

# VnT<sub>E</sub>X — Typesetting Vietnamese

Hàn Thế Thành   Reinhard Kotucha

## Abstract

VnT<sub>E</sub>X is an extension to Donald Knuth's T<sub>E</sub>X typesetting system which provides support for typesetting Vietnamese.

The primary site of VnT<sub>E</sub>X is <http://vntex.sf.net>.

## 1 Where to get Help

The current maintainers of VnT<sub>E</sub>X are:

- ▶ Hàn Thế Thành <[HanTheThanh@gmail.com](mailto:HanTheThanh@gmail.com)>
- ▶ Reinhard Kotucha <[Reinhard.Kotucha@web.de](mailto:Reinhard.Kotucha@web.de)>
- ▶ Werner Lemberg <[wL@gnu.org](mailto:wL@gnu.org)>

There is a mailing list (very low traffic) for questions about VnT<sub>E</sub>X and typesetting Vietnamese. To subscribe to the list, visit

<http://lists.sourceforge.net/lists/listinfo/vntex-users>

## 2 Related Documents

The following files are part of the VnT<sub>E</sub>X distribution

- ▶ *Hàn Thế Thành*, Hỗ trợ tiếng Việt cho T<sub>E</sub>X
- ▶ *Hàn Thế Thành*, Minimal steps to typeset Vietnamese
- ▶ *Hàn Thế Thành và Thái Phú Khánh Hòa*, Dùng font với VnT<sub>E</sub>X

The following files are not part of VnT<sub>E</sub>X but might be part of the T<sub>E</sub>X distribution you are using.

- ▶ *The American Mathematical Society*, Hướng dẫn sử dụng gói amsmath,  
<http://ctan.org/tex-archive/info/amslatex/vietnamese/amslatex-vi.pdf>  
<http://ctan.org/tex-archive/info/amslatex/vietnamese/amslatex-print-vi.pdf>
- ▶ *H. Partl, E. Schlegl, I. Hyna, T. Oetiker*, Một tài liệu ngắn gọn giới thiệu về L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub>,  
Translated by Nguyễn Tân Khoa.  
<http://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.  
<http://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.

Vn $\TeX$  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 to tell  $\TeX$  which one you are using. How to do this is described 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 might be 'a', 'á', or 'ǎ', 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,  $\TeX$ Maker, and  $\TeX$ works. On Windows there are some alternatives, like  $\TeX$ shell, WinEDT, and  $\TeX$ nicCenter.

If you are on Mac OS X,  $\TeX$ shop is a good choice.  $\TeX$ shop 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.  $\TeX$ works is something very similar. But it is supposed to work on all operating systems and will probably be the default editor in  $\TeX$  Live and Mik $\TeX$ .

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  $\LaTeX 2_{\epsilon}$  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  $\TeX$  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,vietnam]{babel}
\begin{document}
  Tiếng Việt,
  \selectlanguage{russian}%
  русский язык,
  \selectlanguage{vietnam}%
  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 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 characters in hexadecimal notation, like  $\^{\wedge}DF$ . 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.

TCX and inputenc are more or less equivalent. The advantage of TCX is that you get Vietnamese characters in messages about overflow/underfull boxes and the like. The main drawback is that it doesn't support UTF-8.

$\text{Vn}\TeX$  provides two TCX tables, viscii-t5 and tcvn-t5. Here is an example:

```
%& -translate-file=viscii-t5
\documentclass{report}
\usepackage[noinputenc]{vietnam}
\begin{document}
  \langle \dots \textit{text in VISCI} \textit{encoding} \dots \rangle
\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.

### 3.1.4 Creating HTML from $\mathbb{E}\TeX$ sources

In order to create HTML documents from  $\text{La}\TeX$  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  $\TeX$ -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  $\text{Vn}\TeX$  distribution any more but it is very likely that they are part of the  $\TeX$  system you are using.

If you have a version of plnfss which doesn't already support Vietnamese, please install

<http://vntex.sf.net/download/vntex-support/plnfss.zip>

### 3.2.2 Using TCX

TCX files can be used as described in the  $\TeX$  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  $\text{Vn}\TeX$  in `texmf*/source/latex/vntex/tests/`. Please read the file README in this directory.

## 4 Vietnamese Fonts

$\text{Vn}\TeX$  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 Computer Modern Fonts and have been designed by Hàn Thế Thành.

*nonfree fonts* VnTeX provides Vietnamese versions of free fonts donated by Adobe, URW, and Bitstream. Some of the URW fonts 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. There is an extra package containing these fonts:

```
http://vntex.sourceforge.net/download/vntex/vntex-nonfree.zip
http://vntex.sourceforge.net/download/vntex/vntex-nonfree.tar.bz2
```

If you are using TeX Live, installation is very simple. Just run

```
getnonfreefonts --help
```

on the command line in order to get more information.

*Latin Modern and T<sub>E</sub>X Gyre* There are many other fonts supporting Vietnamese which are not shipped with VnTeX because they are an integral part of any modern T<sub>E</sub>X distribution anyway. Some quite interesting fonts are the Latin Modern and T<sub>E</sub>X Gyre fonts created by Bogusław Jackowski and Janusz M. Nowacki. They fully support Vietnamese but their main advantage is that they support virtually all Latin scripts used today. Thus, they are well suited for multilingual documents.

*MS core fonts* The Microsoft core fonts are supported too. VnTeX only provides the additional files needed to make them accessible to T<sub>E</sub>X, but not the fonts themselves. If you are on Windows, they are available already.

If you are on Unix, you have to install them yourself.

```
http://sourceforge.net/project/showfiles.php?group_id=34153
```

Though the quality of the original fonts was quite good, the Vietnamese accents were added later, and obviously not by the guys who designed the fonts. The only exception is Palatino Linotype. Inspect the fonts carefully before you decide which one you want to use.

*font samples* There are sample files for all fonts which support Vietnamese, can be used with T<sub>E</sub>X, and can be used freely, even commercially. However, some of them can't be distributed if you charge money for the distribution.

```
http://vntex.sf.net/fonts/samples
```

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

```
http://ctan.org/tex-archive/info/Free_Math_Font_Survey/vn/survey-vn.pdf
```

## 4.1 Selecting Fonts in T<sub>E</sub>X

We describe how to use fonts with T<sub>E</sub>X first. A description of plnfs (plain TeX) is given below.

To use Latin Modern instead of VNR, simply

```
\usepackage{lmodern}
\usepackage[viscii]{vietnam}
```

For Antikwa Toruńska, do

```
\usepackage{anttor}
\usepackage[viscii]{vietnam}
```

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

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
<code>\renewcommand\sfddefault{uag}</code>	VnURWGothicL	AvantGarde
<code>\renewcommand\rmdefault{ubk}</code>	VnURWBookmanL	Bookman
<code>\renewcommand\ttdefault{ucr}</code>	VnNimbusMonL	Courier
<code>\renewcommand\sfddefault{uhv}</code>	VnNimbusSanL	Helvetica
<code>\renewcommand\rmdefault{unc}</code>	VnCenturySchL	New Century Schoolbook
<code>\usepackage{mathpazo}</code>	VnURWPalladioL	Palatino
<code>\usepackage{mathptm}</code>	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:

```
\langle...some text...\rangle \textfplsc{\langle...some text in small caps...\rangle} \langle...some text...\rangle
```

The macro argument should not contain any numbers because they will appear as oldstyle numbers, but you cannot use oldstyle numbers with other font shapes in T5 encoding (Vietnamese) yet. This problem will be fixed in a future release.

*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	<code>\usepackage{lmodern}</code>
Palatino	<code>\usepackage{mathpazo}</code>
Times	<code>\usepackage{mathptm}</code>

There are many others too, please consult:

[http://ctan.org/tex-archive/info/Free\\_Math\\_Font\\_Survey/vn/survey-vn.pdf](http://ctan.org/tex-archive/info/Free_Math_Font_Survey/vn/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 “l”. 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
<code>\renewcommand\sfddefault{ma1}</code>	ArialMT	Arial
<code>\renewcommand\ttdefault{mcr}</code>	CourierNewPSMT	Courier
<code>\renewcommand\rmdefault{lpr}</code>	PalatinoLinotype	Palatino
<code>\renewcommand\rmdefault{mns}</code>	TimesNewRomanPSMT	Times New Roman
<code>\renewcommand\sfddefault{jth}</code>	Tahoma	Tahoma
<code>\renewcommand\sfddefault{jvn}</code>	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. See:

<http://vntex.sf.net/fonts/samples>

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

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

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

Instead of

```
\renewcommand\rmdefault{...}
\renewcommand\sfddefault{...}
\renewcommand\ttdefault{...}
```

you have to select fonts like this:

```
\setrmdefault{...}
\setsfddefault{...}
\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.

- ▶ <http://www.gnu.org/licenses/gpl.txt>
- ▶ <http://www.gnu.org/licenses/lgpl.txt>
- ▶ <http://www.latex-project.org/lppl.txt>
- ▶ <http://tug.org/fonts/utopia/LICENSE-utopia.txt>



## 6 Contributors

$\text{\LaTeX}$  support (input encoding files, font encoding files, babel support and vietnam.sty) had been provided by Werner Lemberg. vntex.sty had been proposed by Huỳnh Kỳ Anh.

Vietnamese fonts for tex4ht had originally been provided by Hàn Thế Thành, but they are now maintained by Eitan Gurari.

plnfss had been removed from  $\text{VnTeX}$  because it supports many other languages as well. You can download it from

<http://ctan.tug.org/tex-archive/macros/plain/plnfss>

## 7 Known Problems

- ▶ In order to use amsart.cls (and other AMS  $\text{LaTeX}$  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  $\text{\LaTeX}$ 's uppercasing commands which might cause bad secondary effects. Note that this problem is not specific to Vietnamese but affects any multibyte encoding.

- ▶ In order to use  $\text{VnTeX}$  with Prosper, you have to install

<http://vntex.sf.net/download/vntex-support/prosper-vn.zip>

- ▶ In order to use  $\text{VnTeX}$  with Foil $\text{\TeX}$ , you have to install

<http://vntex.sf.net/download/vntex-support/foiltex-vn.zip>

- ▶ If you have an old version of  $\text{\LaTeX}$  which does not support UTF-8, you have to use ucs.sty. If ucs.sty is not present on your system you can install

<http://vntex.sf.net/download/vntex-support/ucs-vn.zip>

We will keep this file as long as ucs is unmaintained. If a maintainer is found you should download a fresh ucs distribution from the original site.

This is a small subset of the ucs package which supports Vietnamese only. You can invoke it with `\usepackage[utf8x]{inputenc}`.

Instead of installing ucs we recommend you update  $\text{\LaTeX}$ . The latest version can be downloaded from

- <http://www.dante.de/tex-archive/macros/latex/base.zip> (Germany)
- <http://www.tex.ac.uk/tex-archive/macros/latex/base.zip> (UK)
- <http://tug.ctan.org/tex-archive/macros/latex/base.zip> (USA)