



# Unicode Convertor Reference Manual

© Kamban Software, Australia. All rights reserved.

Kamban Software is a Registered Trade Mark. Registered in Australia,

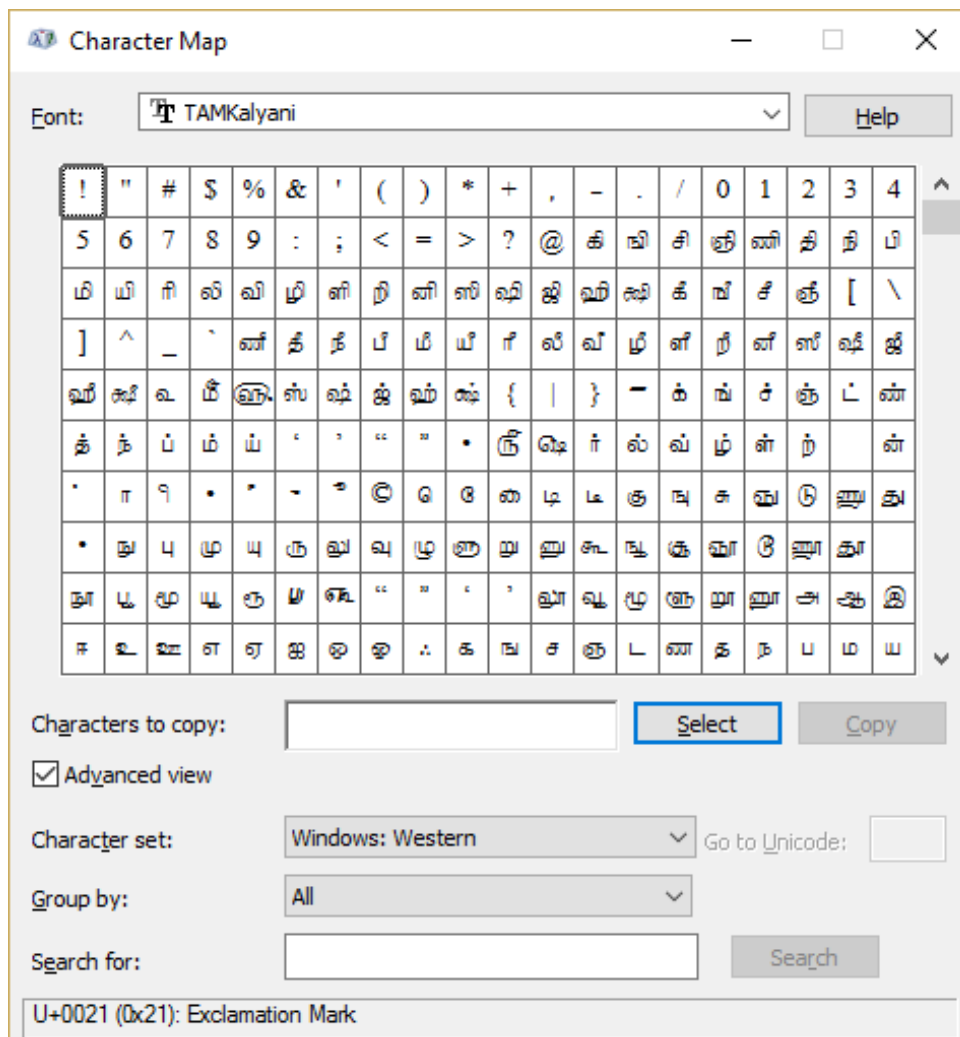
ABN: 18820534037

# Introduction

First let me thank you for purchasing the product. The amount of time spent on building any software product will not justify the cost associated with the product. This is especially true with Kamban Software, where we strive hard to keep the cost within the reach of common people absorbing any losses. This is because of our zeal to serve the community while profit is pushed back. If you can pay for the product rather than use a pirated version, this will alone encourage us to provide more packages keeping the cost very minimal.

## Conversion Basics:

Before UNICODE became standard, many font files with different character map existed. Many manufacturer, including Kamban Software created fonts that are unique and proprietary. Many files have been created with these fonts. Most computers in the 90s used English character set also known as Windows: Western or DOS: Western. They are all limited in mapping much less than 200 usable characters. To overcome this many Software vendors provided tools to manipulate these mappings. While this worked at that time, distribution of these files are limited and the user has to ship the font with the file for others to view.



With the dawn of Web, pages written with these character sets became obsolete. Many browsers did not recognise these character sets and many times the output on the screen looked garbled. To overcome all these issues and to support all the languages under one standard, UNICODE was introduced. This became very popular in early 2000s. Many programs converted to UNICODE standard and today, it is very rare to see the old font format. Kamban Software provided Unicode Editor back in 2000 which is followed by KWPLite - much enhanced version in 2012. Our sites had been converted to Unicode since 2005.

However, all data entered a decade ago became obsolete overnight. Most of them needs to be recreated or re-keyed in - a very tiresome process. Since most of these files are old, even the

availability of font files have become rare. Without the font file it is very difficult to get the character mapping. Given above is a typical character mapping of a simple TAM based font encoding:

Character Map is a standard Windows program normally accessed by clicking on Start and selecting Windows Accessories. You will find the icon for Character Map there. The character sets are mapped on English sets. Hence if the file created with this font is used, you will get all garbled text, when this font is not installed in your system. Some standards did exist for these mappings. Most notable one is TAM / TAB standard provided by Tamil Nadu Government in 1999 known as TN99 of which I was also a part. Prior to that a standard was developed known as TSCII. All these are for Tamil Language. Also, I received support calls from many users requiring to convert their texts using Adhwin-Tamil font. Some of them were NGO's and Ashrams who have many documents based on this font. Hence I decided to include this with the application. Also my friend in UK required convertor for Kannada fonts. I created a separate convertor for him. I thought I will include that as well in this application.

A convertor program was provided by me free for many registered users since 2005. A very basic convertor just to quickly convert Text files. A decade since then, I notice there is still requirement for the conversion. Most important of all is our own package KWP 3.03, which is sold even today. The files created with this application uses TAM / TAB standard fonts, which at one stage will require converting to Unicode.

Current Unicode Convertor, converts all the old mappings to Unicode mapping and restores the files to original albeit with Unicode mapping. This saves a lot of time in re-creating these documents. Moving forward, I ask the **Internet community**, mostly **users like you** who would like any of their non-standard fonts to convert to Unicode. If the request is appreciable level, I am happy to include them with this application. To facilitate this, I have separated the application from Filters used, thus it will be easier to create a simple filter file for each encoding, without having to change the program. Do let me know about your requirement and send me the font file. It can be on any Indian language. I will strive my best to provide a filter for that.

I have kept the cost very low. Amount of saving you obtain from using this will itself pay for the price you pay. As always this Application can be obtained completely free of cost to all Non-Profit Organisations. Please fill the appropriate Form in our Website: <http://kamban.com.au>

Finally, we have seen a lot of piracy in India and lost a lot of our revenue. I cannot stress enough on this. Please, please go to our website and register your product. Once you have registered, we can start providing support. Remember – all our products are life-time subscription and free minor version upgrade. Occasionally depending on when you purchased the product, you are entitled for free upgrade, if not heavy discount on our major version releases. We have changed our website to accommodate for easier registration.

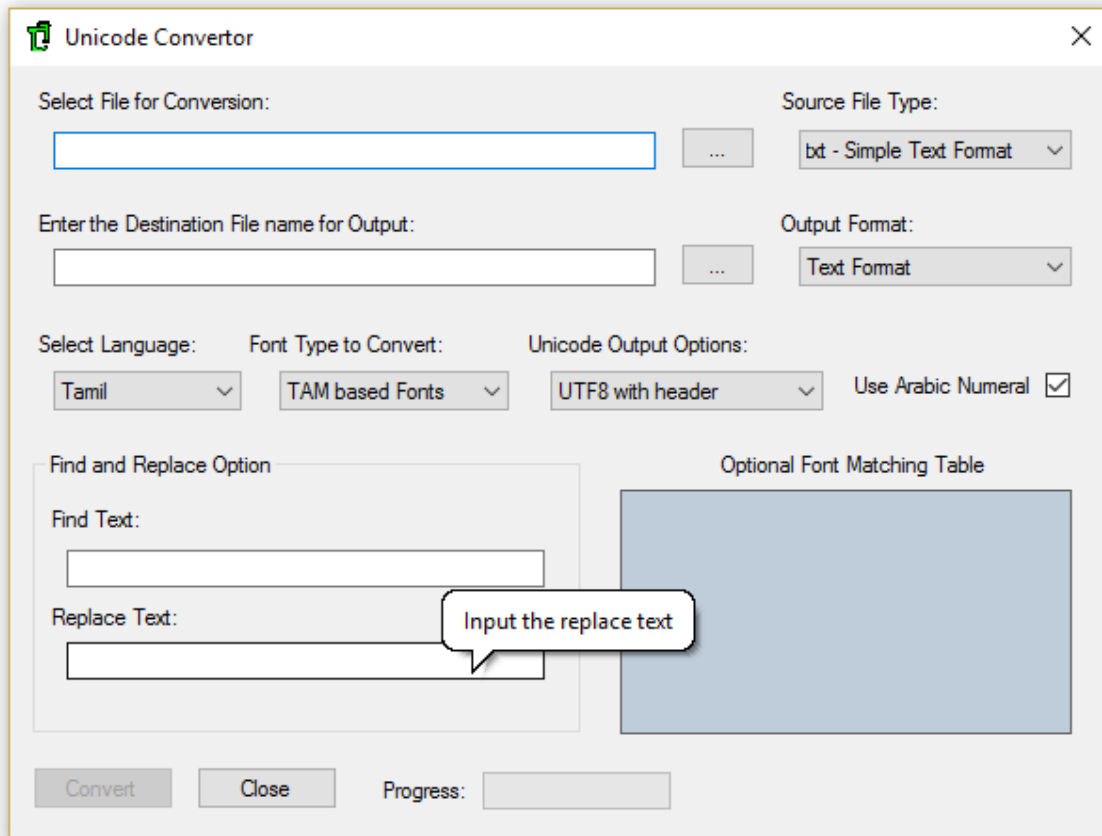
Do provide your feedback. Happy to always to help and serve the community



Vasudevan, Australia

18<sup>th</sup> February 2016

# Getting Familiar with Unicode Editor



The Application is generally self-explanatory. You can find tool tips as shown above when you move your mouse to each control within the Application window. This file provides a quick guide as a reference. If you want a RTF / DOCX version of this file, please write to us.

Here is a brief explanation of working with the application:

## Usage

1. Select the file for conversion. This can be a text file or Rich Text Format. In the dialog box that opens, select the file type. If you select (All files) and select a file, the extension will be considered as Text file.
2. If you select Rich Text format file, then the Language and Font Type Combo box will automatically display “—Auto—”. This is because the Application is intelligent enough to identify the font type and language based on the font represented in text. To aid the automatic selection of fonts, Font Matching Table can be provided. More on this can be found on the topic: “Font Matching Table – Specification”
3. If Text file is selected then you must select a language and Font Type. Wrong selection will result on garbled output.
4. Accept the default for other items and click on <Convert>
5. The output file will have be same file name with “\_uc” added to the original file. It will be saved on the same folder as the original.

**Note:** Most important feature of this utility is, you do not need to have the font installed for conversion. This is very helpful, if you do not have the font handy.

### Output Format:

Two options are available. If you are converting a text file, you can use RTF format for output. This is just a simple RTF conversion. Do not expect any major formatting.

If you select a RTF file for conversion, then you can select Text file format. Though available it is not advisable, as you will lose all the formatting information.

### Select Language:

Only two Languages are currently available. Many more will be added when there is demand for that. If you select a Text file (.txt extension) you must select a Language. This will ensure right encoding could be used. When a Language is selected Font Type option available for that Language will be displayed in Font Type Combo box.

You can leave this option to “—Auto—”, - default option for RTF format. However if you want to convert only a selected Language in the RTF format, you can optionally select the Language.

### Font Type to convert:

Depending upon the Language, available Font Type will be displayed. Select the right Font type for the text file.

You can leave this option to “—Auto—”, - default option for RTF format. However if you want to convert only a selected encoding in the RTF format, you can optionally select the Font Type for RTF files.

### Unicode Output Options:

There are 4 options available, plus one where you can select “No conversion required”. In general, you do not have to change any of these settings. Unless you know the file encodings, you will run into issues, by changing these values. Without going into details, here is the synopsis:

1. **UTF-8 with header**. When used, the encoding will be UTF-8. This is widely used in Internet and generally is the defacto standard. Each byte (or character) in your document will be represented by 3 bytes. Hence expect the file size to be 3 times bigger. The good news is, this can be distributed across the web or to any Operating system without having to worry about compatibility. This is the default setting.
2. **UTF-8 without header**: When the file is saved with UTF-8 encoding a small header is added to the file – 3 bytes. While most Operating system will ignore this header (other than to identify the encoding) MacOS will try to represent this in the text file. Thus at the start of the text file, you will see three letters, which are garbled. Rest of the texts will be fine. If you want to distribute your files to Apple computers, you can use this option.
3. **UTF-16 with header**: This is commonly used by Linux applications. In general, using this encoding will reduce the file size by 30%. This is because, the encoding uses 2 bytes per character as against 3 by UTF-8. This standard is not fully adapted by Internet, though more and more sites are slowly changing. Be aware that without the header file, it will be very difficult for any application to identify the right encoding of the file.

4. **UTF-16 without header:** This is same as above except the header info (2 bytes) is not added. Also as noted above, this option is not advisable since some applications will not be able to decode this properly.
5. **No Conversion Required:** Use this if no conversion required. Does not serve any purpose other than recreating a new file which is the original copy. However if you use “Find and Replace option” as defined in the next section, you can optionally replace all the word or sentence.

#### Find and Replace Option:

This is not a major feature to be tooted about. A very simple option found in all text applications. The only reason this is provided is, as an added advantage to change all occurrences provided in the Find Text Box.

Provide the Replacement word or sentence in the Replace Text Box. After conversion, the application will replace the text with this option. Generally this is much faster than any other program as the Find and Replace option here will work on core text file. Use this with “No conversion required” option to find and replace text in a large text files relatively faster.

#### Use Arabic Numerals

When you convert your files, whether Text file or RTF file, this option allows you to change the way the numerals are denoted in your converted file. Since most of the fonts provide Language based numerals, during conversion process, these numerals are converted to standard Arabic numerals. If you want to retain Language based numerals, then un-check this option.

Optionally in a RTF format, if Arabic numerals are provided, then you can use this option to use Language based numerals or vice-versa. A very useful feature if you want to represent your document with Language based numerals.

#### Optional Font Matching Table:

This is just a list box which provides the defined font matching and encoding. This is read from a file and represented here. Other than providing a visual information, there is nothing you can do here. See next Topic on more detailed explanation and how it is used.

#### Buttons – Convert, Close, Help & Progress Bar:

All self-explanatory. Convert is enabled only if you have selected a file for conversion and provided a destination output file. When Conversion is in progress, the <Close> button will change to <Abort> Thus if there is a lengthy operation in progress you can abort the operation by clicking on it. When there is no conversion, this button can be used to close the application.

Help file on clicking will open this file, provided you have right Windows settings to open up PDF file format. Check your Windows manual for this configuration. This file can always be found in your Start/All Apps/Kamban Software folder.

Progress Bar provides the status of conversion progress. Use this to approximately estimate time it takes for conversion and decide whether you want to continue or abort the process.

# Font Matching Table – Specification

---

First, why we do need this? Well, this is only used by RTF file, when mode “—Auto—” is selected. For all other options this is ignored. The system has no way of knowing which encoding is which font. With TAM / TAB standard, there was a font naming standard introduced. All TAM / TAB fonts must start with TAM or TAB. This is very advantageous in conversion, as the system can now sort out TAM and TAB encoded font without any issue. This luxury is not available for other fonts. One of the reason why the “encoding name” (or Font Type) always follows the font name so the system can properly match.

With this table, you can optionally help the application to identify the right encoding. This will ensure right output for your formatted text and significantly reduce formatting head-aches. Hence this option is provided. It is a simple text file which can be edited with any text editor. However the name and location of the file is important and must always be “**fntmatchlist.dat**” and must be located in the same folder as the executable file.

As mentioned earlier this is a text file. Each entry is separated by a “Carriage Return / LineFeed pair” – layman terms just <Enter> is pressed at the end of each line.

## [fntmatchlist.dat specification](#)

First line must contain the following sentence. Do not change this. This must be the first line. Ensure no space before this line.

Font Matching Listing

After this line you can optionally provide comments. All comments should begin with double forward slash - //. An example is given below:

```
//*****Start of Comment*****
```

You can have as many comment lines as you want. They can be anywhere on the file. These will be ignored by the Application.

## [Matching Table parts:](#)

The main table for matching consists of 3 parts.

1. First part is the name of the font. Please provide the font name with the right spelling. If in doubt use the “Character Map” to get the name of the font. If you have multiple fonts starting with the same name for ex: nudi – k1, nudi –k2, nudi K– lite, nudi K– bold etc, you only need one entry nudi. This will ensure all the variants are automatically considered as Font Type similar to that of Nudi by the application. There is no need to enter all the variation. Just enter the starting characters (at least 4 to be accurate) matching these fonts.
2. Second part consists of the Font Type. Currently defined Font Types are: ta\_TAM, ta\_TAB, ta\_TSC, ta\_ADHA, ka\_SHREE, ka\_NUDI. More font types will be added later, as more requirements arise. The first two letter for ex: ta stands for Language:Tamil while ka stands for Kannada. This will ensure same Font Types are possible for multiple languages. DOE standard generally defines common encoding to all Indian Languages. This two letter prefix will thus be able to identify the appropriate language.

3. Third part is optional. If provided, when the document is converted, the converted document for listed font (see point 1.) will be substituted with this font. Note that this font must be Unicode font. If it is not or if the font is not available default font “Arial” will be substituted.

Note: None of the texts are case sensitive. You can provide any of the words in any case – upper or lower or mixed. Each parts must be separated by | character. Spacing between the separators is ignored. If any of the line does not comply with this requirement, the entire line will be ignored.

#### Some Examples:

```
MaduramTSC | ta_TSC | Maduram
```

The above line will be interpreted by the application as follows:

Wherever the document consists of font starting with font name “MaduramTSC”, consider the language to be Tamil and the Font Type to be ta\_TSC. After conversion, substitute the “MaduramTSC” Font with “Maduram” font.

```
Adhawin-Tamil | ta_ADHA | Kalyani
```

The above line will be interpreted by the application as follows:

Wherever the document consists of font starting with font name “Adhawin-Tamil”, consider the language to be Tamil and the Font Type to be ta\_ADHA. After conversion, substitute the “Adhawin-Tamil” Font with “Kalyani” font.

```
Shree851 Lipi | ka_SHREE
```

The above line will be interpreted by the application as follows:

Wherever the document consists of font starting with font name “Shree851 Lipi”, consider the language to be Kannada and the Font Type to be ka\_SHREE. After conversion, substitute the “Shree851 Lipi” Font with default font.

Note how the language is interpreted from the coding. Also if you want to force the application to substitute a given font provide full name of the font – like Adhawin-Tamil. This will only substitute Adhawin-Tamil with Kalyani font and if there is Adhawin-Hindi, it will be ignored.

# Miscellaneous – Geeky stuff

---

Unicode Convertor is compiled on Visual Studio 2013. The application is built on Dot Net 4.5. Hence you must have .Net4.5 installed. The installation utility will install .Net4.5 if it is not available in your system.

This application is tested on Windows 7, 8.1 update and Windows 10. If you have any other versions of windows, the application may not work. However if it works, let us know. We can add that to the compatibility list.

The only help file is this. Done in MSWord and converted to PDF for easy distribution and also easier to read it on mobile platforms. If you want MSWord format write to us.

The application consists of following files all stored in the executable directory.

**UnicodeConvertor.exe:** The main Executable File. Current Version 1.00.00

**ucfilter.dat:** The core Data file which defines all the languages and Font types. Without this file or if this file is corrupt, the application will not run and will close with a message.

**fntmatchlist.dat** Optional font matching file as described in previous section. The application will run without this file.

**Unicode Convertor Reference.pdf:** This file.

**Uninstall.exe:** Uninstall File

**Uninstall.ini:** Uninstall configuration file.

Apart from the above no other files are distributed. Check your program files directory if you suspect corrupt install.

# Sample fntmatchlist.dat file contents

---

Font Matching Listing

```
//*****Start of Comment*****

//This file is only used for RTF text. For text file conversion, this is ignored.

//First line is always the header. Do not remove it. It is defined as above.

// Use // for all comments.

//Each entry must be separated in its own line.

//Enter the full name of the font file followed by | and font encoding

//Defined Font Encoding are as follows:

//ta_TAM, ta_TAB, ta_TSC, ta_ADHA, ka_SHREE, ka_NUDI

//More font encodings will be added later, as more requirements arise.

//ta stands for Tamil and ka stands for Kannada.

//Thus it is possible to add same font encoding for multiple languages.


//Note: When providing the encoding ensure Character Mapping are same.

//Improperly matched font will provide be garbled output.

//For ex: Adhawin font is used as character mapping. Before you map a font to Adhawin

//ensure the characters in Character Mapping are same.

//If in doubt, use Character Map program to verify in Windows.


//Encoding is not Case Sensitive. Upper or lower or mixed case will be considered

//Font name is also not case sensitive and can have space in between.

//For font matching provide the best combination. For ex if you provide just Nudi -e

//then all font name that starts with Nudi -e are matched.


//Optionally you can provide output font that can be used in RTF file.

//If no font is provided, default font used will be "Arial"


//See help files for more details.

//(c) Kamban Software Australia. Last Modified 15th February 2016

//Sample listing is provided below

//*****End of Comment*****
```

MaduramTSC | ta\_TSC | Maduram

Adhawin-Tamil | ta\_ADHA | Kalyani

Shree851 Lipi | ka\_SHREE | Tunga