

Zpravodaj Československého sdružení uživatelů TeXu

Philip Taylor

Pragmatický přístup k odstavcom

Zpravodaj Československého sdružení uživatelů TeXu, Vol. 4 (1994), No. 3, 110–116

Persistent URL: <http://dml.cz/dmlcz/149715>

Terms of use:

© Československé sdružení uživatelů TeXu, 1994

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>

Vždy ma privádza do úžasu to, ako veľa užívateľov \TeX u (vrátane niektorých celkom vynikajúcich) je oboznámených s najtajnejšími oblasťami \TeX u, ale ak sú postavení pred úlohu, ktorá by mala byť najľahšou zo všetkých — presvedčiť daný odstavec, aby sa správne vysádzal bez vygenerovania správy `underfull \hbox` alebo zneužitia príkazu `\spaceskip` a jemu podobných príkazov — zdajú sa byť neschopní dosiahnuť niečo blízke vyhovujúcemu riešeniu. Vždy keď dostanem \TeX ovský dokument od takéhoto užívateľa a spustím ho, nájdem asi tucet výstražných hlásení, ktoré autor, zdá sa, celkom šťastne ignoroval, alebo ak je celý dokument vysádzaný so zväčšenými medzerami medzi slovami a bez rozdeľovania slov, pýtam sa sám seba, prečo sa táto zdánlivo jednoduchá úloha zdá byť pre mnohých taká ťažká.

Čestne povedané, časť výčítiek súvisí s Knuthom. Zatiaľ čo Boson SlowCoach a jeho súper sa zdajú byť šťastní, ak vytlačia odstavce s najdesivejšími medzerami medzi písmenami, alebo ak vytlačia riadok len s troma slovami a s obrovským medzislovným priestorom, Don sa rozhodol, že \TeX nebude robiť nič také: buď bude odstavec vytlačený správne, alebo nebude vytlačený vôbec! A samozrejme väčšina z nás, ak nie všetci, súhlasíme s Donom; kvôli čomu inému by sme sa mali vyhnúť zázrakom WYSIWYGu Bosonu kvôli kabalistickej zložitosti jazyka \TeX (pokiaľ samozrejme nie sme všetci intelektuálni masochisti, ako sa mi občas zdá). Ale časť výčítiek, aby sme ostali čestní, súvisí tiež s knihou *The \TeX book*; kým viac než primerane opisuje parametre, ktoré riadia \TeX ovskú sadzbu odstavcov, menej však informuje o metódach, pomocou ktorých môžeme získať vhodné hodnoty pre tieto parametre.

V tomto článku chcem ukázať to, čo nazývam „*pragmatický prístup k odstavcom*“, kým nejaký matematický génus nepríde so vzorcom alebo s algoritmom, ktorým môžu byť určené vhodné hodnoty týchto parametrov pre ľubovoľné kombinácie fontov, miery,¹⁾ odsadzovania, vzory

¹⁾ Používam pojem „miera“ v typografickom zmysle, majúci na mysli šírku tlačenej strany bez okrajov; pre viacstĺpcovú prácu to zodpovedá šírke jednotlivého stĺpca bez okrajov a medzier.

delení slov, atď., atď., atď., menej smrtiacim, ako keď budem pokračovať v riešení nezávideniahodnej úlohy tlačenia textu pre niektoré zdanlivo ľubovoľné kombinácie týchto premenných, a presvedčujúcim tých, pre ktorých tlačíme, že to *nepôjde*, kým to nebude spĺňať naše vlastné, do určitej miery bezchybné štandardy estetickej dokonalosti rovnako ako tie Donove a T_EXovské.

Musíme začať uvažovaním tých parametrov, ktoré (a) najviac ovplyvňujú správnosť vysádzania daného odstavca a (b) môžu byť logicky uvažované pod patronátom sadzača na rozdiel od tých, ktoré ovplyvňujú sadzbu, ale sú striktne pod kontrolou dizajnéra. Nasledujúca tabuľka, hoci nie úplná, uvádza niektoré dôležité parametre, ktoré spadajú pod tieto dve záhlavia

Vymedzené sadzačom	Určené dizajnérom
<code>\emergencystretch</code>	<code>\font</code>
<code>\pretolerance</code>	<code>\hsize</code>
<code>\tolerance</code>	<code>\patterns</code>
<code>\hbadness</code>	<code>\parindent</code>
<code>\hfuzz</code>	<code>\fontdimen <i>n</i></code>

Samozrejme, že dizajnérska sa bude pravdepodobne snažiť zadať horné ohraničenia na všetky vstupy v stĺpčeku sadzača, zatiaľ čo sadzačovi bude poradené, aby sa skutočne neplietol do vstupov v dizajnérovom stĺpci (ak on alebo ona chce byť vôbec niekedy znovu najatý!).

Prejdime k metóde, verím že bude jednoduchosť sama, hoci trochu zložitá na vysvetlenie:

1. Spracujte text s citlivými implicitnými hodnotami piatich „sadzačových parametrov“. Vyhovujúce implicitné (default) hodnoty môžu byť:

```

\emergencystretch = 0 pt
\pretolerance = 150
\tolerance = 250
\hbadness = 150
\hfuzz = 0 pt

```

Ak od T_EXu nedostaneme žiadne chybové hlásenie, je všetko v poriadku a úloha je skončená.

Ak nie, potom chybové hlásenia môžeme rozdeliť do dvoch skupín: tie, ktoré reprezentujú `overfull \hboxy` (preplnené rámčeky a teda

skutočné chyby) a tie, ktoré reprezentujú `underfull \hboxy` (nenaplnené rámčeky a sú to teda skôr upozornenia ako chyby). Z uvedených je potrebné sa zaoberať najskôr prvými.

2. Znovu spracujte text, ale teraz nastavte `\tolerance = 9999`; najjednoduchšie sa to dá urobiť vypustením príkazu pre `\tolerance` z preambuly a inicializovať ho priamo v príkazovom riadku, ako napr.:

```
TeX „\tolerance = 9999 \input text\
```

Znovu musíme chybové hlásenia rozdeliť na skutočné chyby a upozornenia, na rovnakom základe ako minule.

Ak sa už nevyskytli žiadne skutočné chyby, prejdite na krok 7; ak tu sú ešte stále skutočné chyby, tak text a formát môžu byť logicky považované za patologicky zlé a budú vyžadovať ďalšie akcie.

3. Overte veľkosť `overfull \hboxov`; ak nie sú väčšie, ako horné ohraničenia, určené dizajnérom na `\hfuzz`, nastavte `\hfuzz` na maximálnu hodnotu preplnenia a prejdite na krok 7.
4. *Ešte stále* sa vyskytujú `overfull \hboxy`. Prezrite log-súbor, zistíte riadok (riadky), v ktorých sa objavili a určte, či je problém v neadekvátnom delení (ten je väčšinou riešiteľný), alebo v širokej tabuľke, nerozdeliteľnom vzorci, a pod. Ak je problém spôsobený nesprávnym delením, prejdite na krok 6.
5. Problém je mimo rámca tohoto článku! Zvážte vytlačenie tabuľky, presahujúcej koniec riadku v menšom fonte, zmenšíte `\tabskip` a pod. Spýtajte sa autora, či nerozdeliteľný vzorec skutočne *nemôže byť* rozdelený. Použite svoju vlastnú heuristiku na riešenie úlohy, a keď prekonáte tieto ťažkosti, pokračujte v tejto procedúre od kroku 7.
6. Problém je spôsobený neadekvátnym delení slov. Zistite, či je potrebné rozšíriť zoznam v `\hyphenation`, alebo je potrebné pridať explicitné delenia do slova, presahujúceho riadok. Pridanie do zoznamu delení bude vhodné, ak v slove nie je nič nezvyčajné, ale vyskytlo sa málo deliacich bodov alebo nesprávne deliace body; explicitné delenie slova by sa vyžadovalo, ak slovo obsahuje určitý znak brániaci deleniu, napr. písmeno s akcentom, alebo ak pred slovom nie je medzera. Rozšírte delenia a pokračujte krokom 2.
7. Úspech! Už nemáte žiadne skutočné chyby. Teraz zostáva už len optimalizovať dokument tak, aby konečná verzia reprezentovala „najlepšiu možnú sadzbu“, v nejakom vágnom zmysle.

Všimnite si najhoršie prípady v `underfull \hboxoch`. Hodnotu `\tolerance` nastavte na túto hodnotu a `\hbadness` na hodnotu o 1 nižšiu a znovu spracujte text. Nemali by sa vyskytnúť žiadne skutočné

chyby, ani `\underfull \hboxy`, ktorých nekvalita (`badness`) prekračuje parameter `\tolerance`.

Teraz prichádza prevrapujúca časť. Mohli by sme si opodstatnene myslieť, že sme určili spodnú hranicu parametra `\tolerance`, ale mnohé dokumenty dokazujú, že to nie je tak. Nastavte `\tolerance` na hodnotu o 1 nižšiu ako v predchádzajúcom kroku a spracujte dokument znova. Ešte vždy *môže* byť vysádzaný korektne! Zmysel tohoto je sotva patrný, ale môže sa stať zrozumiteľným pri predstave, že tento T_EXovský koncept „ideálneho“ odstavca je taký, ktorý minimalizuje celkovú špatnosť (`badness`); T_EX *nemá* záujem o minimalizáciu špatnosti v nejakom jednotlivom riadku. Teda teraz môžeme mať odstavec, v ktorom sú dva alebo viaceré riadky „zlé“ v nejakom zmysle, zatiaľ čo špatnosť najhoršieho riadku bola zmenšená. Celkove je odstavec „horší“, ale z estetického hľadiska sa môže javiť ako „lepší“ (odstavec, pozostávajúci výhradne z kratších riadkov môže vyzeráť lepšie ako odstavec, v ktorom vytŕča jeden extrémne kratší riadok).

Ak je odstavec korektne vytlačený s hodnotou `\tolerance` o jedna nižšou, ako je zrejma dolná hranica, môžeme to rovnako dobre zopakovať. Vždy, keď je odstavec vytlačený bez chýb, nastavte `\tolerance` na hodnotu o jedna menšiu, ako je ohlásená najhoršia špatnosť, a opakujte postup. Môže sa stať, že ďalšie zmenšenie v `\tolerance` nie je možné a objaví sa `\overfull \hbox`; nastavte predchádzajúcu hodnotu a skončíte. Globálna optimalizácia je teraz skončená — určili sme optimálnu hodnotu `\tolerance`. Ak je menšia alebo rovná ako hodnota, ktorú uviedol dizajnér ako hornú hranicu, naša práca je spravená; ak nie, musíme vyvolať `\emergencystretch` a potom, možno, uskutočniť lokálnu optimalizáciu.

8. Zmenšíte `\tolerance` na hodnotu dolnej hranice určenej dizajnérom a nastavte `\emergencystretch` na malú kladnú veľkosť. Správanie parametra `\emergencystretch` a konkrétne jeho interakcia s ďalšími uvažovanými parametrami sú slabo pochopené a v skutočnosti sú predmetom výskumu v projekte L^AT_EX3. Napriek tomu zdravé pravidlo, založené na skúsenostiach, je nastaviť ho na 1 em (založené na základnom textovom fonte); táto hodnota môže byť, hoci sa to môže zdať zvláštne, rovnako vyhovujúca pre obe — širokú aj úzku — miery.

Nastavte `\hbadness` na hodnotu o jedna menšiu, ako je `\tolerance`, a spracujte text znova. Ak boli hlásené `\overfull` boxy, tak máme problém: *môžeme* zväčšiť hodnotu `\emergencystretch`, ale toto náhle vedie k množstvu nenaplnených rámečkov a odstrašujúcej estetike; za

týchto okolností je pravdepodobne lepšie uskutočniť lokálnu optimalizáciu, čo je vlastne ďalší krok.

Pre každý riadok, hlásený ako preplnený alebo nenaplnený, uvážme odpovedajúci odstavec a zistíme, či ďalšie delenie slov môže umožniť T_EXu zbalíť pár prebytočných medzier alebo niekoľko znakov v riadku; tiež zvážte, či jeden alebo viaceré parametre, uvedené v Dodatku A (napr. `\doublehyphendemerits`, `\exhyphenpenalty`) môže prinútiť T_EX uspokojiť sa so sadzbou tohoto odstavca, ktorá by nebola perfektná. Ak to je nevyhnutné, zvážte zmenu jedného alebo viacerých týchto parametrov len počas trvania odstavca tak, že ho uzavriete medzi dvojicu `\begingroup`/`\endgroup` a zmeníte parametre hneď za `\begingroup`. Pamätajte si, že odstavec musíte ukončiť príkazom `\par` alebo prázdny riadok *pred* zátvorkou `\endgroup`, ak majú zmeny parametrov priniesť nejaký efekt.

Pokračujte v lokálnej optimalizácii bez úprav parametrov `\tolerance` alebo `\hbadness`, až kým ďalšie zlepšenie už nebude môcť byť dosiahnuté; ak ostávajú nenaplnené rámčeky, je potrebné postupovať taktne: prezrite chybný odstavec (odstavce) a pozvite autora poopraviť ich len v prípade, ak na to oprávňujú estetické dôvody.

Pamätajte, že parameter `\tolerance` bol zmenšený zo svojej „optimálnej“ hodnoty na dizajnérovu dolnú hranicu; ak autor odmieta chybný odstavec prepracovať, uzavrite odstavec medzi dvojicu `\begingroup`/`\endgroup` rovnako ako vyššie a len na trvanie tohoto odstavca parameter `\tolerance` nastavte na jeho „optimálnu“ hodnotu (a oznámte dizajnérovi, že to bolo nevyhnutné).

Teraz je lokálna optimalizácia skončená. Zmeňte preambulu, aby ste začlenili experimentálne určené hodnoty pre `\tolerance` a `\hbadness`. Vyššie sme nezdôvodnili, prečo je potrebné nastaviť `\hbadness` na hodnotu o jedna nižšiu ako `\tolerance`: je to jednoducho preto, že ak ho nastavíme na doporučenú hodnotu, je možné preveriť, že experimentálne určená hodnota `\tolerance` bola v skutočnosti nevyhnutná, zatiaľ čo ostatné varovania T_EXu potláčame; ak T_EX zaradí do správy `underfull \hbox` so špatnosťou rovnou `\tolerance`, bola spravená nejaká chyba. Samozrejme, pre použitie musí byť `\hbadness` nastavené rovnaké ako `\tolerance`, aby boli eliminované falošné varovné oznamy.

Tento postup sa môže zdať zložitý, ale v skutočnosti je veľmi priamočiarý a je určite intuitívny, ako raz plne oceníme jemnosť opakovaných zmenšení `\tolerance`. Zahrnuté kroky si vyžadujú úplné minimum edi-

ovania, maximálne možno využiť toho, že správanie T_EXu môže byť ovplyvnené parametrami príkazového riadku. Postup bol dôsledne použitý pri písaní tohoto²⁾ článku (a mnohých predchádzajúcich) a je pravidelne dennodenne používaný na mojej fakulte, kde sa produktivita cení dokonca viac ako estetická dokonalosť.

Philip Taylor
The Computer Centre, RHBNC,
University of London, U.K.
<P.Taylor@Vax.Rhbncc.Ac.Uk>

Dodatok A:

Prehľad parametrov súvisiacich s odstavcom

Celočíselné parametre:

- \adjdemerits
- \doublehyphendemerits
- \exhyphenpenalty
- \finalhyphendemerits
- \hbadness
- \hyphenpenalty
- \linepenalty
- \looseness
- \pretolerance
- \tolerance

Rozmerové parametre:

- \emergencystretch
- \hangindent
- \hfuzz
- \hsize
- \parindent

Glejové parametre:

- \leftskip
- \parfillskip
- \rightskip
- \spaceskip
- \xspaceskip

Rôzne parametre:

- \fontdimen 2

²⁾ Pri písaní originálu tohoto článku. (Pozn. prekl.)

```
\fontdimen 3
\fontdimen 4
\fontdimen 7
\parshape
```

Ghostscript versus ps2pk

KAREL HORÁK

V poslední době se stále častěji setkáváme s PostScriptem, a to i na stránkách tohoto občasníku. Minulé číslo našeho zpravodaje se ho dotklo hned třemi příspěvky a mně už před časem napadlo porovnat (nemáje přístup na žádnou obyčejnou PostScriptovou tiskárnu), jak se se zobrazením PostScriptových písmen vyrovnají běžně dostupné emulátory PostScriptu.

V horní části prvního obrázku vidíte v prvním sloupci TimesRoman tištěný z .pk souborů získaných programem ps2pk 1.4 z .pfb souboru ze sady písem pro Corel! 3.0, a to postupně ve velikosti 24.88 pt, 10 pt a 5 pt. Druhý sloupec pro srovnání obsahuje stejné písmo z české verze Adobe Type Manageru zpracované stejným způsobem.

Spodní část obrázku obsahuje stejnou ukázkou zpracovanou verzí Ghostscriptu 3.12.

TimesRoman

TimesRoman TimesRoman

TimesRoman

TimesRoman TimesRoman

TimesRoman

TimesRoman TimesRoman

TimesRoman

TimesRoman TimesRoman

Je jasné, že čím větší rozlišení, tím kvalitnější je kresba bez ohledu na způsob emulace PostScriptu (toto číslo je tištěno z předloh v rozlišení 600 dpi). Proto nám lepší obrázek poskytnou čtyřnásobně zvětšené bitové mapy pětibodových písmenek, jak je vidíte v další tabulce.