

**Nitartha *international*
White Paper**

WebTibetan:
Tibetan-Language
Documents on the World
Wide Web



**“Ancient Wisdom
for the Modern Mind”**

www.nitartha.org

WebTibetan

Due to recent advances in web standards and technology, Tibetan-language documents can be displayed online quickly and easily in freely available Web browsers. Web browsers have overcome initial obstacles preventing a display of documents with Tibetan-language fonts and Tibetan letter-forms (known as “glyphs”). Web browsers now have the ability to display Tibetan-language documents using Tibetan fonts. Proprietary technology is not needed. The Tibetan documents can then be searched using standard word processing search techniques for Tibetan-language documents, and can also be downloaded and edited in Tibetan-language word processing systems.

In this whitepaper, Nitartha *international* shows you:

- Examples of Tibetan web display, with an explanation of system requirements for seeing WebTibetan in your browser;
- How to create your own Tibetan language documents for the web,
- Learn more about the standards that make WebTibetan possible;
and
- Troubleshoot issues that may arise in creating such documents.

On-line WebTibetan examples

The Nitartha international site contains examples of Tibetan-language display, using the Nitartha-Sambhota display font and web documents created for Tibetan-language document word processing, and Microsoft Word for Windows95. In case you do not have the Sambhota program, we are providing a Tibetan-language font for free download. Download the font, unzip the file and follow your system's font installation instructions.

The WebTibetan examples include:

- an excerpt from Khenpo Tsultrim Gyamtso Rinpoche's Presentation of Lorik and Takrik (Nitartha international Publications, 1998).

- an excerpt from the Wangchug Dorje's Mahamudra: The Ocean of Definitive Meaning (Nges don rGya mtsho) (the English translation is due for publication from Nitartha international Publications January 1999)

To view the examples, your web browser must meet the WebTibetan viewing requirements.

WebTibetan Viewing Requirements

There are two requisites to view WebTibetan at Nitartha *international's* site:

1) you must have a browser which correctly displays fonts resident on the client browser operating system (invoked via the "font face" tag),

We have successfully viewed the examples using Microsoft Internet Explorer 3.0 and 4.0 for Windows. There is some problem in the implementation of the fontface tag in Netscape Navigator and Communicator 3.0 and 4.0 for Windows with respect to display of Nitartha-Sambhota. We are continuing to test it.

2) you must have the Sambhota true type font named Samw (file name "SAMW.TTF") loaded in your native operating system.

You may freely download the samw font from the [Nitartha *international* website](#). Unzip the compressed file and install according to your system's font installation instructions. (Those with the Sambhota system already installed may also view the samples, although one character displays incorrectly, as noted below.)

A note on creating WebTibetan documents:

Use of the samw font, or any specific proprietary font, is not required for the display of Tibetan text; in theory, any font that a browser can display will work equally to display Tibetan. For a detailed explanation of how to create WebTibetan documents, go to the next page on "How to create WebTibetan."

Samw is in fact a version of the native Nitartha-Sambhota program font named "Sama". The Nitartha-Sambhota program is an update of Sambhota, one of the oldest and most widely used Tibetan word processing systems. In our development program. We are still working to further enhance the product.

For those who have Sambhota installed on their system, the Sama font will be called as an alternative Tibetan font. Sama will work with the samples listed above with the exception of one character, and the Tibetan documents may be displayed without the necessity of downloading Samw. However, the Samw font provided for download from this page fixes a bug in the Sama display with respect to that broken character. Learn how to create WebTibetan.

Do not be disturbed by the seemingly unreadable code. The Roman characters outside the brackets (highlighted in blue below) "<>" are not part of the HTML code, but Tibetan characters converted into Roman script for HTML purposes.

 **ŸP-FÉC-«äc-®é_-nÉ-\T-tÉN-Eäc-
NéP-"-**

UWê]l-£ÉE-Ré-zÉP-`c-]éN-]zé-T[ÖCc,

The Tibetan script was entered as Tibetan in the original document. One does not need to view the converted Roman character nonsense version at all. Rather, all one need do is use a Tibetan language word processor to prepare a document, and then, in Word, simply save it as an HTML document, which converts the Tibetan script automatically into HTML format and preserves the fontface tags.

The HTML conversion process replaces the Tibetan glyphs with Roman glyphs for purposes of the source code, but reverses the process to display the Roman characters as Tibetan glyphs for display in the browser. This process is similar to what is happening at the system level for any word processing display. (See below for a discussion of the tools we use at this site.)

Note that the above Tibetan text is an image of how the text looks.

Specific tools we use to prepare our WebTibetan examples

The Tibetan documents displayed on this site were prepared using the Nitartha-Sambhota Tibetan-language word processing system and Microsoft Word for Windows97. In Word, the documents are displayed in their native Tibetan. After the example documents were created in their original Tibetan using Nitartha-Sambhota's native Tibetan fonts, and we simply saved them as HTML documents. Word automatically added the proper HTML fontface tags to display the Tibetan correctly. We then made a minor modification to use Samw rather than Sama as the basic font, since Samw contains some modifications for the web.

Nitartha-Sambhota for Microsoft Word for Windows works transparently in Microsoft Word. Microsoft Word97 and above is designed to create native HTML documents (using the ".htm" rather than ".doc" extension). The combination of WinWord as a Tibetan document editor and html document creator is quite useful. Within Word, the documents retain their native formatting using Tibetan fonts, and can be displayed and edited within the standard Sambhota program. The documents saved as html documents are automatically precoded for display on the Web.

Nitartha-Sambhota is a very widely used Tibetan-language word processing systems, soon to be featured as the standard for display on the next Asian Classics Input Program. In our software development program, Nitartha international is retooling Nitartha-Sambhota for a future release in which we will enhance the program's ability to create, edit, and prepare Tibetan-language documents for Web display, as well as for traditional desktop publishing.

We have also successfully tested the Tibetan display with Marvin Moser's Tibetan for Windows.

Though the process was rather straightforward and easy, but there are some design issues you should be aware of if you intend to do extensive WebTibetan publishing.

As noted, moreover, there is the basic requirement that a font be loaded in the operating system which houses the browser. There are developing standards which will address this problem. Please our discussion of Developing Standards for Tibetan Typography for more details about the web technology used to display Tibetan documents on the Web. At some time in the future, Unicode fonts will be available on all operating systems which will natively display Tibetan characters. For an intermediate solution which involves embedding or downloading fonts on demand, see our discussion of Sanskrit diacritics, where font distribution techniques are discussed applicable.

WebTibetan Standards

One major impetus to the development of universal standards for Tibetan typography is the World Wide Web. The ability to universally display Tibetan language text has recently been made possible by the adoption by leading Web browsers of HTML code ("tags") which allow the Web browser client to select any fonts installed on the operating system of the user. We have put examples of Tibetan text on the web in the form of Tibetan-language excerpts of the Ngedon Gyamtso and Lorik and Takrik.

Microsoft Internet Explorer 3.01, 4.0 and above and Netscape Navigator 4.0 and above web browsers now support the font tag, which allows the easy display of Tibetan fonts on Web pages. If the Tibetan font is natively loaded on the operating system in which the browser is resident, an HTML tag ("font face") can call that font for use in the browser. The result is the magnificent, easy display of Tibetan documents directly on the Web using readily available tools. The hidden costs of the platform notwithstanding, this ability to display Tibetan-language documents on the Web is another revolution wrought by the Internet.

The font face tag used to implement the Tibetan display is part of rapidly developing HTML standards for display on the Web. Until a few months ago, Web browser technology limited font selection in a drastic manner, so that all Web displays looked the same and used basically similar fonts. This limitation was intentional, to promote the universality of information viewable on the Web. However, the limitations of the fonts used also had drawbacks, as it precluded basic customization of pages and use of nonstandard glyphs. A rapid series of customization tags were implemented by Netscape and Microsoft, and the notion of using resident fonts began to find its way into HTML standards proposals, such as the original HTML 3.2 proposal, and the CSS-1 proposals. With the adoption of HTML 4.0 with the introduction of Cascading Style Sheet technology, font display technology is available not only for Tibetan, but also for Sanskrit diacritics.

Unfortunately, there continue to be problems with Netscape's implementation of the font face tag when used with certain Tibetan fonts.

For further information on these standards, please see the current standards for Cascading Style Sheets. Both Netscape and Microsoft have implemented parts of these standards in their systems since version 3 of those browsers, and the implementation is continuing to improve and be more complete. Microsoft Internet Explorer 3.01 and above has the most stable display of the font face tag in Windows95 and above. For a discussion of the mechanics of displaying fonts native to the operating system in a Web browser, see, for example, Microsoft's general discussion of typography and their specific discussions of Web typography and using the font face tag.

We have not run tests using Windows 3.1 versions of the browsers. We would be happy to have any input on the capabilities of different browsers. Please send your comments to our webmaster, who will forward them to the appropriate Nitartha personnel.

Issues for Web Designers

When using Microsoft Word and its HTML editor add-in, one can compose and display Tibetan language documents using such word processing program add-ins as Sambhota for Word. The HTML converter in Microsoft Word is thoughtfully designed to implement the font face tag so as to preserve fonts.

The remaining obstacle to Web display has to do with the nature of Tibetan font sets, which are implemented differently in each word processing system. As a result, the Web display of Tibetan documents still requires specific Tibetan fonts to be loaded on the client, and alternate fonts will not work between different font designers.

The problem with inconsistent implementation of incompatibilities from the use in nonstandard font sets of reserved characters also effects Web browser displays, and undermines perfect browser implementations of the font face tag. MS Internet Explorer has two characters that are incorrectly displayed arising from Microsoft's handling of the glyphs for single and double "curly" quotes. Netscape's implementation changes from beta to beta and from "point" revision to point revision.

Further, in each of the browsers the Tibetan text will not wrap on the tsek character. This is not specific to the browsers, but results from a change made to the way the hyphen character is handled. Previously, lines would automatically break at the hyphen, so a hyphen character slot was used for the tsek. At some point Microsoft changed the operating system to preclude such automatic line breaks. Accordingly, the Tibetan text lines at this time must be scrolled or mechanically broken for display. (The next release of Nitartha-Sambhota will fix this problem.)

Finally, there are size and alignment inconsistencies between Tibetan fonts and Roman fonts, as can be seen in the table of contents display in the Lorik and Takrik.

The surprising aspect of these issues is that they amount to relative glitches in comparison to the labored development of Tibetan word processing programs in general. The bottom line is that the future of Web display of Tibetan-language documents looks very bright. Given the scattered locations of the Tibetan scholarly community, the potential for universal access can only provide a great benefit to the preservation and strengthening of Tibetan scholarship.