
TUG 2024 in Prague

Mitchell Gerrard

The entire conference was streamed at no charge on YouTube. Each day’s complete stream, and eventually each talk as an individual video, can be accessed via youtube.com/@TeXUsersGroup.

Thursday, July 18

TUG 2024 was in Prague, Czechia, home of beautiful architecture, beloved authors, and many brilliant \TeX ncians. The conference began with a reception at the event’s venue, the Hotel Grandior.

Friday, July 19

Boris Veytsman opened the conference, welcoming the expectant attendees; then: we were on our way.

Norbert Preining gave a prerecorded video talk about \TeX on arXiv. Operating since 1991 from Cornell, arXiv is the oldest and largest server hosting preprints of largely scientific publications, with no paywall. Norbert talked about supporting old software, including multiple versions of \TeX , and gave an overview of the new submission software, which is being rewritten in a combination of Python and Docker containers, the goal being easy and painless submission to arXiv.

Next, I rooted around through the historical precursors of literate programming. An early WEB-like paradigm called “Holon Programming” was described in a hard-to-find technical report by Pierre-Arnoul de Marneffe, but it is now typeset and freely available online.

Vincent Goulet presented on a literate programming workflow he developed for the R ecosystem, in the context of teaching courses related to actuarial science. A literate setup allows Vincent to keep teaching materials for himself, students and TAs in one file, which includes: lecture notes, questions and solutions, and unit tests. He described a circular dependency issue he ran into in this setup, and provided a clever way out. (During the talk a tiny spider appeared in front of me, suspended from the high conference ceiling, in search of weaving some web.)

Martin Ruckert described a profiler he wrote for \TeX . A profiler is a tool to tell you where software spends the majority of its time when running on some input, so a developer can know where to apply optimizations. Martin said we should never optimize for speed without a profiler. His \TeX profiler is aimed for use by macro \TeX ncians whose macros will be used frequently by many people; the document author doesn’t need a profiler, unless curiosity counts as a necessity. We looked at how to optimize an input

example from the Book of Numbers, and Martin encouraged us to only apply optimizations if they can be done within a day or so.

Tyge Tiessen rewrote the entirety of \TeX 82 in Rust (yes, it even passed the TRIP test) *and* he lived to tell the tale. Tyge’s initial goal of learning Rust through a medium-sized project expanded a bit when he settled on reimplementing \TeX . He began with a straightforward translation of each module, using Rust’s `unsafe` keyword to access globals, and creating a nice approximation of Pascal’s `goto` statements. Eventually this was refactored to remove all explicit global accesses. Like Knuth’s code, Tyge’s version includes no external dependencies. I say: respect.

Didier Verna presented on the design of a parser within his platform for experimenting with typesetting algorithms, which is written in Common Lisp. He spoke about the need for robustness and flexibility in software design, and confessed to being a “flexibility psychopath”: there are now around a dozen recovery options in his parser. I learned of a (flexible!) error handling system in Common Lisp that improves upon traditional try/catch mechanisms by allowing you to restart computation at arbitrary program locations. Didier’s parser validated nearly 80,000 font metric files, finding only two truly unusable fonts among the 770 flagged as non-compliant.

We took a break for lunch and the group photo, which included many shifts up a big set of steps as we worked to get everybody into the camera’s viewfinder.

Boris Veytsman began by highlighting recent work in decolonization on reclaiming cultural traditions, including typography. Specifically — Ukrainian typography, which has a rich history distinct from Russian letterforms. Boris showed a new Ukrainian font, Arsenal, recently added to \TeX Live. The specimens shown came from Boris’ personal collection of favorite Ukrainian poets and writers. Arsenal was designed to be “business-like”; Boris wondered aloud: is it OK to mix business and poetry?

Vít Starý Novotný brought us behind the scenes of producing documents for the International Software Testing Qualifications Board. This is done by populating a YAML file with easy-to-read-and-write Markdown, and then running this file through a small \LaTeX driver to produce output. The Markdown syntax (with some extensions) has covered most use cases for ISTQB, but \TeX ncians can still jump in to this workflow and tweak away if desired.

Wim Obbels spoke about Ximera, a tool for making interactive courses and textbooks. Using one \LaTeX source file, Ximera can generate PDF and

HTML documents. Links to Desmos can be embedded and there's nice integration with Sage. If teachers want to print out PDFs, e.g., for worksheets, a QR code can bring students to interactive explanations. Thanks to \TeX -in-the-browser (is it magic? no, it's WebAssembly!), students just need a browser to start playing with Ximera's creations.

Sarah Lang talked about \LaTeX use and non-use within the digital humanities. She outlined how this field uses \LaTeX : for conference submissions, archaeology catalogues, and print versions of digital scholarly editions (thank goodness books haven't died out as predicted). Unfortunately many people in digital humanities see \LaTeX as exclusionary or too computer-y, but then end up increasing complications by, e.g., hacking arcane solutions in Word. Sarah mused on ways to reduce this reticence by way of approachable tutorials and blogs.

Oliver Kopp began his talk on a cloud-based pipeline for \LaTeX by asking: why in the world would we want such a thing? It's to avoid the shoulder-shrug-frustrations of "dunno, it worked on *my* machine". Oliver stepped through creating a basic package in this pipeline, and on the way featured a package dependency printer, showed how to release to CTAN using GitHub/GitLab and how to use this workflow locally with the help of Docker. The day ended with an interactive tutorial in this pipeline.

Saturday, July 20

Michal Hoftich presented on a tool that can take HTML or ePub input, remove images and ads, and output the pared-down pages to PDF. This tool's responsive design helps pages display nicely on a huge screen or on a tiny phone. The overall feel is similar to "reader mode" in Firefox. The tool can also pre and postprocess HTML, and Michal showed an example of using transformation rules to remove certain HTML elements.

Simon Pfahler opened with an effective demonstration of why disregarding colorblindness can be a problem: his title slide changed color schemes and suddenly the text disappeared! This was a simulation of colorblindness, but problems like this arise frequently, as five percent of people have some form of color vision deficiency. Simon looked at colorblind-safe design, including the most important rule: always provide information in more ways than just color. He urged developers to think about color use when designing defaults, and introduced the `colorblind` package, with color definitions of various colorblind-safe schemes.

Changxu Duan talked about the difficulties involved in getting large language models (LLMs) to

consume scientific papers. Most LLMs made for this purpose are tailored to \LaTeX input, not PDF, which an LLM interprets as one unstructured image. Changxu helps LLMs along with reading PDFs by coming up with clever ways to convert its contents into a mix of "vanilla" \LaTeX and Markdown, allowing LLMs to offer better summaries, related paper recommendations, etc.

Joseph Wright presented on work that's been 24 years in the making: "templates" in \LaTeX . The idea is to give the user "instances" of a standard way to implement something, and the user can then tweak a few small parameters. There are a small number of template types (or "things"). The talk was quite interactive, with lots of calls to use some other word than "template". This caused Frank to put his head in his hands; Joseph snapped a picture of this from the podium.

Norbert Preining gave a second video talk on how arXiv makes research accessible. Online scientific work is mostly in PDF form, but if one is, e.g., a blind researcher, this file format is often unhelpful. HTML is a better solution for accessibility, with responsive design, dark mode, built-in language translations, etc. Norbert described the work done at arXiv to convert the \LaTeX sources submitted with most uploads into HTML. There are still some issues, such as missing `TikZ` support, but the response from the community thus far has been overwhelmingly positive.

Ulrike Fischer presented on progress in tagging PDF documents. But as Norbert just said, HTML is generally more accessible, so why bother with PDFs? Well, not every use of PDF can be replaced by HTML. PDFs are easier to handle offline and to archive; it's a faithful representation: you can save and view it on most machines without the hassle of needing to load multiple files, cookies, session IDs, etc. Ulrike showed the status of tagging efforts in various \LaTeX packages, and it's coming along swimmingly well. She opened with a polar bear meme and ended with a cute teddy bear slide.

Frank Mittelbach began with a simple observation: hooks are devices on which you can hang multiple things. There were no hooks in early versions of \LaTeX , then a few were added (including the `AtBeginDocument` hook), and today there's a general hook mechanism. The new hook system reduces the brittle patching that was/is rampant in many packages. Frank then spoke about sockets, in which only *one* thing can be plugged at a time. These are useful for tightly controlled code which can either be turned "on" or "off". Didier joked about renaming sockets and hooks to "socks" and "feet".

Jeffrey Kuan brought us into the realm of U.S. laws relating to accessible documents. How can we support a typical U.S. mathematician to create documents that meet legal requirements? In just a few years, all public U.S. universities and colleges will be required to make all course material “accessible”; this may or may not include PDFs posted to arXiv. But many documents being used in courses today are still not accessible, and most instructors don’t learn about accessibility, so there needs to be a set of solutions to these problems, and quick. These could include standardized courses on L^AT_EX and on accessibility in graduate school.

Vincent Goulet was tasked by The Canadian Journal of Statistics to create a bespoke class in L^AT_EX that had a distinct look and feel from their publisher’s (Wiley) class. So Vincent revamped the old class (from 1994!), replacing crowded headers with breathable ones; stacked English/French headers with side-by-side ones; and Times New Roman/Helvetica fonts with STIX/Fira ones. Many other features make Vincent’s new class a pleasure to use and, upon T_EXing, to read.

Jean-Michel Hufflen presented on making the proceedings for BachoT_EX and *Cahiers GUTenberg*. Issue #57 of *Cahiers* was published in 2012, and after *ten* years, issue #58 came out. So there was a backlog of submitted articles whose compilation required all the varieties of T_EX engines. And for BachoT_EX, certain extensions were necessary, such as tables of contents in both English and Polish. Jean-Michel used a T_EX parser in Scheme to do certain programmatic tasks, and generated the article files based on a makefile. Boris interjected that “People who don’t want to use `make` just reinvent it.”

Rishi T spoke on challenges in manuscript submission from a typesetter’s perspective. His company receives thousands of articles to typeset on a monthly basis. Well-designed class files are a huge help for both authors and typesetters. But certain metadata such as affiliation fields can come in varying formats requiring a lot of manual checking by the typesetter. To streamline the process, Rishi suggested better education for document authors, automated validators/linters, and publisher participation in T_EX conferences.

Andrew Watters joined us in a video talk on using L^AT_EX in the setting of a small law firm. He didn’t want to be trapped by vendor lock-in of using, for example, Microsoft’s suite of tools; so has turned to the land of free software. We got to see inside the workflow of his firm, and with a priest-like hand gesture, we were all blessed with NDAs. Andrew stepped us through his PHP scripts that generate

L^AT_EX documents that help him create essential legal documents such as prebills for clients and pleadings for the courts.

samcarter presented on a new Beamer theme named Moloch, which is a slight variation on Metropolis, one of the most-used and iconic Beamer themes. Metropolis is lovely but now a bit rusty, being last updated on CTAN in 2017. Moloch resolves many of Metropolis’ incompatibilities, cleans up its code, and uses Beamer tooling when possible. But it’s not a 100% replacement of all the Metropolis features, hence the name change. To try Moloch, most users can simply replace `metropolis` with `moloch` in the `usetheme` command.

Rishi T and Rajagopal CV gave an interactive tutorial on T_EXFolio; unfortunately I didn’t bring my laptop so didn’t participate in this.

In the evening we all went to the banquet at Restaurace Tiskárna, some in the upper level and others of us in the comfortable cave environs. The three-course meal was superb.

Sunday, July 21

At this point in the conference I had developed an affection for the cappuccino machine in the lobby, and felt anticipatory regret that I would soon be leaving it.

Ondřej Sojka talked about expanding his previous work on hyphenation patterns in Czech to other Slavic languages. There was a brief overview of Frank Liang’s hyphenation method and the current trend toward hyphenation based on phonetics rather than on etymology. Ondřej’s goal was to improve upon subpar hyphenation patterns in Slavic languages. To do so, he used wikipedia datasets of word lists and combined phonetic hyphenation with their IPA representations.

Antoine Bossard presented on typesetting Latin verse scansion. What is scansion? It’s the identification of metrical feet within verse. “Feet”? That’s the quantity and duration of syllable groupings. Poetic analysis has a standard notation (in the form of diacritical marks) for these concepts, and Antoine showed how to implement this notation in T_EX. He demonstrated this with a passage from Virgil’s *Aeneid*, which he even recited at Martin’s prompting.

Jan Vaněk and Hàn Thé Thành gave a demo and explanation of their Primo tool, a WYSIWYG structural PDF editor. How does one directly edit a PDF? Well, the displayed document is actually built upon XML in the background and immediately produces PDF output in the editor. The tool is collaborative; it’s like a Google Docs but for academic publishing (whose large houses often use XML to

encode their documents). Jan dove into the details of how the collaborative updates work under the hood, using a “TriLayers” abstraction.

Boris Veytsman extended the `bookshelf` package, from Peter Flynn, to handle font selection for multilanguage libraries. Boris happily declared that the original package and his extension have absolutely no practical value, but: *it’s fun*. This project included hacks in `expl3` and `Biber`, and found ways around limitations on the number of fonts you can open at a time. Boris has plans to make the width of a book’s spine vary depending on its actual size.

Joseph Wright gave a development update on the `siunitx` package. His day job is a chemist, so he works with lots of units. And he has a favorite (joule per mole kelvin). Joseph gave an overview of the long history of this package, whose latest additions include complex numbers, a new model for uncertainty, new SI prefixes, finer rounding control, and better alignment in tables. One truism for a package creator is that you can never anticipate all desires from all people, and there *will* be some odd requests.

Before lunch there was the TUG Annual General Meeting, whose minutes will be given elsewhere in this TUGboat issue.

Didier Verna spoke about extensions to the Knuth-Plaus justification algorithm. The original algorithm uses a cost function that applies demerits based on the context of looking at two consecutive lines at a time. Didier introduces a new contextual demerit: considering the beginnings and endings of consecutive lines. The idea is to avoid consecutive occurrences of common short words such as “and”. He implemented the extension in his ETAP tool, and demonstrated its effectiveness using Grimm and Melville texts.

Martin Ruckert began by thanking `samcarter` for helping him update to the Moloch Beamer theme used for this presentation. Martin showed how his HINT tool allows for a bigger design space for programmatic APIs than something like PDF’s fixed format. This was demonstrated by stepping through possible color specifications within `HiTeX`. Colors can be nested within colors and you can specify amounts of transparency for these overlays. Martin ended festively, showing the `TeX` logo with slightly overlapping harlequin-tinted boxes in the background of each character.

Boris brought the session to a close, thanking our team of Czech hosts: Tom Hejda, Michal Hoftich, Ondřej Sojka, and Petr Sojka; and invited us to enjoy the beautiful city of Prague.

The conference ended with an excursion to a pipe organ concert at St. Salvator church. The organist was Lukáš Vendl (with some page-turning help from conference organizer Tom Hejda), and he charmed us with the following musical incantations.

Georg Muffat, Tocatta Septima

Matthias Weckmann, Magnificat Secundi Toni

Johann Gottfried Walther, Giuseppe Torelli’s Concerto in D minor transcribed for the organ

Dieterich Buxtehude, Toccata in D minor

Johann Pachelbel, Chaconne in F minor

Johann Sebastian Bach, Prelude & Fugue in E minor

After the concert we were invited to take a look into the back of the organ where the pipes are housed, and Tom — an organist himself — gave a mini lesson on the mechanics of these marvelous machines.

Acknowledgment

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

◇ Mitchell Gerrard
mitchell dot gerrard (at) gmail
dot com

