

Luca Padovani

University of Bologna, Italy

We describe the architecture of a syntax-directed editor for authoring structured mathematical documents which can be used for the generation of MathML markup. The author interacts with the editor by typing \TeX markup as he would do in a normal text editor, with the difference that the typed markup is parsed and displayed on-the-fly. We discuss issues regarding both the parsing and presentation phases and we propose implementations for them. In contrast with existing similar tools, the architecture we propose offers better compatibility with \TeX syntax, a pervasive use of standard technologies and a clearer separation of content and presentation aspects of the information.

(Coffee break: 10:30 AM)

10:45 AM**Animations in pdfTeX-generated PDF: A New Method for Directly Embedding Animation into PDF**

Jan Holeček and Petr Sojka

Masaryk University, Brno, Czech Republic

This paper presents a new approach for creating animations in Portable Document Format (PDF). The method of animation authoring described uses free software (pdfTeX) only. The animations are viewable by any viewer that supports at least some features of Acrobat JavaScript, particularly Adobe (Acrobat) Reader, which is available at no cost for a wide variety of platforms. Furthermore, the capabilities of PDF make it possible to have a single file with animations both for interactive viewing and printing.

The paper explains the principles of PDF, Acrobat JavaScript and pdfTeX needed to create animations for Adobe Reader using no other software except pdfTeX. We present a step by step explanation of animation preparation, together with sample code, using a literate programming style. Finally, we discuss other possibilities of embedding animations into documents using open standards (SVG) and free tools, and conclude with their strengths and weaknesses with respect to the method presented.

11:15 AM**Arabic Mathematical e-Documents**

Mustapha Eddahibi, Azzeddine Lazrek and Khalid Sami

University Cadi Ayyad, Marrakech, Morocco

What problems do e-documents with mathematical expressions in an Arabic presentation present? In addition to the known difficulties of handling mathematical expressions based on Latin script on the web, Arabic mathematical expressions flow from *right to left* and use *specific symbols* with a *dynamic cursivity*. How might we extend the capabilities of tools such as MathML in order to structure Arabic mathematical e-documents? Those are the questions this paper will deal with. It gives a brief description of some steps toward an extension of MathML to mathematics in Arabic exposition. In order to evaluate it, this extension has been implemented in Mozilla.

(Lunch break: 12:00 AM)

* * *

Tutorials**6:00 PM****Advanced L^AT_EX**

Apostolos Syropoulos

Greek T_EX Friends Group, Xanthi, Greece

(Abstract not available.)

7:00 PM**Fonts with FontForge**

George Williams

Programmer and creator of FontForge (formerly PfaEdit), Santa Barbara, CA, USA

(Abstract not available.)

* * *

Day 4

Thursday, September 2, 2004

Session 5 – Critical Editions and Philology**9:00 AM****The bigfoot Bundle**

David Kastrup

Ruhr-Universität Bochum, Germany

(Abstract not available.)

9:30 AM **$\frac{\mu\omicron\nu\omicron}{\overline{\overline{\overline{\pi\omicron\lambda\nu}}}}$ (mono2poly): Java-based Conversion of Monotonic to Polytonic Greek**

Johannis P. Likos

ICT consultant

This paper presents a successfully tested method for the automatic conversion of monotonic modern Greek texts into polytonic texts, applicable on any platform. The method consists of combining various freely available technologies, which have much better results than current commercially available solutions. The aim of this presentation is to introduce a way of applying this method, in order to convert thousands of digitally available single-accented modern Greek pages into attractive artworks with multi-accented contents, which can be easily transferred either to the Web or a T_EX-friendly printer. We will discuss the preparatory and postprocessing efforts, as well as the editing of syntax rulesets, which determine the quality of the results. These rulesets are embedded in extendable tables, functioning as flat databases.

10:00 AM**Hyphenation Patterns for Ancient and Modern Greek**

Dimitrios Filippou

Volos, Greece

Several files with Greek hyphenation patterns for T_EX can be found on CTAN. However, most of these patterns are for use with Modern Greek texts only. Some of these patterns contain mistakes or are incomplete. Other patterns are suitable only for some now-outdated “Greek T_EX” packages. In 2000, after having examined the patterns that existed already, the author made new sets of hyphenation patterns for typesetting Ancient and Modern Greek texts with the greek option of the `babel` package or with Dryllerakis’ `GreekTEX` package. Lately, these patterns have found their way even into the `ibycus` package, which can be used with the *Thesaurus Linguae Graecae*, and into Ω with the `antomega` package.

The new hyphenation patterns, while not exhaustive, do respect the grammatical and phonetic rules of three distinct Greek writing systems. In general, all Greek words are hyphenated after a vowel and before a consonant. However, for typesetting Ancient Greek texts, the hyphenation patterns follow the rules established in 1939 by the Academy of Athens, which allow for breaking up compound words between the last consonant of the first constituent word and the first letter of the second constituent word, provided that the first constituent word has not been changed by elision. For typesetting polytonic (multi-accent) Modern Greek texts, the hyphenation rules distinguish between the nasal and the non-nasal double consonants $\mu\pi$, $\nu\tau$, and $\gamma\kappa$. In accordance with the latest Greek grammar rules, in monotonic (uni-accent) Modern Greek texts, these double consonants are not split.

(Coffee break: 10:30 AM)**10:45 AM****mayan: Fonts and Package for Typesetting Mayan Numbers**

Pablo Rosell-González

Universidad Nacional Autónoma de México, Mexico, D.F., Mexico

The Mayans developed a complete positional numeral system. This fact facilitates calculation, particularly adding and subtracting. Their number system was a vigesimal (base 20) system, consisting of three different symbols, a dot which represents ‘one’, a horizontal line which represents ‘five’, and an stylized cacao seed representing ‘zero’. Any number between 0 and 19 can be represented with these symbols, for example, 12 is represented as two horizontal lines, one above the other, and two dots on top of them.

In this paper I will present the `mayan.sty` package as well as `mayan*.mf` METAFONT fonts and their PostScript Type1 variants.

The `mayan` package allows the user to generate mayan numbers by simply invoking, e.g., `\mayan{num}` which in any display math environment will typeset the number in a vertical stack using large symbols while in paragraph mode will separate each level with a diagonal using adequate size symbols depending on the text font size.

Optional arguments and other commands are available for boxing the numbers when stacked vertically, writing the numbers in “vertical style” inside paragraphs, and writing the Arabic values aside each position.

11:15 AM**Using L^AT_EX to Typeset a Marāṭhī-English dictionary**

Manasi Athale and Rahul Athale

Johannes Kepler University, Linz, Austria

We are using L^AT_EX to typeset an old Marāṭhī-English dictionary, dated 1857. Marāṭhī is the official language of Mahārāshtra, a western state of India. Marāṭhī is written using the Devanāgarī script. The printed edition of the dictionary contains approximately 1000 Royal Quarto size (9 1/2" × 12 2/3") pages with around 60,000 words. The roots of the words come from many languages including Sanskrit, Arabic and Persian. Therefore the original dictionary contains at least three different scripts along with many esoteric punctuation marks and symbols that are not used nowadays.

We have finished typesetting 100 pages of the original dictionary. We present our experiences in typesetting this long work involving Devanāgarī and Roman script. For typesetting in Devanāgarī script we used the `devnag` package. We have not yet added the roots in other scripts but that extension can be achieved with the help of ArabT_EX. We want to publish the dictionary in electronic format, so we generated output in pdf format using pdfL^AT_EX. The bookmarks and cross-references make navigation easy. In the future it would be possible to design the old punctuation marks and symbols with the help of METAFONT.

(Lunch break: 12:00 AM)**6:00 PM****Basque: A Case Study in Generalizing L^AT_EX Language Support**

Jagoba Arias, Jesús Lázaro and Juan M. Aguirregabiria

Universidad del País Vasco, Bilbao, Spain

The multilingual support of L^AT_EX presents many weak points, especially when a language does not present the same overall syntactic scheme as English. Basque is one of the official languages in the Basque Country, being spoken by almost 650,000 speakers (it is also spoken in Navarre and the south of France). The origins of the Basque language are unknown. It is not related to any neighboring language, nor to other Indo-European languages (such as Latin or German). Thus, dates, references and numbering do not follow the typical English pattern. For example, the numbering of figure prefixes does not correspond to the `\figurename\thefigure` structure, but is exactly the other way round. To make matters worse, the presence of declension can turn this usually simple task into a nightmare. This article proposes an alternative structure for the basic classes, in order to support multilingual documents in a natural way, even in those cases where the languages do not follow the typical English-like overall structure.

6:30 PM**Implementation tricks in the Hungarian babel module**

Péter Szabó

Budapest University of Technology and Economics, Hungary

`magyar.ldf`, the Hungarian Babel module, was rewritten in the autumn of 2003 to obey most of the Hungarian typographical rules. This article describes some implementation is-

sues, TEX macro programming hacks, and L^AT_EX typesetting trickery used in `magyar.ldf`. All features of the new `magyar.ldf` are enumerated, but only those having an interesting implementation are presented in detail. Most of the tricks shown are useful for developing other language modules.

7:00 PM**Typesetting the Deseret Alphabet with L^AT_EX and METAFONT**

Kenneth R. Beesley

Xerox Research Centre Europe, Meylan, France

The Deseret Alphabet was an orthographical reform for English, promoted by the Church of Jesus Christ of Latter-day Saints (the Mormons) between about 1854 and 1875. An offshoot of the Pitman phonotypy reforms, the Deseret Alphabet is remembered mainly for its use of non-Roman glyphs. Though ultimately rejected, the Deseret Alphabet was used in four printed books, numerous newspaper articles, several unprinted book manuscripts, journals, meeting minutes, letters and even a gold coin, a tombstone and an early English-to-Hopi vocabulary. This paper reviews the history of the Deseret Alphabet, its Unicode implementation, fonts both metal and digital, and projects involving the typesetting of Deseret Alphabet texts.

* * *

Day 5

Friday, September 3, 2004

Closing Session**9:00 AM****The future of T_EX**

Round-table Discussion

(Description not available.)

(Coffee break: 10:30 PM)**10:45 AM****Closing Remarks**

Apostolos Syropoulos

Greek T_EX Friends Group, Xanthi, Greece

* * *