

AUTOMATIZÁCIA TVORBY TABULIEK Z DÁT ULOŽENÝCH VO FORMÁTE .csv

ALEŠ KOZUBÍK (SK)

Abstrakt. Tabulky sú súčasťou väčšiny odborných alebo vedeckých textov. Obvykle bývajú spracovávané v niektorom z tabulkových procesorov. Ich manuálny prenos do cieľového dokumentu tak býva zdrojom mnohých chýb. Príspevok predstavuje balíček `csvsimple`, ktorý umožňuje určitú automatizáciu pre sadzbu tabuliek z dát, exportovaných do súboru vo formáte `csv`.

Kľúčové slová. sadzba, tabulky, `.csv` formát, balíček `csvsimple`.

TYPESSETTING AUTOMATION OF THE THE TABLES CONTAINING DATA STORED IN THE .CSV FORMAT

Abstract. The tables are part of the most scientific texts. Their content is usually processed in some of the spreadsheets. Their manual transmission in the target document is thus the source of many mistakes. The article represents the `csvsimple` package that allows some automation for table typesetting from data exported to a `csv` format.

Keywords. typesetting, tables, `.csv` format, `csvsimple` package.

Úvod

Prakticky žiadny odborný alebo vedecký text sa nezaobíde bez sprievodného tabulkového materiálu. Sadzba tabuliek je pri tom jednou z najnáročnejších súčastí sadzby textu. Je to zapríčinené nielen pomerne náročným spôsobom formátovania tabuliek, ale často aj samotným procesom zostavenia tabulky resp. prenosu jej obsahu do cieľového dokumentu. Práve tento proces býva zdrojom najpočetnejších a pri tom obtiažne odhaliteľných chýb.

Samotné tabulky, najmä ak údaje v nich obsiahnuté sú predmetom či výsledkom rozmanitých výpočtov, je potrebné manuálne opisovať do cieľového dokumentu. Neexistuje totiž nijaká možnosť automatizovaného exportu údajov z tabuliek, či už z tabulkových procesorov, alebo webových zdrojov priamo do formátu typografického systému $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ a ich následného pohodlného vloženia do cieľového dokumentu.

So samotnou tvorbou tabuliek v typografickom systéme $\text{T}_{\text{E}}\text{X}/\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ je možné sa stručne zoznámiť v článku [2] z predchádzajúcich ročníkov našich konferencií. Podrobnejší prehľad sadzby tabuliek je možné nájsť v knihách [4], alebo [3].

V tomto príspevku predstavíme balíček `csvsimple`, ktorý umožňuje určité zjednodušenie a zmechanizovanie prenosu dát do tabuľky z dátových súborov vo formáte `.csv`. Nakoľko ide o pomerne frekventovaný formát ukladania dát a väčšina tabuľkových procesorov disponuje možnosťou exportu dát do tohto formátu, tento balíček predstavuje veľmi užitočnú pomôcku s potenciálom výraznej redukcie chýb pri prepise.

1. Predstavenie balíčka `csvsimple`.

Balíček `csvsimple` je pomerne mladým balíčkom (k dispozícii je od 1.7.2016), ktorého autorom je Thomas F. Sturm. Tento balíček poskytuje jednoduché rozhranie systému `LATEX`, pre spracovanie `.csv` súborov. Balíček silne vychádza zo syntaxe `pgfkeys`, čoho dôsledkom je jeho pomerne jednoduché použitie. Osobitne treba uviesť podporu filtrovania dát a generovania tabuľky. Nakoľko balíček je stavaný ako ľahký nástroj, neposkytuje podporu pre triedenie dát a ukladanie údajových databáz.

Balíček sa do zdrojového súboru vkladá štandardným spôsobom, teda rovnako ako všetky ostatné balíčky príkazom:

```
\usepackage{csvsimple}
```

Zároveň s balíčkom `cvssimple` sú automaticky nahrávané aj balíčky `pgfkeys`, `etoolbox` a `ifthen`. Pri práci sú často potrebné aj balíčky `longtable` a `booktabs`, ktoré však automaticky nahrávané nie sú a je potrebné doplniť ich manuálne.

1.1. Základy práce

Ako sme už povedali, balíček spracúva súbory vo formáte `.csv`, teda údaje oddelené čiarkami. Pre správnu činnosť je potrebné zabezpečiť, aby každý riadok `csv` súboru obsahoval rovnaký počet údajov oddelených čiarkou. Pri tom je možné, aby boli pomocou skupinových `TEX`-ových zátvoriek `{}` vyznačené bloky obsahujúce čiarky, ktoré nemajú byť spracované ako separátory. To môže byť užitočné, pri spracovaní číselných údajov, kedy podľa zvyklostí našej notácie používame desatinné čiarky, narozdiel od anglo-saskej notácie pomocou desatinnej bodky. Tento „zátvorkový diskonfort“ je tiež možné odstrániť voľbou iného oddeľovača, čo väčšina tabuľkových procesorov umožňuje. Ako uvidíme neskôr, aj balíček `csvsimple` umožňuje nastavenie oddeľovačov na iný znak.

Štruktúra súboru `.csv`, je obvykle, aj keď nie vždy, taká, že prvý tzv. hlavičkový riadok obsahuje názvy jednotlivých stĺpcov. Ilustruje to nasledujúci výpis súboru `loko.csv`, ktorý používame na ilustráciu činnosti balíčka. Jeho obsahom sú vybrané údaje o niekoľkých parných lokomotívach, zosumarizované podľa [5].

```

Číslo,Prezývka,Hmotnosť,Výkon
387.0,Mikádo,93 ton,1546 kW
498.1,Albatros,113 ton,1950 kW
556.0,Štokr,99 ton,1472 kW
464.2,Rosnička,112 ton,1250 kW

```

Z týchto údajov je možné automatizovane zostaviť tabuľku pomocou jednoduchého príkazu:

```
\csvautotabular{loko.csv}
```

Odkaz `loko.csv` je samozrejme možné nahradiť vlastným názvom príslušného csv súboru, z ktorého má byť tabuľka vygenerovaná. Výsledkom je potom tabuľka 1.

Číslo	Prezývka	Hmotnosť	Výkon
387.0	Mikádo	93 ton	1546 kW
498.1	Albatros	113 ton	1950 kW
556.0	Štokr	99 ton	1472 kW
464.2	Rosnička	112 ton	1250 kW

Tabuľka 1. Základná tabuľka vytvorená pomocou balíčka `csvsimple`.

Číslo	Prezývka	Hmotnosť	Výkon
387.0	Mikádo	93 ton	1546 kW
498.1	Albatros	113 ton	1950 kW
556.0	Štokr	99 ton	1472 kW
464.2	Rosnička	112 ton	1250 kW

Tabuľka 2. Základná tabuľka vytvorená v spolupráci s balíčkom `booktabs`.

Podobne je možné zautomatizovať aj spoluprácu s balíčkom `booktabs`. Použijeme však mierne modifikovaný príkaz

```
\csvautobooktabular{loko.csv}
```

Výsledok vidíme v tabuľke 2.

1.2. Riadenie vzhľadu tabuľky

V predchádzajúcich jednoduchých ukážkach bolo formátovanie tabuliek plne v kompetencii balíčka `csvsimple`. Ak chceme prevziať plnú kontrolu nad formátovaním tabuľky, použijeme príkaz `\csvreader`. Nasledujúca ukážka ilustruje, ako sa položky z hlavičkového riadku csv súboru automaticky priradia \TeX -ovým makrám, ktoré môžeme následne voľne použiť. Dôležitá je voľba `[head to column names]` príkazu `\csvreader`. Taktiež treba upozorniť, že bola potrebná určitá úprava csv súboru na súbor `loko1.csv`, v ktorom boli odstránené diakritické znaky, ktoré ako pomenovania makier spôsobujú problémy. Príslušný zdrojový kód potom vyzerá takto:

```

\begin{tabular}{|l|l|l|l|}\hline%
\bfseries Číslo & \bfseries Prezývka & \bfseries Hmotnosť &
\bfseries Výkon
\csvreader[head to column names]{loko1.csv}{%

```

```
{\\cislo & \meno & \vaha & \vykon}%
\\hline
\end{tabular}
```

a zodpovedajúci výstup vidíme v tabuľke 3.

Číslo	Prezývka	Hmotnosť	Výkon
387.0	Mikádo	93 ton	1546 kW
498.1	Albatros	113 ton	1950 kW
556.0	Štokr	99 ton	1472 kW
464.2	Rosnička	112 ton	1250 kW

Tabuľka 3. Riadenie vzhľadu tabuľky pri použití príkazu `\csvreader`.

Číslo	Prezývka	Hmotnosť	Výkon
387.0	Mikádo	93 ton	1546 kW
498.1	Albatros	113 ton	1950 kW
556.0	Štokr	99 ton	1472 kW
464.2	Rosnička	112 ton	1250 kW

Tabuľka 4. Zmena vzhľadu tabuľky pri použití príkazu `\csvreader`.

Neprijemnosť s názvami makier je možné obísť úpravou parametrov príkazu `\csvreader`. Ak chceme potlačiť automatické pomenovanie makier a použiť svoje vlastné mená, použijeme príkaz v takejto podobe:

```
\csvreader[late after line=\\hline]{loko.csv}{Číslo=\cislo,
Prezývka=\meno,Hmotnosť=\vaha,Výkon=\vykon}
```

Volba `[late after line=\\hline]` potom určuje, že jednotlivé riadky tabuľky budú oddelené horizontálnou čiarou. Výsledok je možné pozorovať v tabuľke 4

O niečo pohodlnejšie formátovanie je možné pomocou definovania vhodných voliteľných klúčov. To umožňuje definovanie vytvorenie `pgfkeys` štýlu pre celú tabuľku. V našom prípade môžeme príkaz `\csvread` modifikovať napríklad takto:

```
\csvreader[tabular=|l|r|l|,
table head=\hline \bfseries Lokomotíva&\bfseries Hmotnosť&
\bfseries Výkon\\hline\hline,
table foot =\hline]%
{loko.csv}{Číslo=\cislo,Prezývka=\meno,Hmotnosť=\vaha,
Výkon=\vykon}%
{\cislo~\meno & \vaha & \vykon}%
```

Výsledok nášho úsilia potom vidíme v tabuľke 5.

Lokomotíva	Hmotnosť	Výkon
387.0 Mikádo	93 ton	1546 kW
498.1 Albatros	113 ton	1950 kW
556.0 Štokr	99 ton	1472 kW
464.2 Rosnička	112 ton	1250 kW

Tabuľka 5. Zmena vzhľadu tabuľky pri použití voliteľných parametrov príkazu `\csvreader`.

Číslo	Prezývka	Hmotnosť	Výkon
387.0	Mikádo	93 ton	1546 kW
498.1	Albatros	113 ton	1950 kW
556.0	Štokr	99 ton	1472 kW
464.2	Rosnička	112 ton	1250 kW

Tabuľka 6. Tabuľka so vzhľadom riadeným pomocou štýlu definovaného prostredníctvom `\csvstyle`.

2. Pokročilejšie techniky

2.1. Vlastný štýl tabuľky

Nasledujúca ukážka ilustruje, ako si pomocou makra `\csvstyle` môžeme definovať štýl tabuľky. Tu opäť použijeme automatizované pomenovanie stĺpcov tabuľky s využitím voľby `head to column names`. Pri tomto prístupe však položky záhlavia nesmú obsahovať medzery, ani špeciálne znaky. Preto aj v ukážke používame opäť súbor `loko1.csv` s odstránenou diakritikou. Nasledujúci zdrojový kód ilustruje definíciu štýlu tabuľky:

```
\csvstyle{mojaTabulka}{tabular=|c|l|r|l|,
table head=\hline \bfseries Číslo& \bfseries Prezývka &
\bfseries Hmotnosť& \bfseries Výkon\\\hline\hline,
table foot=\hline,
head to column names}
```

Samotnú tabuľku potom vygenerujeme jednoducho pomocou príkazu `\csvread`, tak ako to vidíme v nasledujúcej ukážke:

```
\csvreader[mojaTabulka]{loko1.csv}{}%
{\cislo & \meno & \vaha & \vykon}
```

Výsledkom je potom tabuľka 6.

Syntax `pgfkeys` umožňuje vytvárať vlastné štýly pre opakované použitie a dosiahnutie konzistentného a jednotného vzhľadu. Nasledujúci príklad je upravený tak, aby sa získalo viac alebo menej možností nastavení. Nový štýl, pomenovaný `mojeVlaky`, definujeme pomocou nasledujúcej postupnosti príkazov:

```
\csvset{mojeVlaky/.style={%
tabular=|r|l|r|,
table head=\hline \bfseries Číslo&\bfseries Lokomotíva &
\bfseries #1\\\hline\hline,
late after line=\\\hline,
column names={cislo=\cislo,meno=\meno}
}}
```

Takto definovaný štýl potom môžeme použiť na sadzbu dvoch tabuliek 7 a 8 pomocou príkazov

```
\csvset{mojeVlaky/.style={%
tabular=|r|l|r|,
table head=\hline \bfseries Číslo&\bfseries Lokomotíva &
\bfseries #1\\\hline\hline,
late after line=\\\hline,
column names={cislo=\cislo,meno=\meno}
}}
```

resp.

```
\csvreader[mojeVlaky={Váha}]{loko1.csv}{vaha=\vaha}%
{\cislo & \meno & \vaha}%
```

```
\hfill%
\csvreader[mojeVlaky={Výkon}]{loko1.csv}{vykon=\vykon}%
{\thecsvrow & \cislo~\meno & \vykon}%
```

Číslo	Lokomotíva	Váha
387.0	Mikádo	93 ton
498.1	Albatros	113 ton
556.0	Štokr	99 ton
464.2	Rosnička	112 ton

Tabuľka 7. Tabuľka so vzhľadom definovaným štýlom `mojeVlaky`.

Číslo	Lokomotíva	Výkon
1	387.0 Mikádo	1546 kW
2	498.1 Albatros	1950 kW
3	556.0 Štokr	1472 kW
4	464.2 Rosnička	1250 kW

Tabuľka 8. Pozmenená tabuľka so vzhľadom definovaným štýlom `mojeVlaky`.

3. Záver

V článku boli ilustrované možnosti automatickej tvorby tabuliek z dátových súborov uložených vo formáte `csv`. Základné postupy boli ilustrované na niekoľkých jednoduchých príkladoch tabuliek.

Poďakovanie. Tento príspevok vznikol s príspevom grantu KEGA-041ŽU-4/2017 podporovaného Slovenskou kultúrno-edukačnou grantovou agentúrou.

Literatúra

- [1] BLAŠKO, R.: *L^AT_EX nie je farba na maľovanie*, Otvorený softvér vo vzdelávaní, výskume a IT riešeníach, zborník medzinárodnej konferencie OSSConf 2010, Žilina, 1.–4. júla 2010, str. 43–52, ISBN 978-80-970457-0-8, <http://sospreskoly.org/files/OSSConf2010/ossconf2010-Blasko.pdf>.
- [2] BLAŠKO, R.: *L^AT_EX nie je farba na maľovanie, ale na písanie*, Otvorený softvér vo vzdelávaní, výskume a IT riešeníach, zborník medzinárodnej konferencie OSSConf 2010, Žilina, 1.–4. júla 2011, str. 43–52, ISBN 978-80-970457-0-8, http://ossconf.soit.sk/static/contributions2011/Blasko_farba.pdf.
- [3] KOPKA, H. – DALY, P. W.: *L^AT_EX – Podrobný príručce*, Brno, Computer Press, 2004, ISBN 80-722-6973-9.
- [4] RYBIČKA, J.: *L^AT_EX pro začátečníky*, Brno, KONVOJ 2003, ISBN 80-7302-049-1.
- [5] SCHRÖTTER, J. – FULTNER, B.: *Svět lokomotiv*, Brno, CPress, 2016, ISBN 978-80-264-1268-7.

Kontaktná adresa

RNDr. Aleš Kozubík, PhD., Katedra matematických metód, Fakulta riadenia a informatiky, Žilinská univerzita, Univerzitná 8215/1, 010 26 Žilina, Slovenská Republika,
E-mailová adresa: alesko@frcatel.fri.uniza.sk