Back to the roots: TUG 2019 in Palo Alto

Henri Menke

Thursday, August 8

This year's TUG meeting stood under a very special sign. TEX returned to its birthplace, Stanford University in Palo Alto, California, USA. And not only did it return to its cradle, but also its inventor would be there, none other than the great Donald E. Knuth (DEK) himself. It was a great honor to have such a very special guest at the meeting.

The evening before the meeting, there was a reception and registration at the conference site, the Sheraton Hotel. Many participants showed up and DEK was also there to personally welcome everyone.

Friday, August 9

On the morning of the first day our current president, Boris Veytsman, officially opened the conference. The first speaker to kick off the program was Erik Braun from the CTAN team to talk about goals and difficulties CTAN is facing. A notable quote from the talk is that "the Comprehensive TEX Archive Network is neither comprehensive nor an archive". However, the team wants to change that in the future together with some new web design.

Next up Arthur Reutenauer, a current maintainer of XƎTEX, enlightened us about the past, present, and future state of XƎTEX. As it currently seems, XƎTEX might go into maintenance mode, because the original author Jonathan Kew has moved on to other projects and major contributor Khaled Hosny, who ported XƎTEX from AAT to HarfBuzz, has focused his attention on LuaTEX, leaving no competent developer. With recent progress in LuaTEX and its likely adoption of HarfBuzz, the main use cases for XƎTEX will be covered, making it kind of obsolete. Right now, Arthur tries to isolate XƎTEX-specific code from the engine, to maybe contribute those patches to LuaTEX.

Before the morning break, Frank Mittelbach introduced us to the LATEX "dev" format. The "dev" format is a preview of the next official release of the LATEX format and is shipped with TEX Live for interested users to try out. The LATEX team hopes to attract more beta testers this way, so that problems, especially regressions, can be detected before official releases, reducing the number of required hotfixes.

After the break, Frank Mittelbach continued with a second talk about UTF-8 in LATEX. A few years ago the "Unicode revolution" was completed and almost everything is formatted in UTF-8 nowadays. Unfortunately, the old 8-bit TEX engines do not support UTF-8 natively, so the well-known `inputenc` was introduced, which knows about the "magic" multibyte sequences of UTF-8 and does the right thing by converting the characters into the LATEX internal character representation (LICR). This technique has been integrated into the LATEX format last year and is available to all users.

The next talk was presented by Uwe Ziegenhagen, a LATEX and Python enthusiast from Cologne. After a short introduction to the Python programming language, he presented two ways of combining LATEX and Python: First, how to generate LATEX in Python using a template engine, and second, how to execute Python from within LATEX.

My own talk was the last before lunch, and the audience had a lot to digest, despite their empty stomachs. I talked about how to parse complex data formats in LuaTEX using the integrated LPEG library. The LPEG library provides a domain-specific embedded language for parsing expression grammars (PEG) within Lua using operator overloading. After an introduction to PEG, I demonstrated how to construct a JSON parser, which can be used, for example, to read document metadata from a configuration file.

After lunch, Dick Koch, the principal maintainer of MacTEX, told us about how Apple is continuing the war against its developers. If a macOS application is not digitally signed by its developer, the system will display a warning that the application is not trusted, which is a good thing. However, with the upcoming macOS Catalina, Apple will allow only "notarized" applications to be run, which requires the developer to apply to Apple directly for notarization. Dick outlined which changes were necessary in MacTEX to pass the notarization test, so macOS TEX users will not have to worry.

This was followed by a talk by Nate Stemen, a software developer at Overleaf, about their product. Overleaf is an online LATEX editor which allows multiple authors to edit the same document at the same time. It is useful for beginners as well, because it saves the user the installation of a full TEX distribution on their own machine and it comes with a large pool of example documents. With over four million users world-wide, Overleaf is an expanding business. The company is also very interested in symbiosis with TUG and the TEX community.

The next speaker was Rishi T from STM DOCS in India, presenting Neptune, which is part of STM's proofing framework TEXFolio (`neptune.texfolio.org`). Neptune allows journal editors to send proofs to authors with specific queries for them to address. The talk was mostly an interactive demonstration.

Before the afternoon break, Pavneet Arora spoke about distraction-free writing with Vim and how he leveraged those features to finish his latest novel.

The last session of the day started with a talk about Knuth's book *TeX: The Program.* Because TeX is a literate program, the source code and the documentation can be generated from the same file and it should be possible to read the code like a book. However, the structure of the TeX program is bottom-up, i.e., numerous low-level details are discussed before moving to the big picture, which makes the book hard to read. The speaker, Shreevatsa R, has collected resources to aid with reading at `shreevatsa.net/tex`.

Next up, Petr Sojka reviewed the history and current usage of TeX at his home institution, the Masaryk University in the Czech Republic where the pdfTeX engine was born.

This was followed by Shakthi Kannan, who introduced his free software framework "The XƎTEX Book Template" to publish multilingual books using XƎTEX, based on Emacs Org-mode and TeX. He presented prominent features and shared experience in creating and publishing books using the framework.

The last speaker of the day was Jim Hefferon, who spoke about his experience helping new LaTeX users (or as he called them "the great unwashed") in the `r/latex` forum on Reddit. He presented a survey of which questions are frequently asked, and pointed out that new users tend to frequent social sites on the web much more than the TUG site.

After the sessions had ended it was time for the TUG Annual General Meeting. The main theme was how to raise money for TUG by interacting first and foremost with the institutional members. Notes from Jennifer Claudio follow this report.

Saturday, August 10

On the second day of the conference we were visited by our special guest DEK, who attended all the sessions and the banquet afterwards.

The first session of the second day was started off by Petr Sojka and his son Ondřej, who spoke about their recent effort in generating better hyphenation patterns for the Czech language.

After that Arthur Reutenauer gave another, more historical talk where he reviewed the history of hyphenation patterns in TeX. The hyphenation algorithm by Liang and Knuth has since made its way into many other programs, such as OpenOffice. Therefore it seemed appropriate to unify the mess of hyphenation patterns and make them publicly available outside of CTAN. The project is ongoing and very successful (`hyphenation.org`).

The next talk was an earthshaking announcement by DEK's former PhD student David Fuchs. He has taken TeX and METAFONT and combined them into a single program with graphical output of the generated DVI. When TeX tries to load fonts, these are created ad-hoc by METAFONT and cached. At each page, the state of the TeX engine is recorded and the difference to the previous state is cached. When then editing the text, TeX is able to preview the changes in real time, even for large documents such as *The TeXbook*.

After the morning break we heard the presentation by Tom Rokicki, maintainer of the `dvips` program. Even though `dvips` has been mostly stable for a long time, Tom was unsatisfied with the fact that when bitmapped Type 3 fonts were used, it was not possible to copy and paste text from the output PDF. To this end he implemented a new font encoding routine, which reads optional encoding data to map the font correctly. This change will be available in an upcoming version of `dvips`.

In the next presentation, Martin Ruckert introduced his newest creation, the HINT file format. The HINT file format, produced using the HiTeX program, is very similar to the output of `\showlists`, as it captures most of the important information. This information can then be used by a viewer to generate screen output which is very similar (or even equivalent) to TeX, without implementing all of the algorithms. This allows reflowing the text, resulting in a great application for eBook readers, which so far have notoriously bad typesetting quality.

The last talk before the lunch break was given by Doug McKenna about his implementation of a TeX interpreter as a library, JSBox. This library is bundled with iOS applications to typeset interactive eBooks ad-hoc. This was demonstrated using his latest book *Hilbert Curves*.

Before the program resumed after lunch it was time for the group photo. After that Jennifer Claudio entertained us with different shapes of the letter 'E' that she had seen on her way to the conference.

This was followed by a talk by Federico Garcia-De Castro, a professional composer and typesetting enthusiast, who has designed an algorithm for typesetting so-called slurs in sheet music with META-FONT. It was amazing to see how much attention he spent to detail and how superior his approach is compared to commercial scoring software.

Afterwards William Adams presented a fun little project in which he manufactured a small wooden box with a CNC machine at home. The blueprints for this box were prepared with TeX.

Henri Menke

The last session of the day was opened by Boris Veytsman, who talked about another method to prepare commented editions of a text. In this case the target was a mathematics textbook in which the teacher's version contains extra comments all around the page. This was achieved by means of his new package `commedit`.

The next presentation was a history lesson by Behrooz Parhami about the evolution of Persian and Arabic scripts from the early days of handwritten script, over the introduction of movable type, and eventually typewriters, to computer typesetting and the problems with bitmapped fonts due to small features in the glyphs.

This talk set the stage perfectly for the next, by Amine Anane, who introduced his software "Visual METAFONT", which can trace the outlines of scanned glyphs and turn them into a variable font. The newly introduced extensibility of glyphs should eventually be respected by TEX when breaking lines to offer superior typesetting for Arabic script.

The final talk of the day was presented by Takuto Asakura. The use of mathematical markup is very heterogeneous and does not necessary reflect the corresponding semantics in scientific documents. To this end, the speaker designed a synthetic analysis which harnesses the written descriptions of formulae in natural language to assign meaning to the markup.

After the official program, Herbert Schulz ran a workshop on the macOS (LA)TEX editor TeXShop.

In the evening of that day a lovely and delicious banquet took place at the Sheraton hotel. We were honored by the presence of our special guest DEK. Dinner was followed by a raffle of TAOCP set of physical books and two e-book vouchers, all donated by Pearson, as well as two TEX lion plushies donated by Jill Knuth, and a framed original black and white conference drawing. For her efforts on the local organizing committee, Boris presented Jennifer Claudio with this year's Duane Bibby signed original color conference drawing. In addition, Cheryl Ponchin and Sue DeMeritt's long service to TEX and TEX Users Group was recognized with a personalized gift certificate, as they retired from the TEX Users Group board this year. Barbara Beeton, a charter member of TEX Users Group and the TEX Users Group board, among many other TEX and TEX Users Group activities, was also recognized, with the first lifetime membership to TEX Users Group, on the occasion of her retirement from the AMS. Barbara was also given a personalized gift certificate and other group memorabilia. Finally, Don Knuth was given unusual books of organ music as a small token of appreciation, on behalf of the entire TEX community.

Sunday, August 11

The last day of the conference was begun by Antoine Bossard, who teaches at Kanagawa University in Japan. There he is confronted with typesetting mixed CJK and Latin content, so he presented his minimal approach for TEX macros to facilitate this.

The next talk was delivered by Jaeyoung Choi, who in collaboration with others designed a module for the FreeType library to render METAFONT-generated and TEX-oriented bitmap fonts. FreeType is used on many platforms including Android and Apple operating systems.

In the last talk of the first morning session, Jennifer Claudio reported on a project she undertook with her student Emily Park. They studied whether machine learning techniques could be used to detect transliterated English words in Korean text with the Hangul alphabet. As of now it seems that measures such as grayness are not sufficient to distinguish.

The second sessions started with Rishi T from STM DOCS in India, with the second presentation on STM's proofing framework TEXFolio, which is a complete journal production system that supports LATEX and XML input and HTML5 and ePub output.

Then we moved on to the next talk by Boris Veytsman. He had applied machine learning techniques to BIBTEX datasets. Starting from an annotated set of BIBTEX records he had collected from online sources, a neural network was trained to identify author, title, journal, etc., from the generated output. So far the results are mixed because citation styles vary in a rather inconvenient fashion. Most tend to abbreviate authors with initials, and physics journals often omit titles.

Before lunch we heard about another TEX format that does not come up very often but is nevertheless very important — Texinfo. It is a format for software documentation that can produce a number of different outputs including HTML and PDF. The speaker Didier Verna is applying Texinfo to the Common Lisp ecosystem to automatically generate documentation for all of the available libraries (numbered in the thousands): `quickref.common-lisp.net`.

After lunch Uwe Ziegenhagen presented his second talk, this time on creating and automating exams with LATEX using the `exam` package. Again using Python, he created different versions of the same exam to make it harder for students to copy.

This was followed by another talk on exams, by Yusuke Terada, who wants to optimize marking of the Japanese national exam called the "center test". To this end he created machine-readable exam sheets using TEX and matching software which presents the

extracted answers to an examiner in an anonymized fashion to remove bias. The marks are collected electronically and reunited with the personalized information to generate an evaluation sheet. So far this system has only been implemented at Yusuke Terada's school, but should eventually become the national standard.

The conference concluded with two talks on accessibility, the first of which was delivered by Chris Rowley of the LaTeX3 team. He reported on the current state of accessibility in LaTeX and introduced the `tagpdf` package by Ulrike Fischer which aids in tagging LaTeX-generated PDFs with the proper structural elements. There are still a lot of open problems, especially concerning mathematics.

After that Ross Moore, who joined via video from Australia, demonstrated that accessibility is already possible in LaTeX if one is aware of certain difficulties, using the example of a research report he prepared for the U.S. National Park Service.

Conclusion

In summary, the TUG 2019 meeting in Palo Alto was great. Many topics were touched on and it was amazing to see which recent developments are taking place. There were a lot of lively discussions, especially with participants from the big Silicon Valley companies and it was a great honor to meet DEK. Next year's TUG 2020 will take place at the Rochester Institute of Technology in Rochester, New York.

Acknowledgment

I'd like to thank the TUG bursary for funding, which supported me in attending this conference.

⬦ Henri Menke
  Dunedin, New Zealand
  henrimenke (at) gmail dot com