Guide TeX It: Uneasy Beginnings of Typesetters from the Perspective of Non-Typesetters

Nech TeX nás být: pohled na krušné začátky sazečské z pohledu nesazečů

Libor Sarga

Abstract: The article describes the process of typesetting a proceedings in TeX from the perspective of prospective typesetters along with challenges and obstacles encountered and solved during the work. Focused on the problems of generating a desired Table of Contents and captions of graphic objects, it further lists minor annoyances and tricks used to solve them. Also described is a field-proven electronic content management and synchronization system for different file versions utilized while working on the project in a decentralized fashion.

Key words: typesetting, TeX, tocloft, caption, tables, figures

Abstrakt: Článek popisuje tvorbu sborníku příspěvků v systému LaTeX z pohledu amatérských sazečů spolu s problémy a těžkostmi, které byly při této práci zaznamenány a vyřešeny. Zvláštní pozornost je věnována peripetiím při tvorbě obsahu a popiskům grafických objektů. Nastíněny jsou také drobné problémy a triky, které je vyřešily. Rovněž je popsán v praxi vyzkoušený systém správy a synchronizace verzí souborů, použitý při decentralizované manipulaci s hlavní šablonou sborníku.

Klíčová slova: sazba, TeX, tocloft, popisky, tabulky, obrázky

Introduction

Before we delve into how we end up typesetting a proceedings and what we went through while doing so, it is necessary to flesh out my prior experience with TeX as this serves an important point of providing a background. Albeit short, the section is intended to provide information regarding my initial level of knowledge.

In September 2010 I began my first year as a doctoral student at the Department of Statistics and Quantitative Methods, Faculty of Management and Economics, Tomas Bata University in Zlín. I was aware even then that for producing quality scientific output it would be viable to choose a more sophisticated

157

system than widely-available commercial software residing at the university's computers. I started looking and very soon bumped into TeX. I downloaded *The Not So Short Introduction to LaTeX 2_ε*[1] and promised myself to at least peek at it in the near future.

This I did and thus came to know things such as preamble, what to write in there, how to load packages and lots of other useful bits and pieces of what constitutes the system, which I had unfortunately no means of practically testing due to lack of infrastructure. That changed in early October when necessary equipment, including a fresh install of TeX Live 2010 became a reality. I was pleasantly surprised that Lubor Homolka, another first-year doctoral worker at the department, and Pavel Stříž were both into TeX as well and have often provided me with numerous theoretical and practical examples. I was then and am still now nothing but a beginner who decided to slowly familiarize himself with the 'tex' (phonetic transcript of my pronunciation at that time) using TeXMaker editor. When I compiled my first document, I was elated:

```
\documentclass[a4paper]{article}
\begin{document}
Hello world!
\end{document}
```

I was pleased with myself, experimented as instructions told me, loaded packages, inserted pictures, switched languages. Lubor entered the realm of math typesetting. I was looking, jealous, as he compiled his dozens of lines long documents without any error while I explored aesthetically pleasing red messages the editor threw at me with steel precision and determination. Today as well as then I'm waiting, palpably discomforted, if the compilation, starting with innocuous

```
Process started
```

will end in victory

```
Process exited normally.
```

or if I get a dreaded

```
Process exited with error(s).
```

We saw our part of the last message at the time of typesetting, often more so than our families, friends, or beer kegs.

Proceedings

A relatively peaceful situation quickly gave way to something else after an e-mail came from Pavel Stříž, asking whether we would be interested in typesetting a proceedings in care of his brother's printing house[2]. This sign of trust had me

---

[1] http://www.ctan.org/tex-archive/info/lshort/english/lshort.pdf
[2] http://www.striz.cz/

158

excited: finally I will be able to show my extensive knowledge of PDF crafting, hampered by a mistake here and there, to the whole wide world. However, I was not completely able to filter out a notion of haste in the e-mail's text. Evidence, such as 'urgent work' and 'two or three weeks tops' left no doubt the process need have to be a quick one. We were further told the proceedings comprises of approximately 250 pages, 60 figures and 50 tables. My inner voice, warning me of possible repercussions, was quickly hushed and the deal was struck shortly before the middle of October. Pavel Stříž promised us backing in case we needed a hand which we decided to use only if no other option was evident to us. Out of the sample we were given the main issues were identified to be with the tables which would probably be necessary to convert into a unified format; and pictures, requiring conversion to PDF (we were yet to find out how).

On October 15, 2010 the proceedings in the `.doc` file was acquired. At 3.822 MB it contained all the articles in English along with figures and tables. We began to work. Since that day a frequently used communication channel was established between our side and Pavel Stříž's brother, Martin. He provided us with multitude of advice on basic configuration, a type of binding to be used and other mantras we have never previously heard of (we couldn't have since we had not been in the printing business, we concluded). The book's format was to be A5 (148×210 mm) as reflected in the settings of a template. We also had to admit weakness and reject making the cover using PSTricks, a task too complex for either of us.

The first thing was to change the PDF viewer since Acrobat Reader 9 was not suitable for our purposes. Every PDF file the program opens is prevented from being changed, including with the editor commands. I found out later this is due to it not supporting DDE (Dynamic Data Exhange), an omission in many other viewers as well (Foxit PDF Reader). A solution is to create a simple batch file containing instructions to first terminate the process and then re-open it (a type of indirect data exchange) and link to this script instead to the viewer's absolute path. We did not know it then and decided to use Sumatra PDF that filled the gap.

Then, the settings of the editors were synchronized, especially font encoding (UTF-8) because Lubor decided against using commercial operating systems in favor of Linux, and these two different environments could cause problems. Also without UTF-8 in place comments in Czech became a gibberish of question marks and other exotic characters not put there previously by either of us even on the same versions of Windows. The `inputenc` package with `[utf8]` options provided a solution. Cropping the images was to be done by using a licensed version of Adobe Acrobat 8 Professional, listing the functionality in its features. An alternative is to use an internal command `viewport=` with the parameters being coordinates, allowing for automation of the process. We did not go this way as the thought of such declarations quickly sapped the color out of our cheeks.

The first step of image extraction was to convert the document into PDF, then using an open source tool PDF Split and Merge (`http://www.pdfsam.org/`) to extract the pages with the figures and finally unleash Adobe Acrobat Professional to crop them to external files. An alternative to the first step is to use pdftk (`http://www.pdflabs.com/tools/pdftk-the-pdf-toolkit/`). This ensured scalability unachievable by formats such as JPEG and PNG as per our own experience. Moreover, JPEG format is lossy by nature – not counting in the JPEG2000 variant which supports lossless compression – and suffers from generation loss, i.e. increasing of noise (specifically noise to information ratio) which is easily detectable visually when zoomed in. PNG rectified the problem, but PDF provided the ultimate scalability. However, as Martin poignantly quipped, 'the cost of the software is higher than the average salary in the Czech Republic'. The program was the only commercial software used during typesetting, however (apart from operating systems).

We discovered a 'Vlna' macro[3] by Petr Olšák which seeks solitary conjuctions and prepositions 'a, i, o, k' placed at the end of the lines (a mistake by all means) and links them with the following words. The command

```
vlna -l -m -n dokument.tex
```

assures automation of inserting such tildes ('vlnka' in Czech). The `-l` switch enables the LaTeX mode, `-m` ignores the math environment and `-n` does the same for the verbatim environment.

The macro is intended for Czech language text only, depressing us as the proceedings was in English. So we put emphasis on looking for numbers and units divided on two lines and placed several tildes in the early stages of our work. It became clear only later that any modification in the text renders many of them unusable and unneeded. A conclusion was that is is not advisable to solve all problems at once but instead create a hierarchy.

Trials and Tribulations

Since October the uphill struggle with the system has begun. One point became clear early on: the way to operate (that is, the way generating the least amount of errors) is to create a main template, inserting chapters by invoking `\input{*}` which produces much more clear and concise code structure. As for the packages, there were two of note: `tocloft` a `caption`.

---

[3]`http://ftp.linux.cz/pub/tex/local/cstug/olsak/vlna/`

Lists: tocloft

`tocloft`[4] is a widely known and used set of macros for lists generation. Of those we needed three: chapters (ToC – Table of Contents), figures (LoF – List of Figures) and tables (LoT – List of Tables). The last two gave us our first serious headache. When we changed caption labels ('Figure' from 'Obrázek' and 'Table' from 'Tabulka') by means of altering the `caption`, the lists were subsequently populated only by their respective numbers, not the words themselves. We attempted to counter this behavior by leaving the default labels and redefining them globally by

`\renewcommand*{\figurename}{Figure}`
`\renewcommand*{\tablename}{Table}`

This produced the same result. I am not sure whether the problem is solvable only by changing the `babel` option to `[english]` instead of `[czech]` but I sincerely hope we tried that, too. Finally, we had to use a dirty trick. First, we indented all the items in the list by

`\addtolength{\cftfignumwidth}{25pt}`
`\addtolength{\cfttabnumwidth}{20pt}`

and then inserted the words 'Figure' and 'Table' to the existing space permanently by

`\renewcommand{\cftfigpresnum}{Figure }`
`\renewcommand{\cfttabpresnum}{Table }`

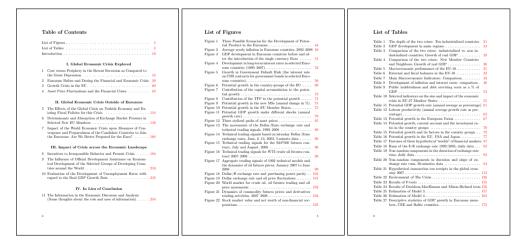By fine tuning the values, we obtained the solution as seen in Figure 1.

Figure 1: ToC and a sample of LoF and LoT.

---

[4] http://www.ctan.org/tex-archive/macros/latex/contrib/tocloft/

This was our first major milestone which, apart from the fact it was achieved by a former WYSIWYG[5] users, was the point where we realized even basic operations require a great deal of work.

I also concluded the documentation the authors of the packages provide are the source of all the information one needs despite some of them being a bit too technically demanding, i.e. `tocloft`. That's why we made it a habit of leaving one person (me) in charge of all changes made to (`tocloft` a `caption`). Lubor volunteered for the equally important task of cropping the pictures that, although hardly technical, was all the more demanding as to patience and accuracy required.

Labels: captions

We approached the `caption`[6] package by Axel Sommerfeldt with a certain uneasiness due to our previous dealings with English labels but at the end it proved to be straightforward and painless thanks to its easily comperehensible and descriptive documentation. We were able to completely get rid of `\renewcommand{*}{*}` and change the labels from Czech to English by using an elegant solution in the preamble:

`\usepackage{caption}`

`\captionsetup{figurename=Figure,tablename=Table}`

We strived to parametrize as much as possible by the packages' options rather than using global modifiers that could possibly lead to further, wholly unexpected problems.

Hyperlinks in captions made us equally worried. Before we utilized the `url` package each slash had to be preceeded by a backslash, proving to be our bane as after we decided on `url` all the backslashes had be to be removed again. Without using `url` the hyperlinks were further treated as plaintext and therefore unclickable in the PDF. Another conclusion abound: familiarize myself with the possibilities TeX offers before attempting to devise my own dirty solutions.

The resulting labels are shown in Figure 2:

To all graphic objects the string `[!hpbt]` was appended, courtesy of Pavel Stříž who proposed it after our complaints that all pictures are positioned at the end of the text, putting a reader in a dire situation especially when the author repeatedly commented on the object within the text body. Later on, I found out the characters serve as a guide on where to put the objects in TeX's notation.

My last conclusion came from seeing advantages of using `\textwidth` and `width` instead of `\scale`, a point we proved repeatedly. While the former scales the object in relation to the width of the text on a page, the latter does so absolutely. We had to very often tweak the `scale` manually as the object was

---

[5]What You See Is What You Get editor allows to see the changes made to the document without the need to re-compile it.

[6]`http://www.ctan.org/tex-archive/macros/latex/contrib/caption/`

Table 7 Main Macroeconomic Indicators: Comparison
Source: EC (2008)

Figure 41 ISE National 100 Index (Based on the Closing Prices)
Source: The Undersecretariat Turkish Treasury, Ekonomi Sunumu, http://www.hazine.gov.tr/irj/go/km/docs/documents/TreasuryWeb/Statistics/EconomicIndicators/egosterge/Sunumlar/Ekonomi_Sunumu_ENG.pdf, p. 129

Figure 2: An example of figure and table labels.

positioned outside the bounds of A5. By using any value lower or equal to one in the first declaration (for instance `\width=0.5\textwidth`) the problem is avoided.

Tables

Tables were another challenge we had to face when we finished with figures. There were two principial options in the matter: either convert all of them to TEX or safely crop them as we did previously. Because my greatest achievement in this regard at that time was the `{c|c|c}` declaration, causing all the text in three columns to be centered, and because several tables looked like even A3 format was too small to accomodate them, the second way was definitely preferred.

The problems arose immediately, first of which was the output format. Several extremely large tables were divided on two consecutive pages and therefore needed to be joined before we were able to crop them at all. I could not achieve this on my non-widescreen monitor so I set the page format in Microsoft Word to an extremely large value (B0, or 1000×1414 mm was sure to work) instead, then using a virtual printer PDFCreator[7] to set the same output format and printed to PDF.

Before I stumbled upon the correct value, though, some time has passed. In Adobe Acrobat, the resulting behemoth had to be scaled down for the lack of 1.4 meters wide monitor in our possession. We were proud... until we found out the tables were not converted to the same format which required us to do just that and then repeat the process of virtual printing and cropping again. Fortunately, Martin Stříž came to our help and diligently sit through the task, helping us when we realized the results of several days worth of work are useless.

Miscellaneous

During the course of the work file management and naming were a constant issue. We had several versions of files, tables, figures, templates, working and

---

[7]http://sourceforge.net/projects/pdfcreator/

to-be-discarded experiments in the e-mails' attachments, removable media and folders. Maintaining order was imperative especially because we were working in a decentralized fashion with moth of us having different way of keeping order in our repositories. This resulted in some working experiments to be repeated because the template the code was re-written by another piece of working code elsewhere. What I finally came to do was to save a working template at the beginning of each day and after the work was done copying the additions to the backup file before compiling it. Thus we always had a working version as a core component, modified as needed. In case of procuring a file for control, the maximum offset would be 24 hours.

As for file naming convention, the spontaneous system we devised and tested consisted of `figXX` for figures, and `tabXX` for tables. When the table was divided into two pages, the files were named `tabXXa` a `tabXXb`, the joined table `tabXX` as usual. The notation for invidual chapters was `kapXX` (meaning chapter), the main template was `sablona_hlavni` (meaning main_template). Even by quickly looking at the contents of the folder it was possible to make out the structure easily.

Several marginal challenges appeared, solved by dilligently looking through forum boards of LATEX users. A simple action such as putting a dot after chapter name still resides in my bookmarks. Answers to such questions, sought after by beginners, are to professionals a matter of invoking a corresponding declaration (`\renewcommand{*}{*}`) or loading a package containing the functionality (`titlesec`). We certainly forgot lots of things as we haven't used them since, and it was also possible to solve a lot of these problems more easily than we did back then. For instance, we loaded the `fancyhdr` package and used it only to produce a plain page with no numbering, a feature easily achieved by setting `\thispagestyle{empty}`. We handled clubs and widows[8] by increasing the scale of the preceeding picture by avoiding them while the correct solution is to set the correct value in `\widowpenalty=XX` and `\clubpenalty=XX`.

One of the things we did not pay attention to initially were quotation marks. We were not too experienced in this regard in Czech language and asked for the correct way of handling them. The answer in the form of warning was that they are problematic and that we had to make sure we had them right. We promised we would be careful, and did the mistake anyway. The situation was complicated by the fact the authors of the articles were foreigners (meaning the quotation marks were not 'Czech'), the proceedings was printed in the Czech Republic, though (suggesting Czech quotation marks should have preference). We chose the second argument to base our assumptions on. This caused us to sift through the whole text again and correcting that mistake. To avoid further complications

---

[8]In typography, a club is the first line of a paragraph positioned on the last line of the page; a widow is the last line of a paragraph positioned on the first line of the page.

when seaching according to `Alt+` codes, we copied the quotation marks off Czech and English version of Wikipedia. Czech quotation marks look like „this", English like "this", or to put it simply, Czech quotation marks resemble the pair 99 down and 66 up while the English ones match 66 and 99 up.

However, we soon discovered a catch in the form of inserted quoation marks which also exhibits this peculiar property. Czech inserted quoation marks „look like ‚this'", the English ones "look 'like this'", however (so, in terms of pairs, 6 down 9 up for Czech language, 6 and 9 up for English). New sifting through the text and assuring all inserted quoation marks are the English version. Afer this was done we were confident we got through the worst of it. It came to us as a surprise when we found out some authors did not respect typograhic conventions and instead of standard characters as seen above used the apostrophe (') with nearly unrecognizable difference in appearance in the source text. Third sifting and close examination of the English keyboard layout in search of another characters possibly representing quotation marks bore no fruit and we were confident this was the last time we heard of any problem concerning quotation marks. Only later we found there was another 'type' of inserted quoation marks in form of acute. The final sifting was done in a matter of minutes.

Conclusion

Typesetting the proceedings greatly enriched us both. With the help of Pavel Stříž and his brother Martin the work was finalized before the deadline. After proofreading, additional corrections and production of the cover by a third party we received a message announcing the book is finished. When we got the physical copy we were rightfully proud as every equation, table, figure, quotation mark, punctuation mark and dot was revised multiple times. Even though several mistakes came to light before the printing we were not having too ambitious a goal of leaving no mistakes behind but rather to find out if we can do it.

We did and learned a lot of things in the process we now use when preparing presentations for the students whom we also try to enlighten as to the ways of open source typography, particulary TeX. They should be allowed to see that it is possible to typeset not only using commercial software but with free solutions as well. Taking into account further tools for statistics, such as programming language R, gretl or gnuplot graphic utility, we may safely conclude quality science can be done with minimal (zero) costs involved.

sarga@fame.utb.cz
Faculty of Management and Economics
Tomas Bata University in Zlín
Mostní 5139, 760 01 Zlín