

Zpravodaj Československého sdružení uživatelů TeXu

Joachim Schrod
Komponenty TeXu

Zpravodaj Československého sdružení uživatelů TeXu, Vol. 3 (1993), No. 1, 1–10

Persistent URL: <http://dml.cz/dmlcz/149650>

Terms of use:

© Československé sdružení uživatelů TeXu, 1993

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>

T_EX vyžaduje pomerne veľa doplňujúcich komponentov (súborov i programov), ktorých zmysel a vzájomná súčinnosť nie sú často známe. Tento článok osvetľuje jednotlivé zložky a vzájomné väzby jadra systému T_EX, ktoré sú viditeľné pre užívateľa T_EXu.

O obsahu tohto článku

T_EX je sadzačský systém, ktorý ponúka autorovi ľahké využitie sadzačských vymožeností tlače na úrovni umenia sadzby pomocou počítačov. Toto však nie je len výsledkom programu T_EX: značný počet doplňujúcich programov a súborov vytvára spolu kompletný sadzačský a autorský systém. Spolu s tými programami, ktoré patria priamo do T_EXu, existujú ešte dva ďalšie veľké programy, ktoré boli vytvorené DONALDOM KNUTHOM v spojitosti s T_EXom a ktoré musia byť zahrnuté do popisu celého systému: METAFONT pre tvorbu fontov a WEB, dokumentačný a vývojový „jazyk“ pre programovanie. T_EX a METAFONT sú písané vo WEBE.

V tomto texte popíšeme toto „jadro“ T_EXu z hľadiska užívateľa: na konci čítania by ste mali mať prehľad o súčiastiach systému T_EX a o súboroch a podporných programoch, ktoré sú pre Vás ako užívateľa podstatné. Text však nie je úvodom do schopností T_EXu alebo návodom, ako spustiť T_EX na počítači.

Budem používať „rohové“ poznámky na identifikáciu miest, kde budú po prvýkrát vysvetlené pojmy. Skratky pre typy súborov — obyčajne identifikované spoločnou príponou alebo rozšírením — budú vysádzané pomocou „strojového“ písma. Uvedomte si, prosím, že tieto skratky sa niekedy nestotožňujú s príponou súboru (pozrite tiež tabuľku).

Tento článok je úvodným v sérii článkov, ktoré budú popisovať hore uvedené subsystemy a ich komponenty. V tejto sérii sa každý článok sústreďuje na jeden subsystem z istého pohľadu. Jeho výsledkom by nemal byť gigantický popis, ktorý povie všetko (a teda nič). Podľa môjho názoru nasledujúce témy by mohli byť zaujímavé:

- štruktúra a štandardná inštalácia T_EXu

- dvi drajvery a fonty
 - možnosti implementácie grafiky do $\text{T}_{\text{E}}\text{X}$ ovských dokumentov
 - komponenty METAFONTu
 - štruktúra a štandardná inštalácia METAFONTu
 - systém WEB — koncept *gramotného programovania*
 - iné (aj keď ešte nenaplánované) témy prípadného záujmu
 - ▷ rozdiel medzi $\text{T}_{\text{E}}\text{X}$ om a DTP systémami
 - ▷ spôsob, ako pracuje $\text{T}_{\text{E}}\text{X}$ (na túto tému existuje niekoľko dobrých kníh)
 - ▷ hranice možností $\text{T}_{\text{E}}\text{X}$ u
 - ▷ $\text{T}_{\text{E}}\text{X}$ ako programovací jazyk
- Jednotlivé články budú publikované v uvedenom poradí.

Čo je to $\text{T}_{\text{E}}\text{X}$

$\text{T}_{\text{E}}\text{X}$ je sádzací systém s veľkými možnosťami značkovanie, pre sadzbu formul. Jeho základným princípom je, optické značkovanie, že štruktúry sú vyznačené v texte, a potom trans- logické značkovanie formované do sadzačského výstupu. Vyznačenie takejto informácie v štruktúre dokumentu sa nazýva *značkovanie*.¹⁾ Ak značky popisujú vzhľad dokumentu, tak sa nazývajú *optickými*, ak označujú štruktúru dokumentu, tak hovoríme o *logických značkách*. $\text{T}_{\text{E}}\text{X}$ podporuje obidva typy značiek. To znamená, exaktné riadenie vzhľadu častí dokumentu a ich umiestnenia, ako aj označkovanie štruktúry formul alebo zložiek dokumentu. Logické značky sú $\text{T}_{\text{E}}\text{X}$ om pretransformované do optických, takže vzhľad môže čitateľovi slúžiť na identifikáciu štruktúry. Vzhľad²⁾ a vo všeobecnosti knižný dizajn nepredstavujú neuzitočnú krásu. Dobrý knižný dizajn musí v prvom rade podporovať pochopenie obsahu v prospech čitateľnosti textu. Je *estetický* v jeho najlepšom slova zmysle, lebo spája formu a obsah a vytvára novú kvalitu.

Jadrom sádzacieho systému $\text{T}_{\text{E}}\text{X}$ je formátovací program $\text{T}_{\text{E}}\text{X}82$, ktorý sa často nazýva proste $\text{T}_{\text{E}}\text{X}$. My tu prijmeme túto zvyklosť, pokiaľ rozdiel medzi celým systémom a formátovačom nie je dôležitý alebo je zřejmý. $\text{T}_{\text{E}}\text{X}82$ je veľký monolitický program, ktorý je sádzacia mašina, $\text{T}_{\text{E}}\text{X}82$, balík makier

¹⁾ v anglickom origináli markup

²⁾ anglicky layout

publikovaný v knihe *TEX: The Program* od DONALDA KNUTHA. Jeho základné rysy môžu byť rozdelené do dvoch rovín:

- 1 TEX82 formátuje text, t.j. zalamuje odseky (včítane automatického rozdeľovania slov) a stránky.
- 2 Dáva k dispozícii programovací jazyk TEX, ktorý obsahuje mechanizmy pre tvorbu makier. To umožňuje zabudovanie nových príkazov na podporu značkovania na vyššej úrovni. DONALD KNUTH prezentuje jeden takýto príklad v TEXbooku: `plain TEX`. Súhrn makier, ktoré podporujú špecifický cieľ a majú (snáď) spoločnú filozofiu použitia, sa nazýva *balík makier*.

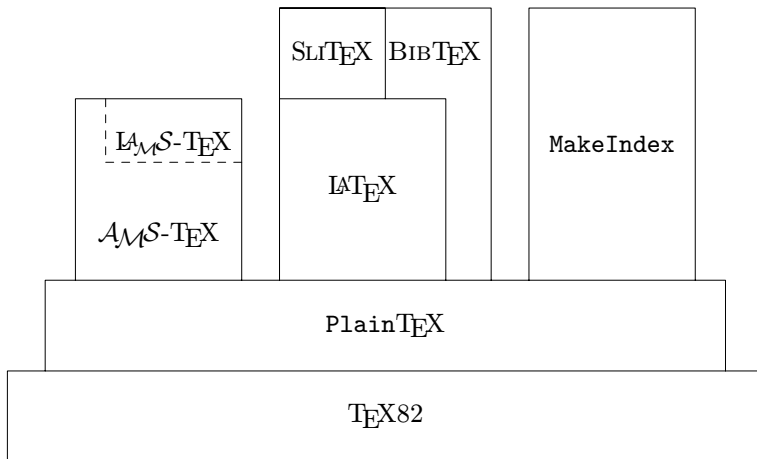
Rysy vyššej úrovne optického značkovania tak, ako sú reprezentované `plain TEX`om, umožňujú zabudovanie dodatočných úrovní pre úplné logické značkovanie. V súčasnosti sú široko rozšírené dva balíky makier pre logické značkovanie: `AMS-TEX` a `LATEX`. Obidva systémy sú viac alebo menej postavené na `plain TEX`u a užívateľ môže v prípade potreby k logickému označovaniu použiť aj optické z `plain TEX`u. To má za následok, že autor môže použiť zmes štruktúrnych informácií a explicitných informácií pre vzhľad — teda situáciu, ktorá pripúšťa veľkú škálu možností, ktoré môžu viesť (a aj vedú) k množstvu typografických nezmyslov.

Pretože TEX bol vytvorený len pre sadzbu textov, a pre podporu realizácie nových značkovacích štruktúr, chýba mu veľa spôsobilostí, ktoré požadujú autori. Na podporu ďalších možností, ako zostavovanie indexov, zoznamov literatúry alebo na vťahnutie grafiky, museli byť napísané dodatočné programy, ktoré využívajú informáciu získanú počas formátovania programom TEX82, spracujú ju a pripravujú pre následné spracovanie programom TEX82. Veľmi rozšírené a dostupné pre mnoho typov a kombinácií počítač/operačný systém sú dva doplnkové programy: `BIBTEX` pre tvorbu zoznamov literatúry z bibliografických súborov a ***MakeIndex*** pre zostavovanie indexu.

Špeciálnym prípadom spracovania informácie, ktorú poskytuje TEX počas svojho behu, je zostavovanie obsahu a využitie krížových referencií v texte. K tomu potrebujeme len informáciu o čísle strany, čísle časti atď. Tieto údaje poskytuje TEX82, môže ich aj sám spracovať, takže TEX82 sa používa v tejto situácii aj ako vlastný post-processor.

Videli sme, že sádzací systém TEX predstavuje súbor prostriedkov, ktoré pozostávajú zo sádzacej ‚mašiny‘ TEX82, balíkov makier (pri-

padne viacerých, ktoré využívajú ďalšie makrá) a doplnkových programov, ktoré sú používané spolu s týmito balíkmi makier. Tieto súvislosti sú ilustrované na obrázku 1.



Obr. 1: Komponenty $\text{T}_{\text{E}}\text{X}$ u

Formátovanie

Proces formátovania $\text{T}_{\text{E}}\text{X}$ om vyžaduje informáciu o rozmeroch znakov, ktorá je použitá pri zalamaní odstavcov. Súbor znakov je zoskupený do *fontov*. (Toto je isté zjednodušenie, pretože pojem ‚font‘ by mal byť používaný pre realizáciu istých znakov v pevnej veľkosti pre špecifikované výstupné zariadenie.) Rozmery znakov fontu sa nazývajú ‚metrika fontu‘.³⁾

Formát, v ktorom je použitá metrika fontu $\text{T}_{\text{E}}\text{X}$ om, bol definovaný DONALDOM KNUTHOM a nazýva sa TFM formát (*$\text{T}_{\text{E}}\text{X}$ font metrics*). V tomto formáte je každý znak popísaný ako priehradka s istou výškou, hĺbkou a šírkou. $\text{T}_{\text{E}}\text{X}$ potrebuje len tieto rozmery a nezaujíma ho tvar znaku. Je dokonca možné, aby znak presahoval túto priehradku, čo môže viesť k prekryvaniu s inými znakmi. Rozmery znaku sú špecifikované v rozme-

³⁾ anglicky *font metrics*

roch nezávislých od zariadenia,⁴⁾ pretože zalamovací algoritmus \TeX u je nezávislý od výstupného zariadenia.

Počas zalamovania odseku \TeX automaticky delí slová, a toto zalamovanie môže byť robené spôsobom, ktorý je skoro nezávislý od jazyka. Na jeho prispôbenie k rôznym jazykom sú potrebné *deliace vzory*,⁵⁾ ktoré parametrizujú deliaci algoritmus.

Výsledkom \TeX ovského formátovacieho behu je `dvi` drajver DVI dokument, v ktorom je špecifikovaný typ a pozícia na stránke každého výstupného znaku. Rozlišovacia schopnosť, ktorá je použitá, je tak malá, že každé možné výstupné zariadenie bude mať hrubší raster, takže umiestnenie je naozaj nezávislé od zariadenia. `dvi` dokument špecifikuje len znak, ale nie fonty samotné, takže názov `dvi` (,device independent‘) je presný. (Toto meno je problematické, lebo ,DVI‘ je teraz chránená značka spoločnosti Intel, ale názov `dvi` v \TeX u existoval skôr.) Na sprístupnenie výsledku formátovacieho procesu musí byť `dvi` súbor spracovaný tzv. `dvi` drajverom na požadovanom výstupnom zariadení.

Ak sa objavia problémy v procese formátovania, chybové hlásenia a varovania sa vypíšu na terminál. Každá správa, ktorá sa objaví na termináli, je tiež zapísaná do sprievodného protokolového súboru, ktorý sa nazýva LOG súbor. Do tohto LOG je možné uložiť aj ďalšie informácie, ktoré by boli príliš rozvláčne pre výstup na terminál. Ak sa tak stane, \TeX to oznámi užívateľovi na konci formátovania. Hlásenia \TeX u nie sú zabudované do programu, ale sú uložené v (refazcovom) POOL súbore. Tieto hlásenia musia byť načítané na začiatku formátovacieho procesu.

Balíky makier

Základný balík makier je `plain` \TeX , ktorý vyvinul DONALD KNUTH spolu s programom \TeX 82. Tento parametrizuje sádzacu mašinu \TeX 82, takže táto môže sádzať anglicky text v sade znakov písma Computer Modern. `plain` \TeX navyše poskytuje možnosti pre optické značkovanie. `plain` \TeX je dostupný ako jeden zdrojový súbor `plain.tex`.

⁴⁾ anglicky *device independent*

⁵⁾ anglicky *hyphenation patterns*

Všetky iné autorovi známe balíky makier sú založené na `plain TEX`u, t.j. obsahujú zdrojový súbor `plain.tex`, či už celý alebo s modifikáciami menej dôležitých častí. K `plain TEX`u najbližšie najdôležitejšie (voľne) prístupné balíky makier sú `AMS-TEX` od MICHAELA SPIVAKA a `LATEX` od LESLIE LAMPORTA. Ostatné voľné balíky makier majú často len lokálnu dôležitosť (napríklad `BlueTEX`, `TEXT1` alebo `TEXsis`), alebo sa používajú len vo veľmi špeciálnych prostrediach (napríklad `texinfo` v projekte GNU alebo `webmac` pre `WEB`). Dôležité komerčné balíky makier sú `MacroTEX` autora AMY HENDRICKSON a `LAMS-TEX`, ktorého autorom je tiež MICHAEL SPIVAK.

`AMS-TEX`,
`LATEX`,
`MacroTEX`,
`LAMS-TEX`

Tieto balíky makier obyčajne pozostávajú z jadra, ktoré poskytuje dodatočné primitívne pojmy pre značkovanie. Z takýchto primitívnych pojmov môžu byť zostavené štýly dokumentov,⁶⁾ ktoré realizujú logické značkovanie prostredníctvom príslušného vzhľadu. Tento vzhľad môže byť často upravovaný cez opcie *podštýlov* a štýlov, ktoré môžu poskytnúť ďalšie značkovanie.

Balíky makier vytvárajú doplnkové súbory, ktoré obsahujú informáciu o zalamovaní stránok alebo o značkovaní dokumentov. Tieto informácie môžu byť využité podpornými programami — napríklad špecifikácia citácie z bibliografickej databázy alebo špecifikácia položky indexu so zodpovedajúcim číslom strany pre zostavenie indexu. Špeciálnym prípadom je informácia o krížových referenciách alebo nadpisoch pre zostavenie obsahu, lebo táto informácia môže byť zoskupená a opätovne použitá priamo `TEX`om.

`SLITEX` je špeciálnou zložkou `LATEX`u, ktorá je určená na prípravu fólií s prekryvmi. V TUG-Boatu zväzok 10, číslo 3 (1989) bol ohlásený `LAMS-TEX`, ktorý realizuje možnosti `LATEX`u v `AMS-TEX`u. `MacroTEX` je súhrn makro ‚modulov‘, ktoré môžu byť použité na realizáciu nových značkovaní, ale nakoľko sa stal dostupným len nedávno, nie je ešte veľmi rozšírený.

Pri použití týchto (a ostatných) balíkov makier je potrebné sa presvedčiť, či nevyžadujú dodatočné fonty, ktoré nepatria do rodiny písma Computer Modern. Napríklad

⁶⁾ document styles

pre $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ sú potrebné fonty pre dodatočné symboly a neviditeľné znaky (pre prekryvy fólií), zatiaľ čo $\mathcal{A}\mathcal{M}\mathcal{S}\text{-T}_{\text{E}}\text{X}$ potrebuje niekoľko dodatočných sád fontov pre matematické znaky a azbuku.

Podporné programy

Tu pojednáme len o dvoch podporných programoch: $\text{BIB}\text{T}_{\text{E}}\text{X}$ e od ORENA PATASHNIKA pre prípravu bibliografií a *MakeIndex* od PEHONGA CHENA a MICHAELA HARRISONA pre prípravu a triedenie indexu. Pre obidve úlohy existujú iné, funkčne ekvivalentné podporné programy. Ale spomínané dva sú dostupné pre veľa operačných systémov a majú ‚oficiálny‘ status, pretože LESLIE LAMPORT povzbudzuje ich použitie s $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ om v jeho dokumentácii a TUG ich podporuje pre všeobecné použitie.

Neexistuje úplne prenosný mechanizmus pre zahrnutie všeobecnej grafiky do $\text{T}_{\text{E}}\text{X}$ ovských dokumentov, a preto nie je k dispozícii strojovo nezávislý podporný program.

$\text{BIB}\text{T}_{\text{E}}\text{X}$ sa používa na spracovanie literatúry zhrmaždenej v BIB súboroch. $\text{T}_{\text{E}}\text{X}$ vytvorí doplnkové súbory, ktoré obsahujú informáciu o požadovaných odkazoch na literatúru a $\text{BIB}\text{T}_{\text{E}}\text{X}$ vygeneruje z nich utriedený zoznam literatúry v BBL súbore, ktorý vzápätí môže byť použitý $\text{T}_{\text{E}}\text{X}$ om. Usporiadane a spôsob citovania sú určené *bibliografickým štýlom*, ktorý je špecifikovaný v BST súboroch. Hlásenia z behu $\text{BIB}\text{T}_{\text{E}}\text{X}$ u sú zapísané do BLG log-súboru.

MakeIndex načíta IDX podporný súbor, obsahuje položky indexu a príslušné čísla strán, usporiada a zjednotí ich a zapíše ich do IND súboru, ktorý je vstupným súborom pre $\text{T}_{\text{E}}\text{X}$. Štýl formátovania môže byť špecifikovaný indexovým štýlom. Hlásenia z behu *MakeIndex*u sú zapísané do ILG súboru.

Zlepšenia výkonu

Veľa práce, ktorú vykoná $\text{T}_{\text{E}}\text{X}$, je zhodnej pre každý dokument:

- 1 Celý text musí byť zalomený do riadkov. Slová častí textu v tom istom jazyku sú delené podľa tých istých vzorov.
- 2 Základné značkovanie príslušného balíka makier musí byť prístupné.

3 Požadované metriky fontov sú podobné pre veľa dokumentov, pretože použité sady fontov sa obyčajne nelíšia o veľa.

Kvôli zlepšeniu výkonu $\text{T}_{\text{E}}\text{X}$ u sú deliace vzory a popisy metrik fontov konvertované z externej reprezentácie do internej, ktorá je ľahko použiteľná programom $\text{T}_{\text{E}}\text{X}82$. V prípade (1) a (2) je externá reprezentácia textová. Je rozumné previesť túto transformáciu len raz a nie pre každý dokument. Interná reprezentácia je uložená v FMT súbore. Uloženie sa uskutočňuje $\text{T}_{\text{E}}\text{X}$ ovským príkazom `\dump`, a preto sa FMT súbory často nazývajú ‚dampované formáty‘.⁷⁾ FMT súbor môže byť načítaný na začiatku behu $\text{T}_{\text{E}}\text{X}82$ a tak je dostupný pre spracovanie aktuálneho textu.

Nakoľko FMT súbor sa vytvára nepravidelne — INIT E_{X} ,
obyčajne pre zaktualizovanie balíka makier — produkčná
formátovanie textu sa môže uskutočniť pomocou re- verzia
dukovanej verzie programu $\text{T}_{\text{E}}\text{X}82$, ktorá neobsahuje uloženie a časti programu pre transformovanie vzorov pre delenie a pre dampovanie. Úplná verzia programu $\text{T}_{\text{E}}\text{X}82$ je potrebná len v inicializačnej fáze, a preto sa volá INIT E_{X} . Dodatočné zlepšenia výkonu môžu byť dosiahnuté použitím produkčnej verzie programu $\text{T}_{\text{E}}\text{X}82$, z ktorého sú odstránené časti pre štatistickú analýzu a odladovanie.

Tie verzie $\text{T}_{\text{E}}\text{X}$ u, ktoré nemajú prednatiahnuté Vir $\text{T}_{\text{E}}\text{X}$
žiadne dampované formáty, majú schopnosť na-
tiahnuť dampovaný formát (t.j. FMT súbor) a nemajú schopnosť dam-
povať FMT súbor (t.j. nie sú INIT E_{X}), sa často nazývajú Vir $\text{T}_{\text{E}}\text{X}$, čo je
skratka z virgin $\text{T}_{\text{E}}\text{X}$.

TYP		PRÍPONA, ROZŠÍRENIE
SÚBORU	VYSVETLENIE	SÚBORU, ATĎ.
TEX	vstup textu	<code>tex, ltx</code>
DVI	výstup z $\text{T}_{\text{E}}\text{X}82$, formátovaný text	<code>dvi</code>
LOG	log súbor z $\text{T}_{\text{E}}\text{X}82$	<code>log, lis, list</code>
HYP	vzory delenia	<code>tex</code>
TFM	metriky fontov	<code>tfm</code>
POOL	reľazcový púl	<code>pool, poo, pol</code>
FMT	formátový súbor	<code>fmt</code>
MAC	súbor $\text{T}_{\text{E}}\text{X}$ makier	<code>tex, doc</code>
STY	štýlový súbor $\text{T}_{\text{E}}\text{X}$ u	<code>sty, tex, st, doc</code>

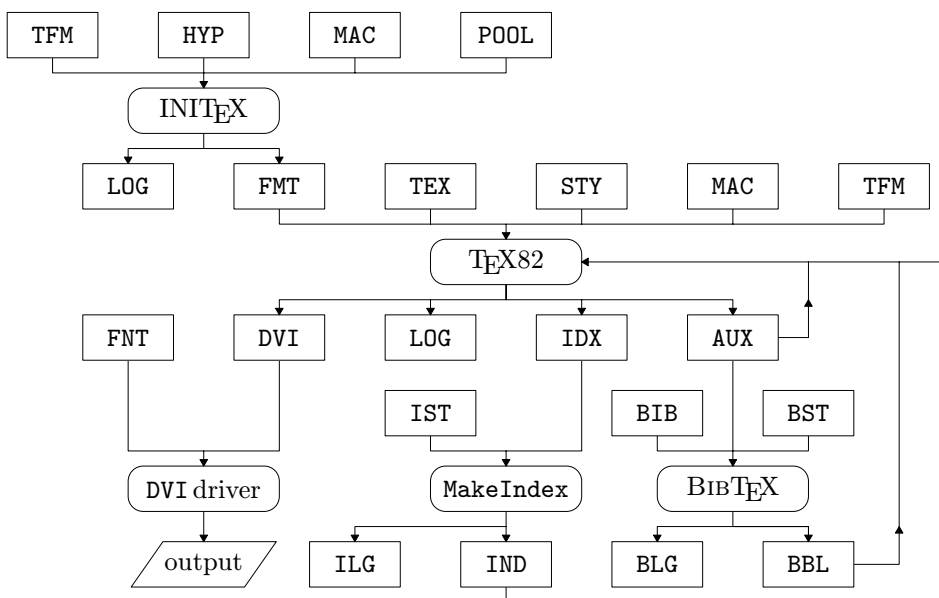
⁷⁾ anglicky dumped formats

AUX	podporný súbor	aux, toc, lot, lof glo, tmp, tex
BIB	bibliografický súbor	bib
BBL	literatúra alebo bibliografia	bbl
BLG	log súbor z BIB _T E _X u	blg
BST	štýlový súbor z BIB _T E _X u	bst
IDX	neutriedený index	idx
IND	utriedený index	ind
IST	špecifikácia značkovania indexu	
ILG	log súbor z <i>MakeIndexu</i>	ilg

Tab. 1: Typy súborov

Vzťahy medzi typmi súborov a komponentami

V predchádzajúcich častiach boli popísané komponenty systému $\text{T}_{\text{E}}\text{X}$ a spomenuté súbory, ktoré sú načítavané alebo písané týmito komponentami. Vzťahy medzi nimi sú graficky demonštrované na obrázku 2. Na



Obr. 2: Vzťahy medzi komponentami a typy súborov

tomto obrázku sú typy súborov reprezentované obdĺžnikmi a programy oválmi. Šípka znamená ‚je načítaný pomocou‘ alebo ‚vytvorený pomocou‘. Skratky pre typy súborov sú vysvetlené v tabuľke, ktorá tiež uvádza identifikáciu súborov (prípony alebo rozšírenia), ktorú tieto súbory používajú (ale všimnite si, že sa používajú aj iné identifikácie súborov).

Podakovanie.

Rád by som poďakoval CHRISTINE DETIGOVEJ, ktorá láskavo urobila anglický preklad. NELSON BEEBEOVÁ navrhla urobiť preklad. KLAUS GUNTERMANN mal hodnotné pripomienky k prvej (nemeckej) verzii. NICO POPPELIER prispel novou verziou obrázku 2, ktorá je lepšia než moja prvá. Publikovanie a distribúcia tohto článku je povolená len za podmienok analogických k GNU General Public Licence. Predchádzajúce revízie tohto článku boli publikované v ‚Die T_EXnische Komödie‘ a v ‚Baskerville‘. © Joachim Schrod

Z anglického originálu publikovaného v T_EXline č. 14
do slovenštiny prevedl Štefan Porubský

Tabuľky v plainT_EXu

JIRÍ VESELÝ

Často se na naši hlavu snášejí výtky, že opomíjíme zařazovat do T_EXbulletinu články pro ty, kteří s T_EXem začínají nebo nedávno začali. Jim je určen následující přehledný popis tvorby tabulek v PlainT_EXu. Jde spíše o seznam jednoduchých příkladů a ukázek než o teoretický výklad. Sazba tabulek v minulosti patřila v tiskárnách k nejnáročnějším operacím a ani při použití T_EXu není triviální. Je ale spíše pracná než složitá. Pokud se budete muset sazbou tabulek zabývat častěji, vyplatí se poohlédnout po nějakém makru, které některé věci zjednoduší. Začneme „od Adama“: použijme vstupní text

```
\halign{#&##&#\hfil\cr
Jméno      & Příjmení & Datum narození\cr
Adam       & Andrle   & 1.~1.~1911\cr
```