VnTeX — Typesetting Vietnamese

Hàn Thế Thành   Reinhard Kotucha

Abstract

VnTeX is an extension to Donald Knuth's TeX typesetting system which provides support for typesetting Vietnamese.

The primary site of VnTeX is `https://vntex.sf.net`.

1 Where to get Help

The current maintainers of VnTeX are:

- ▶ Hàn Thế Thành ⟨`HanTheThanh@gmail.com`⟩

- ▶ Reinhard Kotucha ⟨`Reinhard.Kotucha@gmx.de`⟩

- ▶ Werner Lemberg ⟨`WL@gnu.org`⟩

There is a mailing list (very low traffic) for questions about VnTeX and typesetting Vietnamese. To subscribe to the list, visit:

`https://lists.sourceforge.net/lists/listinfo/vntex-users`

2 Related Documents

The following files are part of the VnTeX distribution

- ▶ *Hàn Thế Thành,* Hỗ trợ tiếng Việt cho TeX

- ▶ *Hàn Thế Thành,* Minimal steps to typeset Vietnamese

- ▶ *Hàn Thế Thành và Thái Phú Khánh Hòa,* Dùng font với VnTeX

The following files are not part of VnTeX but might be part of the TeX distribution you are using.

- ▶ *American Mathematical Society*, Hướng dẫn sử dụng gói amsmath,
  `https://ctan.org/tex-archive/info/amslatex/vietnamese/amsldoc-vi.pdf`
  `https://ctan.org/tex-archive/info/amslatex/vietnamese/amsldoc-print-vi.pdf`

- ▶ *H. Partl, E. Schlegl, I. Hyna, T. Oetiker*, Một tài liệu ngắn gọn giới thiệu về LaTeX 2_ε,
  Translated by Nguyễn Tân Khoa.
  `https://ctan.org/tex-archive/info/lshort/vietnamese/lshort-vi.pdf`

- ▶ *Wolfgang May, Andreas Schlechte*, Mở rộng môi trường định lý. Translated by Huỳnh Kỳ Anh. `https://ctan.org/tex-archive/info/translations/vn/ntheorem-doc-vn.pdf`

3 Typesetting Vietnamese

In order to typeset Vietnamese, you need a text editor which supports Vietnamese. In particular, it should support an *input encoding* and an *input method* suitable for Vietnamese.

If you are not familiar with encodings, here is a brief explanation: Each key on your keyboard is assigned to a letter. Computers don't understand letters, they only understand numbers. The table which assigns letters to numbers is called *input encoding*.

input encoding A popular input encoding system used in Vietnam is VISCII. The problem is that only 256 characters can be used at the same time. It's sufficient for typesetting Vietnamese, however, it's not well suited for multilingual texts. A better approach had been provided by the Unicode Consortium: UTF-8. This is a very efficient encoding system which supports all writing systems of the world. You can have Vietnamese, Arabic, Korean, Ethiopian, Hindi,…characters in one and the same file. UTF-8 is the encoding system of the future and it becomes even more popular in Vietnam.

VnTeX supports many input encodings such as VISCII, TCVN, or UTF-8, but there is no support for VNI (nor will there ever be).

You can use the input encoding of your choice, but you have tell TeX which one you are using. How to do this is decribed below.

font encoding There is a similar issue with fonts. A font is a collection of *glyphs*. A glyph is the graphical representation of a character. Graphical representations of the character a might be 'a', '*a*', or '**a**', for example. Fonts never contain characters; they contain only glyphs, sometimes more than a single glyph for a given character. A font usually contains more than 256 glyphs, but TeX can only access 256 characters at the same time. The table which maps characters to glyphs is called a *font encoding*.

However, if you are using LaTeX, a name is assigned to each character read from the keyboard. This way it can deal with an arbitrary number of characters internally. You can specify more than one font encoding and LaTeX switches between them automatically. In most cases it's sufficient to know that font encoding `T1` supports Western European languages and `T5` supports Vietnamese.

input method But how to enter all the characters if you have only an American keyboard? You have to select an *input method*. An input method allows you to access characters which are not supported by your keyboard. If you select VIQR as an input method, you can write "`Ha` `No^.i`" on your keyboard but you see "`Hà Nội`" on screen and you get "Hà Nội" in your typeset document.

However, input methods are quite system dependent. If your operating system doesn't support anything appropriate, check whether your editor or TeX shell supports them.

editors It's not easy to propose a particular editor. If you are using a reasonably powerful editor for writing your own programs, then use it for TeX too.

Editors which are supposed to work on all operating systems are vim, Emacs, TeXMaker, and TeXworks. On Windows there are some alternatives, like WinEDT, and TeXnicCenter. TeXnicCenter supports UTF-8 as of version 2.0.

If you are on Mac OS X, TeXshop is a good choice. TeXshop is aimed at beginners but it is extremely powerful though. It provides a very fast PDF viewer and if you click on a particular word in the PDF file, the cursor moves to this word in the text editor, and vice versa. TeXworks is something very similar. But it is supposed to work on all operating systems and is shipped with TeX Live and MikTeX.

There are some different flavours of TeX, such as Plain TeX, LaTeX, and Context. LaTeX is the most popular one and there are many books available about it.

3.1 Typesetting with LaTeX

The idea of LaTeX is to treat content and layout separately. If you never used LaTeX before, please read Một tài liệu ngắn gọn giới thiệu về LaTeX 2_ε first.

3.1.1 Using vietnam or vntex

There are two packages, vietnam and vntex. They are quite similar, the only difference is that the default input encoding is VISCII in vietnam and UTF-8 in vntex. However, both packages allow you to specify any supported input encoding. The following encoding systems are supported:

viscii	use VISCII input encoding
mviscii	use MVISCII input encoding
tcvn	use TCVN input encoding
vps	use VPS input encoding
utf8	use UTF-8 input encoding (LaTeX)
utf8x	use UTF-8 input encoding (ucs package)
noinputenc	do not load the inputenc package (use of TCX is assumed)

Examples:

```
\documentclass{report}
\usepackage{vietnam} % use VISCII input encoding
\begin{document}
⟨...text in VISCII encoding...⟩
\end{document}
```

```
\documentclass{report}
\usepackage{vntex} % use UTF-8 input encoding
\begin{document}
⟨...text in UTF-8 encoding...⟩
\end{document}
```

```
\documentclass{report}
\usepackage[tcvn]{vntex} % use TCVN input encoding
\begin{document}
⟨...text in TCVN encoding...⟩
\end{document}
```

Both packages, vietnam and vntex, have the following additional options:

nocaptions	do not define Vietnamese captions
varioref	load the varioref-vi package
cmap	load the cmap package

If the option nocaptions is set, then captions are typeset in English. On the other hand, if you are using the varioref package, you might want to set the varioref option in order to get "ở trang liền sau" instead of "on the following page", for example.

The cmap packages makes the PDF file searchable.

3.1.2 Using **babel** instead of **vietnam/vntex**

For multilingual documents it's better to use the babel package, which is part of the LaTeX core. Though the inputenc package allows you to select the input encoding of your choice, UTF-8 is the preferred encoding for multilingual documents.

```
\documentclass{report}
\usepackage[T2A,T5]{fontenc}
\usepackage[utf8]{inputenc}
\usepackage[russian,vietnamese]{babel}
\begin{document}
    Tiếng Việt,
    \selectlanguage{russian}%
    русский язык,
    \selectlanguage{vietnamese}%
    tiếng Việt.
\end{document}
```

Note that last optional argument passed to babel is the language which is active at the beginning of your document.

The result of the example above is: | Tiếng Việt, русский язык, tiếng Việt. |

3.1.3 Using **hyperref**

In order to use Vietnamese characters in the bookmark panel or in the "Document Properties" box, hyperref must be loaded with the unicode option.

```
\usepackage[unicode]{hyperref}
\hypersetup{pdftitle={VnTeX - hỗ trợ tiếng Việt cho TeX}}
```

3.1.4 Using **TCX files**

TeX itself can't use non-ASCII characters when writing error messages to screen or to the log file. Instead, it prints non-ASCII chacters in hexadecimal notation, like ^^DF. But there is an extension called TCX. If you activate TCX, a translation table is loaded, and all files TeX reads are translated before they are processed. If you are using TCX, you can't use the inputenc package because the translation can be done only once.

If you are using an engine which supports UTF-8 natively, like X∃TeX or LuaTeX you can't use TCX (and you don't need to).

VnTeX provides two TCX tables, viscii-t5 and tcvn-t5. Here is an example:

```
%& -translate-file=viscii-t5
\documentclass{report}
\usepackage[noinputenc]{vntex}
\begin{document}
⟨. . .text in VISCII encoding. . .⟩
\end{document}
```

The very first line says that the option -translate-file=viscii-t5 is passed to TeX when compiling the document. It has the same effect as if you run

```
latex -translate-file=viscii-t5 foo.tex
```

on the command line. Using TCVN is similar.

3.1.5 Creating HTML from LaTeX sources

In order to create HTML documents from LaTeX sources, run

```
tex4ht "html,uni-html4,charset=utf8" yourfile.tex
```

on the command line. You can't use TCX with tex4ht.

3.2 Typesetting with plain TeX

Unfortunately, there is no package for UTF-8 input encoding in plain TeX yet.

3.2.1 **plainenc and plnfss**

plnfss provides a LaTeX-like interface for font selection.

```
\input t5code
\input plnfss
\input plainenc
\fontencoding{T5}
\inputencoding{viscii} % or any other encoding mentioned
                       % above except utf8
\setfontencoding{T5}
\selectfont
⟨...text in VISCII encoding...⟩
\bye
```

plainenc and plnfss are not part of the VnTeX distribution any more but it is very likely that they are part of the TeX system you are using.

3.2.2 Using **TCX**

TCX files can be used as described in the LaTeX section.

```
%& -translate-file=viscii-t5
\input t5code
\input plnfss
\setfontencoding{T5}
\selectfont
⟨...text in VISCII encoding...⟩
\bye
```

3.3 Using **texinfo**

TCX is required:

```
%& -translate-file=viscii-t5
\def\fontprefix{vn}
\input t5code.tex
\input texinfo
⟨...text in VISCII encoding...⟩
```

There are some test files for VnTeX in `texmf*/source/latex/vntex/tests/`. Please read the file README in this directory.

4 Vietnamese Fonts

VnTeX provides a lot of Vietnamese fonts. If you are using T5 font encoding but do not specify any font (as in the examples above) you get Vietnamese Computer Modern. These VNR fonts are extensions to Donald Knuth's Computern Modern Fonts and were designed by Hàn Thế Thành.

4.1 Acquiring Vietnamese Fonts

4.1.1 Fonts provided by VnTeX

The following fonts are part of VnTeX. Vietnamese Glyphs were added by Hàn Thế Thành.

▶ Arev (a version of Bitstream Vera Sans)

- ▶ Bitstream Charter
- ▶ Computer Modern
- ▶ Computer Modern Bright
- ▶ Concrete
- ▶ txtt
- ▶ URW Grotesk
- ▶ urwvn (URW version of Adobe's LaserWriter fonts)
- ▶ Vntopia (based on Adobe Utopia)

4.1.2 VnTeX nonfree Fonts

Some of the fonts donated by URW can be used freely but they can't be distributed if money is charged for the distribution. These fonts are not part of the VnTeX core distribution because otherwise VnTeX can't be in TeX Live or in Linux distributions.

These fonts are:

- ▶ URW Classico (URW version of Hermann Zapf's Optima)
- ▶ URW Garamond

There is an extra package containing these fonts:

    https://vntex.sourceforge.net/download/vntex/vntex-nonfree.zip
    https://vntex.sourceforge.net/download/vntex/vntex-nonfree.tar.xz

If you are using TeX Live, you can download and execute install-getnonfreefonts from

    https://tug.org/fonts/getnonfreefonts

and run `getnonfreefonts --help` on the command line for more information.

4.1.3 Microsoft Core Fonts

Support for Microsoft's *Web Fonts* was removed from VnTeX because the actual fonts cannot be provided for legal reasons. Please consult the VnTeX homepage for more information.

4.1.4 Other Fonts supporting Vietnamese

There are many other fonts supporting Vietnamese which are not shipped with VnTeX because they are an integral part of any modern TeX distribution anyway.

font samples There are sample files of all fonts which support Vietnamese, can be used with TeX, and can be used freely, even commercially. However, some of them can't be distributed if you charge money for the distribution.

> https://vntex.sf.net/fonts/samples

Not every font supports maths. If you have to typeset math formulas, consult:

> https://ctan.org/tex-archive/info/Free_Math_Font_Survey/vn/survey-vn.pdf

4.2 Font Selection

We describe how to use fonts with LaTeX first. A description of plnfss (plain TeX) is given below.

4.2.1 Selecting Fonts in LaTeX

Some fonts provide a LaTeX macropackage which loads the necessary fonts.

To use Latin Modern instead of VNR, simply

```
\usepackage{lmodern}
\usepackage{vntex}
```

For Antikwa Toruńska, do

```
\usepackage{anttor}
\usepackage{vntex}
```

...or use inputenc and babel instead of vietnam or vntex.

Some font packages do not provide such a LaTeX macro package. An example is urwvn.

It is recommended to specify a roman font, a sans-serif font and a typewriter font separately. You do not have to specify all of them. It makes sense, for instance, not to specify a typewriter font — you get Computer Modern Typewriter then, which is a good choice.

Command	PostScript Name	Font Family Name
\renewcommand\sfdefault{uag}	VnURWGothicL	AvantGarde
\renewcommand\rmdefault{ubk}	VnURWBookmanL	Bookman
\renewcommand\ttdefault{ucr}	VnNimbusMonL	Courier
\renewcommand\sfdefault{uhv}	VnNimbusSanL	Helvetica
\renewcommand\rmdefault{unc}	VnCenturySchL	New Century Schoolbook
\usepackage{mathpazo}	VnURWPalladioL	Palatino
\usepackage{mathptm}	VnNimbusRomNo9L	Times

small caps There is also a real small caps font for VnURWPalladioL, made by Ralf Stubner and extended by Hàn Thế Thành. There are still some support files missing.

By default, you get the faked small caps but you can use real small caps with some restrictions. To make use of them, put the following macro definition into the preamble of your document:

```
\newcommand{\textfplsc}[1]{\bgroup\usefont{T5}{fpl}{m}{sc}#1\egroup}
```

You can use it like this:

> ⟨...*some text*...⟩ \textfplsc{⟨...*some text in small caps*...⟩} ⟨...*some text*...⟩

The macro argument should not contain any numbers because they will appear as oldstyle numbers.

math fonts If you have to typeset math formulas, be aware that not all fonts support math. The following fonts support math very well:

Font	Command
Computer Modern	do nothing
Latin Modern	\usepackage{lmodern}
Palatino	\usepackage{mathpazo}
Times	\usepackage{mathptm}

There are many others too, please consult:

> https://vntex.sf.net/fonts/samples/survey-vn.pdf

However, some of the fonts borrow math symbols from other fonts and it's worthwhile to check whether all the symbols you need blend well with the base font you are using. Be very careful when using sans-serif fonts in math formulas. It's very painful if there is no significant difference between "l" and "I". Do you see any difference at all? The first one is a lowercase "L", the second one is an uppercase "i".

MS core fonts If you are using Windows, you also can use the fonts provided by Microsoft:

Command	PostScript Name	Font Family Name
\renewcommand\sfdefault{ma1}	ArialMT	Arial
\renewcommand\ttdefault{mcr}	CourierNewPSMT	Courier
\renewcommand\rmdefault{lpr}	PalatinoLinotype	Palatino
\renewcommand\rmdefault{mns}	TimesNewRomanPSMT	Times New Roman
\renewcommand\sfdefault{jth}	Tahoma	Tahoma
\renewcommand\sfdefault{jvn}	Verdana	Verdana

None of the Microsoft fonts supports mathematics. Though the quality of the fonts is quite high, not much care had been taken in the design of Vietnamese accents (except in *Palatino Linotype*). See:

> https://vntex.sf.net/fonts/samples

Unless someone insists that you use these fonts, you can use

VnNimbusMonL	instead of CourierNewPSMT	Courier
VnNimbusSanL	instead of ArialMT	Helvetica/Arial
VnNimbusRomNo9L	instead of TimesNewRomanPSMT	Times/Times New Roman
VnURWPalladioL	instead of PalatinoLinotype	Palatino

4.2.2 Selecting Fonts in plain TeX

If you are using plain TeX, you can use plnfss.tex to select fonts.

```
\setrmdefault{...}
\setsfdefault{...}
\setttdefault{...}
```

See the plnfss documentation for more details.

5 Licenses

The URW and Bitstream Type1 fonts are copyrighted under the GNU GPL, .map files are public domain, varioref-vi.sty is under LGPL, Vntopia is under the Adobe/TUG Utopia

license agreement, all other files are under LPPL, version 1.3 or newer.

- ► https://www.gnu.org/licenses/gpl.txt
- ► https://www.gnu.org/licenses/lgpl.txt
- ► https://www.latex-project.org/lppl.txt
- ► https://tug.org/fonts/utopia/LICENSE-utopia.txt

6 Contributors

The author of VnTeX is Hàn Thế Thành. Current maintainers are Reinhard Kotucha and Werner Lemberg.

LaTeX support (input encoding files, font encoding files, babel support files and viet-nam.sty) were provided by Werner Lemberg. vntex.sty was proposed by Huỳnh Kỳ Anh.

Vietnamese fonts for tex4ht originally were provided by Hàn Thế Thành, but they are now part of the tex4ht distribution.

plnfss was written by Hàn Thế Thành and Michal Konečný. It was removed from VnTeX because it supports many other languages as well.

7 Known Problems

- ► In order to use amsart.cls (and other AMS LaTeX document classes) with Unicode you must add the following lines immediately before \begin{document}:

```
\def\firstofone#1{#1}
\let\uppercase\firstofone
\let\MakeUppercase\firstofone
```

This completely disables LaTeX's uppercasing commands which might cause bad secondary effects. Note that this problem is not specific to Vietnamese but affects any multibyte encoding.