

Zpravodaj Československého sdružení uživatelů TeXu

Dave Walden; Hàn Thế Thành

Rozhovor s Hàn Thế Thànhem, tvůrcem a správcem pdfTeXu

Zpravodaj Československého sdružení uživatelů TeXu, Vol. 19 (2009), No. 4, 184–190

Persistent URL: <http://dml.cz/dmlcz/150093>

Terms of use:

© Československé sdružení uživatelů TeXu, 2009

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* <http://dml.cz>

Rozhovor s Hàn Thê Thànhem, tvůrcem a správcem pdfT_EXu

DAVE WALDEN, HÀN THÊ THÀNH

Abstrakt

Tento článek je překladem rozhovoru, který byl poprvé publikován v „koutku rozhovorů“ na stránkách T_EX Users Group, <http://tug.org/interviews/>, 24. července 2008. Tázaný i tazatel dali souhlas k překladu a přetištění.

Klíčová slova: T_EX, PDF, Adobe, pdfT_EX, T_EX2PDF, Masarykova univerzita, mikrotypografie, *hz*-program, T_EXbook, Web, Web2C, Kpathsea, programovací jazyk C, plain T_EX, L^AT_EX, TUGboat, t_EX Live, Prolog, TUG, vietnamština, písmo, Computer Modern, METAFONT, Type 1, FontLab, MetaFog, InDesign, FMP, a2ac, vnT_EX, TDS, CTAN.ORG.

Hàn Thê Thànhem je autorem pdfT_EXu a i po mnoha letech jej stále udržuje a vyvíjí. Rozhovor s ním vedl Dave Walden, člen výboru TUG.



Dave Walden: *Můžete mi říci něco o sobě, věci nesouvisející s T_EXem?*

Hàn Thê Thànhem: Narodil jsem se v roce 1972 ve Vietnamu, kde jsem žil do roku 1990. Tehdy se mi naskytla příležitost studovat v České republice, tehdy ještě Československu. Studoval jsem na Masarykově univerzitě v Brně, od roku 1991 do roku 2001. Získal jsem titul magistr (Mgr.) a později doktorát (Ph.D.) v oboru informatika. Potom jsem se vrátil do Vietnamu a pracoval na Pedagogické univerzitě v Ho Či Minově městě (známém také jako Saigon). Od roku 2006 žiji s manželkou v německém Bielefeldu, kde manželka studuje.

Dave Walden: *Jakou práci jste na Pedagogické univerzitě dělal?*

Thànhem: Učil jsem úvod do programování a pracoval jsem jako správce sítě.

Dave Walden: *Můžete říci, kdy a jak jste se poprvé setkal s T_EXem?*

Hàn Thê Thànhem: Během prvních let na Masarykově univerzitě jsem od spolužáků občas o T_EXu slyšel – že je to skvělý sázecí systém, mocný, ale také těžký na zvládnutí. Ale sám jsem T_EX vůbec nepoužíval až do doby, kdy jsem

*This is a translation of the interview which was first published in the T_EX Users Group's Interview Corner, <http://tug.org/interviews/>, on July 24, 2008. Reprinted by permission of the interviewee and interviewer. Translation and corrections by Michal Mádr and Pavel Stríž.

si ve čtvrtém ročníku musel vybrat téma diplomové práce. Témat na výběr bylo hodně, já jsem si zvolil „Automatizované sázeční systémy“ nebo tak nějak. Představa vedoucího práce byla, že přepíšu $\text{T}_{\text{E}}\text{X}$ pomocí programovacího jazyka vyšší úrovně. Později se ukázalo, že by to pro studenta byl příliš obtížný úkol, a tak vedoucí téma změnil na „Sázeční systém $\text{T}_{\text{E}}\text{X}$ a PDF“. Cílem bylo vytvořit to, co $\text{pdfT}_{\text{E}}\text{X}$ dnes skutečně dělá – upravit $\text{T}_{\text{E}}\text{X}$ tak, aby jako svůj výstup produkoval PDF. Nejdříve jsem se ale musel s $\text{T}_{\text{E}}\text{X}$ em vůbec seznámit – do té doby jsem o něm slyšel, ale nikdy jsem ho nepoužil. Tohle se odehrálo v roce 1994.

Dave Walden: *Dalo by se říci, že klíčem k dnešní životaschopnosti $\text{T}_{\text{E}}\text{X}$ u je existence právě programu $\text{pdfT}_{\text{E}}\text{X}$. Z toho, co jste řekl, se zdá, že jste se k práci na $\text{pdfT}_{\text{E}}\text{X}$ u, který dnes používá každý, dostal náhodou – že vás vedoucí diplomové práce nasměroval, spíš než že byste o tuto oblast měl zájem vy sám. Je to tak?*

Hàn Thê Thành: Ano, takto opravdu začátky $\text{pdfT}_{\text{E}}\text{X}$ u vypadaly. Můj vedoucí, profesor Jiří Zlatuška, byl v té době velmi aktivním uživatelem i programátorem $\text{T}_{\text{E}}\text{X}$ u. Jiří je také fanouškem logického programování. Jeho původním záměrem tedy bylo, abych $\text{T}_{\text{E}}\text{X}$ přepsal pomocí deklarativního jazyka typu Prolog a výsledek použil pro další vývoj. Sotva jsem tehdy tušil, co to bude obnášet. Téma jsem si vybral, protože: 1) logické programování se mi líbilo; 2) podle toho, co jsem o $\text{T}_{\text{E}}\text{X}$ u slyšel, vypadalo téma zajímavě; 3) zajímavější téma k dispozici nebylo.

Po několika měsících experimentování s přepsáním $\text{T}_{\text{E}}\text{X}$ u bylo mně i Jiřímu jasné, že implementace $\text{T}_{\text{E}}\text{X}$ u v Prologu by pro mě byla příliš obtížná. Jednoho dne mě Jiří pozval k sobě do kanceláře, dal mi vytištěnou specifikaci PDF verze 1.0 a řekl, že bych mohl upravit $\text{T}_{\text{E}}\text{X}$ tak, aby produkoval výstup v PDF (později jsem se dozvěděl, že tento nápad vznikl v diskuzích Jiřího s Philem Taylorem a Donaldem Knuthem na Stanfordské univerzitě.). Tento nápad se mi líbil, navíc původní zadání se zdálo nerealistické. Takže jsme náš plán změnili. Začal jsem pročítat specifikaci PDF a zjišťovat, jak přizpůsobit $\text{T}_{\text{E}}\text{X}$ s Knuthovým systémem Web, Web2C, Kpathsea a dalšími. Po několika měsících jsem byl schopen vytvořit PDF soubor „Ahoj, světe!“ z $\text{T}_{\text{E}}\text{X}$ ového zdroje – to byl pro nás docela vzrušující okamžik. Ale že dnes bude $\text{pdfT}_{\text{E}}\text{X}$ tak rozšířený, mě tehdy nenapadlo. (A myslím, že ani Jiřího, i když se můžu mýlit.)

$\text{T}_{\text{E}}\text{X}$ jsem se učil obtížně: Začal jsem čtením knihy $\text{T}_{\text{E}}\text{X}$ book, kterou mi na začátku Jiří dal. Potom jsem začal používat plain $\text{T}_{\text{E}}\text{X}$, protože jsem slyšel, že $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$ není tak dobrý v případě, že se chcete naučit detaily $\text{T}_{\text{E}}\text{X}$ u – kontrolovat každou část sazby atd. Takže jsem pomocí plain $\text{T}_{\text{E}}\text{X}$ u sázel periodika, diplomové práce kamarádů a další příležitostné materiály. Ale později jsem začal $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$ používat, protože dělat všechno v plainu je docela komplikované. Takže většinou dávám přednost $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$ u a plain $\text{T}_{\text{E}}\text{X}$ používám jen pro pár specifických aplikací, na které se hodí lépe.

Učení se formátu PDF nebylo příliš obtížné, protože specifikace PDF verze 1.0 se vlezla do útlé knížečky. (Kdybych ale musel vycházet z verze 1.3 nebo některé

pozdější, hned bych to vzdal.) Ale učení se Webovým změnovým souborům, Web2C a podobným věcem bylo docela obtížné: musí se udělat příliš mnoho kroků a v případech problémů není lehké najít chybu.

Jiří chtěl, abych se držel přístupu literárního (kultivovaného) programování a všechno dělal pomocí mechanismu změnových souborů. To ale bylo čím dál obtížnější, a tak jsem se rozhodl naprogramovat část věci v C. Pro rozhodování, co ponechat v Pascalovském Webu a co naprogramovat v C, jsem použil toto kritérium: věci týkající se „back-endu“ se udělají v C, zbytek se udělá ve Webu. Jiřímu se tento přístup moc nelíbil, ale víceméně ho akceptoval (přinejmenším mě nechal ho zrealizovat).

Dave Walden: *Dokončil jste fungující verzi pdf_{TEX}u už v rámci diplomové práce, nebo až později?*

Hàn Thê Thành: Už si nepamatuji, kdy přesně se pdf_{TEX} stal „použitelným“, protože jeho vývoj byl povolný a v různých oblastech k němu přispívalo mnoho lidí. Ve chvíli, kdy jsem dokončil diplomovou práci, byl pdf_{TEX} víceméně ve stavu popsaném v článku Petra Sojky: podpora pro začleněné fonty Type 1, virtuální fonty, odkazy, LZW kompresi (později byla místo LZW použita ZIP-komprese). Ale vkládání obrázků ještě podporováno nebylo!

Dave Walden: *Máte na mysli článek „The Joy of _{TEX}2PDF – Acrobatics with an Alternative to DVI Format“ od Petra Sojky, vás a Jiřího, který vyšel v roce 1996 v TUGboatu, ročníku 17, čísle 3?*

Hàn Thê Thành: Ano.

Dave Walden: *Jak se o vašem programu pdf_{TEX} dozvěděli lidé okolo _{TEX}u a jak došlo k tomu, že byl zařazen do všech _{TEX}ových distribucí?*

Hàn Thê Thành: Poprvé jsem se do kontaktu s lidmi okolo _{TEX}u dostal poté, co si Jiří ohledně pdf_{TEX}u (tehdy ještě stále nazývaném _{TEX}2PDF) dopisoval se Sebastianem Rahtzem. Sebastian pdf_{TEX} zaujal, začal si s ním hrát a různě ho podporovat: založil mailing list pdf_{TEX}u, přeložil a testoval pdf_{TEX} na dalších platformách, seznámil s ním další _{TEX}ové uživatele atd. Sebastian vývoji pdf_{TEX}u v jeho začátcích hodně pomohl.

Později byl pdf_{TEX} ukázán Donaldu Knuthovi během jeho návštěvy Masarykovy univerzity a Knuth se o něm vyjádřil pochvalně. To pro nás (mě a Jiřího) bylo velmi povzbudivé. Prvním článkem na téma pdf_{TEX} byl ten, který jsem už zmínil, jehož autorem byl Petr Sojka. Mailing list pdf_{TEX}u byl v počátcích velmi užitečným místem pro diskuze o vývoji pdf_{TEX}u. A jednoho dne se na mailing listu objevil Hans Hagen, začal s pdf_{TEX}em experimentovat, upozorňovat na problémy, diskutovat nové nápady a rysy atd., což mělo na pdf_{TEX} také velký vliv. To, že se pdf_{TEX} líbil tak známým a aktivním členům _{TEX}ové komunity (Sebastianovi, Hansovi atd.), bylo klíčové pro to, aby se o pdf_{TEX}u vědělo více.

Potom ho Thomas Esser zařadil do TeXu . A jakmile se něco dostane tam, obvykle to pak bude akceptováno i ostatními TeXovými systémy.

Dave Walden: *Můžete vyjasnit rozdíl (pokud nějaký je) mezi pdf TeXem a microtypografickým rozšířením TeXu , které bylo popsáno ve vaší disertační práci a publikováno v TUGboatu jako článek „Microtypographic extensions to the TeX typesetting system“?*

Hàn Thê Thành: Cílem mé diplomové práce bylo umožnit, aby TeX produkoval výstup v PDF. Když jsem začínal své doktorské studium (Jiří byl opět mým vedoucím), věděli jsme jen, že budeme pracovat na něčem týkajícím se sazby. Během zhruba prvního roku mého doktorského studia jsem stále vyvíjel pdf TeX a zároveň se poohlížel po tématu disertační práce. Pár témat mě napadlo, potom ale Jiří rozhodl, že zůstaneme u mikrotypografických rozšíření. Myslím, že to opět bylo rozumné rozhodnutí.

Dave Walden: *Řekl jste, že jste na Masarykově univerzitě studoval v letech 1991 až 2001. Můžete upřesnit, kdy jste ukončil magisterské studium a kdy začalo vaše doktorské studium?*

Hàn Thê Thành: Magisterské studium jsem dokončil v létě 1996. O pár měsíců později jsem začal své doktorské studium.

Dave Walden: *Můžete blíže popsat váš přístup k studiu mikrotypografie, k tvorbě nutných rozšíření pdf TeXu a k výzkumu komponent nezbytných pro úspěšnou disertační práci?*

Hàn Thê Thành: Necítím se moc kvalifikovaný radit, jak udělat úspěšnou disertační práci, protože se svou vlastní jsem zápasil.

Dave Walden: *Promiňte, nechtěl jsem vám připomínat stresující okamžiky. Zajímalo mě jen, jak jste se naučil to, co jste potřeboval vědět o mikrotypografii.*

Hàn Thê Thành: Už si přesně nepamatuji, jak jsem mikrotypografii studoval – byl to pozvolný proces, asi jako pro většinu lidí. Začal jsem čtením knih a článků a potom jsem hledal další relevantní zdroje informací. Ty nejužitečnější, na které si vzpomínám, byly články Hermanna Zapfa „About microtypography and the *hz*-program“ a brožura o *hz*-programu od URW, německé písmolijny. Také jsem hodně experimentoval s InDesignem firmy Adobe, o kterém se tvrdí, že obsahuje integrované moduly *hz*-programu. Je zajímavé, že některé ideje *hz*-programu byly původně inspirovány TeXem samotným.

Dave Walden: *Vím, že než jste dokončil doktorské studium, byl už pdf TeX „zhruba“ funkční. Pokračoval jste v jeho vývoji i po návratu do Vietnamu?*

Hàn Thê Thành: Po návratu do Vietnamu jsem kvůli problémům s přístupem k síti a dalšími věcmi na dlouhou dobu s vývojem pdf TeXu přestal. Později jsem příležitostně našel čas dělat drobná rozšíření, ale už to nebyl aktivní vývoj

jako dříve. Na pdf \TeX u samozřejmě pracovali také další lidé. Nejvýznamnější příspěvky v minulých letech udělal Hartmut Henkel, velmi tichý člověk. Jeho opravy pdf \TeX výrazně vylepšily v mnoha ohledech: v rychlosti, stabilitě, čistotě a přehlednosti zdrojového kódu, lepší funkcionalitě atd.

Dave Walden: *Zdá se, že se nyní na vývoji pdf \TeX u opět podílíte výrazněji.*

Hàn Thê Thành: Ano, od chvíle, kdy jsem se s manželkou přesunul do Německa, jsem s pdf \TeX em v intenzivnějším kontaktu. V Německu pracuji z domu, jako konzultant společnosti River Valley Technologies. Dělán podporu síťové administrativy, automatizování editačních procesů a také nasazení pdf \TeX u.

Dave Walden: *Předpokládám, že vývoj pdf \TeX u funguje v jistém smyslu jako vývoj Open Source Softwaru. A minulý týden jste mi napsal, že musíme interview na pár dní přerušit, protože Karl Berry po vás chce rychlou opravu v pdf \TeX u pro připravované vydání \TeX Live. Můžete přiblížit, jak je vývoj pdf \TeX u organizován a koordinován?*

Hàn Thê Thành: Vývoj pdf \TeX u se časem ustálil a vypadá zhruba takto: projekt pdf \TeX má zřízenou stránku na <http://sarovar.org/>, kam lidé hlásí chyby, požadavky na novou funkcionalitu a opravy. Pro zájemce o vývoj pdf \TeX u existuje také mailing list. A pak je zde vývojářské jádro (Hans Hagen, Taco Hoekwater, Hartmut Henkel, Martin Schröder a já), kde pdf \TeX ové záležitosti diskutujeme.

Dave Walden: *Jestli se nemýlím, strávil jste hodně času tvorbou podpory pro vietnamštinu u mnoha písem. Jak jste se k tomu dostal a jak postupujete?*

Hàn Thê Thành: Ve chvíli, kdy jsem se \TeX učil, jsem ho mimo jiné zamýšlel použít i na sazbu vietnamštiny. V té době už existoval balík nazvaný vcmr od Wenera Lemberga, který poskytoval docela dobrou podporu. Mně se ale nelíbily tvary vietnamských znaků, a tak jsem se rozhodl tyto znaky sám přidat do písem Computer Modern. Byla to pro mě jen zábava, nemám žádnou uměleckou minulost. Učil jsem se z existujících písem, z textů, ke kterým jsem se dostal a z komentářů od zkušených lidí. Vietnamská písmena jsem do písem Computer Modern přidal pomocí METAFONTu. Ke konverzi fontů do formátu Type 1 jsem zkombinoval několik nástrojů: MetaFog, FMP (od YandY), a2ac a své vlastní skripty v Perlu. Stejně nástroje jsem použil pro přidání vietnamských písmen do existujících Type 1 písem, akcenty jsem však nakreslil pomocí FontLabu. Na vn \TeX u se podílí další lidé: Werner Lemberg a Vladimir Volovich (podpora pro \LaTeX) a Reinhard Kotucha (testování a údržba balíku, kompatibilita s TDS a poskytování věcí vyžadovaných od \TeX Live a CTAN.ORG, např. soubor README, copyright atd.) Vývoj vn \TeX u už nepokračuje, protože už bylo vytvořeno docela velké množství písem. Existuje dokonce vietnamský překlad průzkumu o matematických písmech pro \TeX od Stephena Hartkeho, což dokládá, že většina písem zmíněných v průzkumu má verzi pro vietnamštinu.

Dave Walden: *Děkuji, Thành, že jste si našel čas na náš rozhovor. Bylo mi ctí být v kontaktu s někým, kdo má tak výraznou zásluhu na trvajícím oblíbení T_EXu.*

~ ~ ~

Lidé kolem pdfT_EXu: Hàn Thế Thành, Jiří Zlatuška, Phil Taylor, Donald Ervin Knuth, Petr Sojka, Sebastian Rahtz, Hans Hagen, Thomas Esser, Hermann Zapf, Hartmut Henkel, Karl Berry, Taco Hoekwater, Martin Schröder, Werner Lemberg, Vladimir Volovich, Reinhard Kotucha, Stephen Hartke, Pavel Janík, Heiko Oberdiek, Jiří Osoba, Ricardo Balbino Sánchez Cármenes, Robert Schlicht.

~ ~ ~

Závěrečná poznámka: Při kontrole pořádaného rozhovoru Thành upozornil na tyto důležité spolutvářce pdfT_EXu: Pavel Janík přidal podporu pro TIFF (která byla později odstraněna); Heiko Oberdiek přidal podporu pro color stack [uložení aktuální barvy při přechodu na novou stranu PDF]; Jiří Osoba přidal podporu JPG, Ricardo Sánchez Cármenes přidal podporu šifrování (později odstraněnou); Robert Schlicht vytvořil L^AT_EXový balík s podporou pro mikrotypografii; Martin Schröder po mnoho let udržoval pdfT_EX. A samozřejmě, pdfT_EX je jen rozšířením programu T_EX. Kdyby Donald Knuth nenapsal T_EX, nevznikl by ani pdfT_EX. Pokud jsme na někoho zapomněli, velice se omlouváme!

Z pdfT_EXu vychází ambiciózní projekt LuaT_EX, který bude postupně představován v dalších článcích Zpravodaje.

Zmíněná a doporučená literatura

- [1] *Interview with Hàn Thế Thành.* [Interview s Hàn Thế Thànhem.] [online, vytvořeno 24. 7. 2008] Dostupné z URL: <http://www.tug.org/interviews/interview-files/han-the-thanh.html>
- [2] *T_EX People: Interviews from the World of T_EX.* [Mistři T_EXu: Rozhovory s lidmi ze světa T_EXu.] Berry, Karl (editor); Walden, David (editor). USA, T_EX Users Group, 2009. 312 stran. ISBN 9780982462607. Více informací o knize na URL: <http://tug.org/interviews/book/>
- [3] Sojka, Petr; Thế Thành, Hàn; Zlatuška, Jiří. The Joy of T_EX2PDF – Acrobatics with an Alternative to DVI Format. [Úspěchy programu T_EX2PDF – akrobacie s alternativním výstupem do formátu DVI.] *TUGboat*, vol. 17(3): 244–251, 1996. ISSN 0896-3207. Článek je dostupný z URL: <http://www.tug.org/TUGboat/Articles/tb17-3/tb52sojk.pdf>
- [4] Thế Thành, Hàn. Experience from a Real-World Application of Microtypography with pdfT_EX. [Zkušenosti s mikrotypografií v pdfT_EXu z reálných úloh.] *The Asian Journal of T_EX*, vol. 2(1): 1–10, Duben 2008. Proceedings of the Asian T_EX Conference 2008, the Korean T_EX Society. ISSN 1976-1228. Dostupné z URL: <http://ajt.ktug.kr/assets/2008/5/1/0201thanh.pdf>

- [5] Thế Thành, Hàn. Experiences with Micro-Typographic Extensions of pdf \TeX in Practice. [Praktické zkušenosti s mikrotypografickými rozšířeními pdf \TeX u.] *Die \TeX nische Komödie*, vol. 18(2): 159–164, 2006. Deutschsprachige Anwendervereinigung \TeX e.V. Proceedings Euro \TeX 2005, Pont-à-Mousson, France. ISSN 1434-5897. Dostupné z URL: <http://www.tug.org/TUGboat/Articles/tb27-0/tb85complete.pdf>
- [6] Thế Thành, Hàn. Margin Kerning and Font Expansion with pdf \TeX . [Optické vyrovnávání okrajů a natahování znaků pomocí pdf \TeX u.] *TUGboat*, vol. 22(3): 146–148, 2001. Proceedings of the 2001 Annual Meeting. ISSN 0896-3207. Dostupné na URL: <http://www.tug.org/TUGboat/Articles/tb223/tb72thanh.pdf>
- [7] Thế Thành, Hàn. Micro-typographic extensions to the \TeX typesetting system. [Mikrotypografická rozšíření v typografickém systému \TeX .] Disertační práce, Masarykova university v Brně, 2000. Školitel práce: Jiří Zlatuška. Přetiskáno v časopisu *TUGboat*, vol. 21(4): 317–434, 2000. ISSN 0896-3207. Dostupné na URL: <http://www.tug.org/TUGboat/Articles/tb21-4/tb69thanh.pdf>

Summary: Interview with Hàn Thế Thành, the Creator and Maintainer of the pdf \TeX

Translation of an interview conducted by Dave Walden with Hàn Thế Thành, the creator and maintainer of the pdf \TeX program. Discussed were the origins of the program pdf \TeX , its early development, introduction to the wider \TeX community and the organization of the current development and maintenance.

Key words: \TeX , PDF, Adobe, pdf \TeX , \TeX 2PDF, Masaryk University, Microtypography, *hz*-program, \TeX book, Web, Web2C, Kpathsea, the programming language C, plain \TeX , \LaTeX , TUGboat, te \TeX , \TeX Live, Prolog, TUG, Vietnamese, Typeface, Computer Modern, METAFONT, Type 1, FontLab, MetaFog, InDesign, FMP, a2ac, vn \TeX , TDS, CTAN.ORG.

Překlad a dokončení:

Michal Mádr, m.madr@seznam.cz

Pavel Stráž, bulletin@cstug.cz

ČSTUG c/o FEL ČVUT, Technická 2

Praha, CZ-166 27, Czech Republic