

Zpravodaj Československého sdružení uživatelů TeXu

Zdeněk Wagner
LaTeXe v CTeXu

Zpravodaj Československého sdružení uživatelů TeXu, Vol. 6 (1996), No. 1, 15–22

Persistent URL: <http://dml.cz/dmlcz/149749>

Terms of use:

© Československé sdružení uživatelů TeXu, 1996

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>

V tomto článku bude ukázáno, jak je v C_ST_EXu implementován L^AT_EX 2_ε. Zaměříme se zejména na problematiku národní podpory. T_EX, a tudíž i L^AT_EX, je postaven tak, aby pracoval stejným způsobem na všech operačních systémech. Přesto se jednotlivé implementace pro různé operační systémy poněkud liší. Tento článek se tudíž více dotkne té implementace, kterou používá autor, a to je emT_EX pro OS/2.

Instalační kontroverze

Při instalaci se mírně protichůdně stýkají čtyři různé koncepce. První z nich je koncepce autorů L^AT_EXu, druhou koncepcí je přístup autora emT_EXu – Eberharda Mattesa, další proud tvoří názor Petra Olšáka jako hlavního tvůrce C_ST_EXu a závěrem jsou to rozhodnutí spoluautora národní podpory.

L^AT_EX 3 team poskytuje dokumentovaný zdrojový kód. Ten se rozbalí T_EXovým programem DOCSTRIP a výsledné soubory se mají přesunout „tam, kde je T_EX očekává“. Pak se spustí iniT_EX a formát je hotov. Rozbalovací fáze trvá na méně výkonných počítačích dlouho, proto se L^AT_EX dodává i v rozbalené podobě.

E. Mattes poskytuje původní i rozbalenou verzi. Dokumentovaný zdrojový kód je v adresáři `emtex\dist\latex\base\`, všechny soubory, které vyprodukoval DOCSTRIP, jsou v `emtex\texinput\latex2e\`. Tam také `tex*.exe` hledá vstupní soubory. Po vytvoření formátu však zůstává v pracovních adresářích řada souborů, které již nejsou potřebné a zbytečně zabírají místo těm, kteří mají na svém disku těsno. L^AT_EX ovšem používá řadu souborů interně. Běžný uživatel, který nechce číst dlouhou dokumentaci, tedy neví, co může beztréstně smazat.

Petr Olšák správně tvrdí, že uživatel by měl vytvořit formáty, načež může smazat celý strom `emtex\dist\`, aby mu na disku nepřekážely soubory, které nepotřebuje. Vyžaduje ovšem, aby soubor `cslatex.ini` a další konfigurační soubory byly přítomny v adresáři

`emtex\texinput\latex2e\cslatex\`. Tam také zůstanou po vytvoření formátu a zvláště v DOSu na FAT blokuji tyto kraťoučké soubory zbytečně mnoho místa. Z hlediska šetření místa by jejich umístění v `emtex\dist\cslat2e\` bylo logičtější, ale to by vyžadovalo nemalé zá-
sahy do struktury $\mathcal{C}\mathcal{T}\mathcal{E}\mathcal{X}$ u.

Současná instalace české a slovenské verze $\mathcal{L}\mathcal{A}\mathcal{T}\mathcal{E}\mathcal{X}$ u musí splňovat řadu částečně protichůdných požadavků. Předně musí být slučitelná se zbytkem $\mathcal{C}\mathcal{T}\mathcal{E}\mathcal{X}$ u. Musí proto vyhovovat přání Petra Olšáka. Musí být schopna koexistence s $\mathcal{L}\mathcal{A}\mathcal{T}\mathcal{E}\mathcal{X}$ em 2.09, takže je nutno dodržet adresářovou strukturu, kterou vytvořil Eberhard Mattes. Navíc se zde uplatňuje to, co v úvodním článku tohoto čísla napsal Petr Olšák – autoři vytvářejí tento produkt v jiném systému než je MS DOS. Původní verzi národní podpory napsal Jiří Zlatuška a podle jmen souborů je zřejmé, že byla vytvářena na UNIXu. Ode mě pocházejí úpravy, jejichž cílem je přiblížení celé implementace potřebám $\mathcal{C}\mathcal{T}\mathcal{E}\mathcal{X}$ u. Přitom se vychází z mých vlastních potřeb.

Národní podpora je implementována pomocí konfiguračních souborů. Ty vytvořil Jiří Zlatuška podle svých požadavků. Já jsem je dále modifikoval, takže i tyto konfigurační soubory jsou dále konfigurovatelné. Pokud tyto konfigurační soubory „druhého řádu“ nejsou přítomny, vytvoří se stejný formát, jaký má Jiří Zlatuška. Tím se snad přispěje k lepší unifikaci.

Moje verze běhá pod OS/2, kde nejsou problémy s omezením paměti na 640 kB (velikost virtuální paměti je omezena pouze volným místem na pevném disku), přičemž požadují vzory dělení pro řadu neobvyklých jazyků, jako je dánština apod. Takový formát je pro běžné české a slovenské uživatele nepotřebný a svým rozsahem dokonce může až obtěžovat. Je tedy účelné vytvořit pro $\mathcal{C}\mathcal{T}\mathcal{E}\mathcal{X}$ formát jednodušší, který pro kompatibilitu se světem bude obsahovat vzory dělení pro americkou angličtinu a pro nás vzory dělení české a slovenské. To přináší rozdíl mezi instalací „oficiální“ a soukromou, což může být zdrojem chyb. Proto je při instalaci prohlížen další adresář, kam uživatel (hlavně já) může podstrčit svůj konfigurační soubor a vytvořit si tak jinou verzi s jinými vzory dělení. Pak ovšem nelze použít instalační dávku `inibat` z $\mathcal{C}\mathcal{T}\mathcal{E}\mathcal{X}$ u, ale musí se použít postup, který bude uveden později. Takový konfigurační soubor musí být v adresáři `emtex\dist\cslatex.cfg\`. Výhoda umístění uživatelských konfiguračních souborů mimo standardní větev spočívá v tom, že se při instalaci nové verze $\mathcal{L}\mathcal{A}\mathcal{T}\mathcal{E}\mathcal{X}$ u nepřemažou. Tímto mechanismem

se v autorově dílně vytváří jak formát soukromý, tak formát oficiální, a díky vlastním dávkám v OS/2, které byly nyní v nouzi úspěšně převedeny i do DOSu, lze oba formáty provozovat současně. Přitom dlouhá jména souborů nejsou naštěstí problémem. Při rozzipování pod DOSem se překousnou podle konvence 8+3 a emTeX si umí zkrátit jména uvedená v makru `\input` tak, že se soubory správně najdou.

Proč používat L^AT_EX 2_ε?

Tato otázka sice patří úplně do jiného dokumentu, ale přesto ji zodpovíme zde. L^AT_EX 2_ε totiž při instalaci způsoboval jisté obtíže a mnohým nezasvěceným uživatelům by se mohlo zdát, že jejich překonávání za tu námahu nestojí a mohou si klidně vystačit se starou verzí 2.09. To ovšem platí pouze pro ty uživatele, kteří se spokojí s nenáročnými texty.

Základním problémem L^AT_EXu jsou fonty. V původní verzi s old font selection scheme jsou dosti pevně zadrátovány fonty computer modern. Jednotlivý font lze zavést makrem `\newfont` i primitivem `\font`. Pak ale nefungují přepínače `\bf`, `\small`, apod. Plnohodnotné zavedení nových fontů do L^AT_EXu 2.09 není triviální záležitostí. Tento problém, mimo jiné, řeší new font selection scheme. Nový font se zavádí pomocí jednoduchých deklarací. Ortogonální přepínače umožňují použití *tučné kurzívy* prostým uvedením `\it\bf`. Za tuto výhodu však platíme nekompatibilitou. Řada dokumentů, psaných pro NFSS, nelze s OFSS zpracovat, dokumenty psané pro OFSS poskytují nesprávný výstup v NFSS, zejména v matematickém režimu. A právě to řeší L^AT_EX 2_ε.

Všichni víme, že L^AT_EX 2_ε se trvale mění a dvakrát ročně je nová verze. Je to tím, že konečným cílem je L^AT_EX 3 a L^AT_EX 2_ε je vlastně testovací formát pro nové koncepce. Kromě kompatibility starých a nových dokumentů je zde pohodlně podporována vícejazyčná sazba, kdy lze současně používat fonty v různých kódováních. Dokonce i pro jeden jazyk lze použít v jednom dokumentu různě kódované fonty, přičemž dělení funguje ve všech kódováních správně. Logickou nevýhodou jsou značné nároky na paměť a výrazně pomalejší překlad. Uživatelé s předpotopným PC pak mohou neúspěšné pokusy o vytištění i nejjednodušších dokumentů odradit. Pro urychlení existují programy, s jejichž pomocí lze rychle přelatexovat označenou část zdrojového textu a zobrazit v prohlížeči. Na současných počítačích to však není nutné. Na 66 MHz 80486

v OS/2 pracuje `tex386.exe` tak rychle, že jsem při překladu celé knihy rád, když mi počítač dopřeje pár sekund odpočinku.

Z toho vyplývá následující doporučení: pokud chcete vážně pracovat s \TeX em, chcete využívat PostScriptové fonty, sázíte vícejazyčné dokumenty, máte fonty z různých zdrojů s různým kódováním a nechcete ztrácet čas vytvářením virtuálních skriptů pro překódování, zapomeňte na DOS a Windows, pořídte si pořádný operační systém (OS/2 nebo UNIX) a používejte \LaTeX 2_ε. Pokud snad chcete zůstat u procesoru 80286 či horšího a přejdete na \LaTeX 2_ε, nezbyde vám nic jiného než nečinně přihlížet, jak vaše stránky zvolna přezvykují big \TeX .

Vstupní a výstupní kódování

\TeX byl v původní verzi sedmibitový a zpracovával jen první polovinu ASCII tabulky. Akcentované znaky se musely zadávat pomocí maker, což je pro české i slovenské texty velmi nepohodlné. Naštěstí Don Knuth vytvořil osmibitovou verzi, při jejímž použití lze psát texty v editoru s českou či slovenskou klávesnicí.

Původní cm-fonty však nemají akcentované znaky, obsahují pouze samostatná diakritická znaménka a česká i slovenská písmenka se z nich dají složit. To se ale dosahuje různými triky. Zdrojový text se může překonvertovat filtrem tak, aby obsahoval pouze sedmibitové znaky a makra a tento soubor se pak předloží \TeX u. Tím ale trpí kvalita dělení. Proto Oldřich Ulrych podstrčil \TeX u upravené metriky a programem `dvi2dvi` provedl příslušné změny až v DVI-souboru. Vše by vyřešily osmibitové fonty, ale původní dc-fonty nerespektovaly zcela česká typografická pravidla. Proto byly vytvořeny \CzTeX -fonty.

Železná opona kdysi vytvořila kódovou džungli. Snad každý začal na nějakém PC používat češtinu Kamenických, která ale z technických důvodů nešla použít na nejstarších laserových tiskárnách. IBM, Hewlett-Packard a MicroSoft vytvořily společně kódovou stránku 852, kterou oficiálně používá DOS i OS/2. V RVHP byla normalizována KOI-8-ČS, která se ale téměř nepoužívá. Na UNIXových strojích zvítězilo kódování ISO 8859-2. Windows, aby to nebylo tak jednoduché, používají kódovou stránku 1250, která se bezdůvodně liší od ISO 8859-2 v pěti znacích. Windows '95 jsou samostatná kapitola a raději se o nich rozepisovat nebudu. \CzTeX tuto džungli bohužel ještě prohloubil. Při tvorbě \CzTeX u '94

se mylně předpokládalo, že v UNIXu musí být kódování fontů shodné se vstupním kódováním. \mathcal{G} -fonty mají proto kódování IL2, což znamená, že spodní polovina má stejné kódování jako původní cm-fonty, horní polovina je kódována podle ISO 8859-2. \mathcal{G} -fonty lze tedy bez obav použít místo cm-fontů, ale jinak nejsou kompatibilní s ničím. Pokud chce uživatel společně s \mathcal{G} -fonty použít dc-fonty nebo jiná písma, dostupná pravděpodobně v kódování T1 (kód dc-fontů), musí se s tím nějak vypořádat.

První řešení je čistě \TeX ovské. Lze vytvořit virtuální skript, jímž se font překóduje. Práci sice usnadní Mattesův `qdTeXvp1`, ale nějaký čas to stejně zabere. Navíc se to s každým novým fontem musí provádět znova.

Jinou možnost nabízí standardní \LaTeX 2 ϵ . Příkazy pro změnu fontů totiž zahrnují i přepínač pro změnu kódování. Pak lze akcentované znaky zadávat pomocí maker a během sazby je \LaTeX 2 ϵ převede na osmibitový kód aktuálního fontu. Dělení tedy funguje spolehlivě, ale pro tvorbu českých a slovenských dokumentů opět potřebujeme filtr. Chceme-li se filtru vyhnout, musí být akcentované znaky aktivní a musí se expandovat na příslušná makra, která pak \LaTeX 2 ϵ zkonvertuje. Toto řešení vede k téměř 100% přenositelnosti zdrojových textů, ovšem nefunguje správně `\uppercase`.

UNIXový \TeX na rozdíl od `emTeXu` neumí tcp-tabulky, takže následující vlastnost v něm nelze zařídit snadno (vyžaduje překlad s jiným změnovým souborem). V tcp-tabulce `emTeXu` lze totiž specifikovat, že akcentované znaky se mají při čtení nahradit \TeX ovskými makry. Pak lze míchat fonty s různým kódováním a funguje i `\uppercase`. Bez problému mohu napsat i `ň` a `č`, což umí moje óesdvojková klávesnice, a \LaTeX 2 ϵ vše vysází správně, přestože v \mathcal{G} -fontech tyto znaky nejsou. Každá výhoda však má i nevýhody. Pokud si zobrazíte český text makrem `\typeout`, uvidíte místo českých písmenek na obrazovce \TeX ovská makra. To lze ještě přelustit, daleko horší jsou hlášení o přetečených a podtečených boxech. Zde nám \TeX nabídne pouze hexadecimální výstup se zobáky, což se čte dosti těžko. To ale pořád není to nejhorší.

Představte si, že máte starý dokument, který jste kdysi vysadili v \LaTeX u 2.09 a nyní jej sázíte znovu. Použijete \LaTeX 2 ϵ , tcp-tabulka převede akcentované znaky na makra, spustíte prohlížeč a s hrůzou zjistíte, že prostředí `tabbing` se zcela rozpadlo. Tam totiž má `\'` zcela jiný význam. Naštěstí se `tabbing` vždycky dá přepsat pomocí balíku `longtable`. Pro `verbatim` však taková jednoduchá pomoc neexistuje.

Podobný problém přinesou i aktivní akcentované znaky. Lze sice napsat chytřejší makra, která by principiálně ošetřila `tabbing` i `verbatim`, ale nejsem si jist, zda to již současná verze dělá.

Jednoduché řešení tedy neexistuje. autor používá více různých formátů podle okamžitých potřeb. Podle počtu použitých vzorů dělení zabírá formát `LTEX`u 700–800kB, což si lze celkem dovolit i na malém disku.

Instalace

Základní instalace v `CSTEX`u spočívá v tom, že se rozzipují příslušné soubory a spustí se dávka `inibat`. Ta vytvoří standardní formát `cslatex`. Ten obsahuje vzory dělení pro americkou angličtinu, češtinu v kódování IL2 i T1 a slovenštinu v kódování IL2. Více se do paměti malého `TEX`u nevejde. Standardní `tcp`-tabulka převádí vstup z kódu Kamenických do IL2. Formát lze vygenerovat i malým `TEX`em, ale řadu dokumentů, zejména v režimu kompatibility s `LTEX` 2.09, nelze malým `TEX`em zpracovat.

Pokud máte jiné požadavky na konverzi vstupních znaků nebo chcete instalovat vzory dělení pro jiné jazyky, musíte vygenerovat jiný formát. Je vhodné nazvat jej jinak, i když bude kompatibilní s formátem `cslatex` a běžné české texty se v obou formátech vysází stejně. Bohužel nevím, jak upravit `CSTEX`, aby vygeneroval jiný formát. Lze to však učinit použitím Rexxové dávky `cslatex.cmd` pro OS/2 nebo `cslatex.bat` pro DOS. Obě dávky jsou v adresáři `emtex\dist\cs1t2e\`. Jsou dosti složité a jejich přesný popis by zabral příliš mnoho místa. V následujícím textu bude pouze naznačeno, jak si lze `LTEX` 2_ε překonfigurovat. Podrobnější údaje jsou pro zájemce uvedeny v dokumentaci. V OS/2 si můžete přečíst hypertextový dokument `emtex\book\czech\cs1txins.inf`, pro ostatní je k dispozici „lineární“ návod `emtex\doc\czech\cs1txins.*`, kde přípona specifikuje kódování textu.

Instalační dávky plní řadu úkolů, v OS/2 poměrně efektivně, neboť Rexx je dosti silný programovací nástroj, v DOSu hloupě a vyžaduje velký environment. Dokud jsem ještě používal DOS, musel jsem nastavit prostor pro environment na 2048, aby mé dávky fungovaly. To je první věc, kterou musíte zajistit, když se rozhodnete pro použití instalační dávky `cslatex.bat`. V OS/2 tento problém odpadá.

Mé spojení na Internet je pomalé. Proto se mi někdy nevyplácí čekání, až se mi konečně natáhne rozbalená forma a raději si pustím DOCSTRIP sám. Instalační dávka tedy sama testuje, zda je $\text{\LaTeX} 2_{\epsilon}$ již rozbalen. Podobně se zachází i s podporou češtiny a slovenštiny; zde je však důvod jiný. Drobná potíž je i s dc-fonty. Nový $\text{\LaTeX} 2_{\epsilon}$ podporuje nové dc-fonty. Pokud má někdo staré dc-fonty, musí rozbalit další soubor. I to kontrolují mé instalační dávky a vše provedou automaticky.

Jak již bylo uvedeno dříve, $\text{\LaTeX} 2_{\epsilon}$ je konfigurovatelný prostřednictvím konfiguračních souborů. Ty standardní najdete převážně v adresáři `emtex\texinput\latex2e\cslatex\`. Pokud je smažete, pokusí se \LaTeX vytvořit stejný formát, jako má Jiří Zlatuška, a při tom \TeX v DOSu i v OS/2 zhavaruje. Pro svůj formát musíte tudíž napsat vlastní konfigurační soubory. Pokud přepíšete standardní konfigurační soubory, pak můžete o své změny při instalaci nové verze \CTEXu přijít. Proto instalační dávky prohlížejí přednostně adresář `emtex\dist\... \cslatex.cfg\`, do něhož žádná distribuce nic nezapisuje. Standardní dávka `inibat` se ovšem do tohoto adresáře nekouká. Pokud si měníte konfiguraci, nezbyvá vám nic jiného, než pročíst instalační návod a použít mé instalační dávky.

Ještě je nutno zmínit se o vzorech dělení. \CTEX používá parametr `/mt25000`, aby i malý \TeX měl dost paměti pro ostatní objekty. Pokud budete přidávat další jazyky, musíte také něco vyhodit. Bez zvětšení hodnoty parametru `/mt` se obejdete jen stěží. Pak ovšem bude nutné zasáhnout do dávek \CTEXu , jinak si bude \TeX stěžovat, že nemůže za běhu zvětšit trie size.

Ještě jednou kódování

Současným standardem jsou cm-fonty s kódováním OT1. \CTEX fonty jsou jejich rozšířením a jsou standardní součástí \LaTeXu v \CTEXu . \CTUG bude, doufáme, i nadále podporovat \CTEX -fonty, zatímco svět přejde na dc-fonty s kódováním T1. Někteří uživatelé udělali tento krok již nyní a $\text{\LaTeX} 2_{\epsilon}$ z \CTEXu jim to umožňuje. Stačí změna konfiguračního souboru. Na jednu stranu je to výhodné, ale přináší to problém. Snadno se totiž v makrech pro českou a slovenskou podporu něco přehlédne a \LaTeX pak nefunguje tak, jak má. Programování také není jednoduché, když není předem dáno, jaké kódování bude použito standardně. Navíc se

L^AT_EX 2_ε neustále mění. Uživatel to sice nepozná, ale programátor vnitřních maker musí s každou novou verzí poměrně dost předělávat. Tím vznikají různé problémy a byl bych rád, kdyby postupem času ubývaly.

Některé soubory chybí

Podobně jako vlastní L^AT_EX 2_ε, dodává se řada dalších rozšiřujících nástrojů ve formě dokumentovaných zdrojových kódů, které si uživatel musí sám rozbalit programem DOCSTRIP. Tyto nástroje již nejsou tak objemné, takže se na CTAN rozbalené nenajdou. Snad proto je E. Matthes dodává v původním tvaru a najdou se v adresáři `emtex\dist\...` `...latex\tools\`. Již několik uživatelů přišlo na to, že při pokusu o tisk dokumentace se nenajdou soubory `verbatim.sty` a `multicol.sty`. Uvažujeme o tom, že užitečné nástroje rozbalíme a připravíme do vhodných adresářů v C^ST_EXu. Také jsem byl upozorněn na to, že jsem do C^ST_EXu zapomněl vložit PSNFSS, takže PostScriptové fonty může nyní použít jen expert. Předpokládám, že v době, kdy čtete tento článek, jsou již uvedené problémy odstraněny.

Jak naložit s chybami?

Přes všechnu snahu chyby byly, jsou a budou. Pokud nějakou najdete, spojte se s autorem tohoto článku. Některé problémy lze dočasně napravit nějakou náplastí. Vždy však pomůžete sobě i ostatním, protože se můžete trefit do něčeho, co autor nepoužívá a chybu by tudíž nenašel. Další příčinou může být, že autor má nějaké soubory doma, všechno funguje, ale zapomněl je zabalit do distribuce. E-mailovou adresu najdete v podpisu i ve všech distribucích.

Zdeněk Wagner
wagner@mbox.cesnet.cz