

Creating Macrons for the Internet

Authored by Karaitiana N Taiuru – karaitiana@taiuru.maori.nz

Authored by Karaitiana N Taiuru

ISBN: 0-9582597-3-9

This document is © 2002 Karaitiana N Taiuru. This document may be freely redistributed in it's entirety.

Contents

Creating Macrons for the Internet.....	1
Contents	2
History of the macron in the Māori language	3
What is a Māori font	4
What is a keyboard driver (macron keyboard)	5
Where to obtain Māori Keyboard drivers	6
What is a Māori Font	7
ISO Latin Extended-A / Unicode / UTF8	8
UTF-8 macrons for the Web	9
Macron Display Problems?.....	10
Internet Explorer	10
Netscape Navigator.....	10
Multilingual and Unicode Support for browsers	10
Macintosh.....	11
Printing Unicode characters	13
Unicode Character Description Tag.....	14
Unicode in Meta Tags.....	15
Unicode code chart	16
ISO Latin1/Unicode	17
HTML Code for Umlaut	18
Unicode in Browser Comparisons	19
Browser Information.....	21
Internet Explorer Fact Sheet	21
Netscape Browser Fact Sheet.....	21
Lynx Fact Sheet	22
Opera Browser Fact Sheet	22
Changing your browser options to read macrons correctly	23
Embedded macrons for older computers	24
Macrons in Web addresses/URL's	25
Macrons in Meta tags.....	26
How do Search Engines interpret macrons.....	27
Using Graphics for macrons.	28
Cascading Style Sheets Font Family (CSS2).....	29
Weft32.....	30
References.....	31

History of the macron in the Māori language

Before the arrival of the Pākehā the Māori language was primarily oral, possessing no formal written code. Early missionaries, keen to have the scriptures written in Māori, developed a system using 10 consonants and 5 vowels. From a linguistic perspective the system was very good however its one major shortcoming was its inability to mark the vowel length. Consequently early writings in Māori did not distinguish vowel length.

Vowel length in Māori is phonemic, i.e. vowels are either pronounced short or long. Meanings of words can alter dramatically depending on which vowels in the word are or aren't lengthened. Correct pronunciation can only be learnt if correct vowel length is displayed. It is indisputably agreed that it is essential that vowel lengths be indicated in the writing of the Māori language and that their usage be consistent.

The lengthened vowel was first displayed in the keywords of Williams's Dictionary For Māori Language in its 5th edition, published in 1917. The symbol decided upon was from the Latin character set (which also has a lengthened vowel). It was a short line or bar over the vowel signifying a long length and an absence of the bar signifying a short length.

The selection of this symbol has caused frustration in the ensuing years as typewriters and early computers could not reproduce the symbol easily. Furthermore the 5 symbols for the lengthened vowels do not appear in a standard 8 bit ASCII character set. Fortunately this is not a problem with today's computers. Virtually all computers in use today, when supplied with a few simple software tools can easily display, and easily type the correct macron symbol to signify the lengthened vowel.

This convention of displaying the lengthened vowel with a macron character is one that has been encouraged by the Māori Language Commission. It is the convention that is used in the overwhelming majority of educational institutes and it is the convention that will be supported in this web site.

Author Te Taka Keegan.

What is a Māori font

A Māori font is exactly the same as an ordinary font except it has the ability to display a macron by responding to a users command to insert it.

A Māori font replaces two dots/umlaut into a macron symbol.

It should be noted that a Māori font can correctly display all English characters, thus it can be used for both Māori and English.

What is a keyboard driver (macron keyboard)

A keyboard driver is a small piece of software (sometimes only a few k in size) that reconfigures what characters are displayed when certain keystroke sequences are entered on a keyboard. Keyboard driver is the correct technical term for this software but is often referred to as a keyboard macron.

The Māori language keyboard drivers do not effect any English character sequences.

Once a Māori font is installed on a computer the easiest way to type the macron symbol is with the use of a keyboard driver.

Where to obtain Māori Keyboard drivers

The author of this paper has created two Māori keyboard drivers. A UTF-8 (true macron) and a Umlaut keyboard (two dots that require a Māori font to display a true macron).

The two drivers and other exciting Māori language editing software can be found at <http://www.maoriSpellChecker.net.nz>

What is a Māori Font

A Māori font is exactly the same as an ordinary font except it has the ability to display a macron by responding to a users command to insert it.

A Māori font replaces two dots/umlaut into a macron symbol.

It should be noted that a Māori font can correctly display all English characters, thus it can be used for both Māori and English.

The Māori Language Commission's view about macrons

The Māori Language Commission (Te Taura Whiri i te Reo Māori), established in 1987 under the Māori Language Act 1987, has been given the statutory responsibility of promoting Māori as a living language.

The Commission sees itself as having two principal objectives. The first objective is to sustain and nurture the existing language, the second to extend the linguistic range of Māori by weaving together into new combinations current speech, and words and phrases which have fallen out of everyday use.

A full version of the conventions can be obtained from Te Taura Whiri i te Reo Māori. Ph (04) 471 0244 ,
http://www.tetaurawhiri.govt.nz/english/pub_e/conventions.htm

A quote from the conventions follows:

Māori ORTHOGRAPHIC CONVENTIONS

The following is a set of writing conventions that the Māori Language Commission recommends is observed by writers and editors of Māori language texts.

MACRON

A. Long vowels are marked by the macron:

B. Loan-words (from English for the most part) follow the same rules as indigenous words:

C. Where two vowels of like quality fall together at a morpheme boundary, the two vowels are written separately in place of a macronised single vowel.

ISO Latin Extended-A / Unicode / UTF8

ISO Latin 3/ Latin Extended A/ UTF-8 or more commonly Unicode is a major breakthrough for writing in the Māori Language.

Unicode has the correct code to create a true macron without the need for Māori fonts. This document uses Unicode for all macrons.

UTF-8 macrons for the Web

More details of what Unicode is -

<http://www.unicode.org/unicode/standard/WhatIsUnicode.html>

In HTML view, insert the code into where the macron should go , for example for the word Māori in html would look like this Māori .

Macron Display Problems?

If you are having problems viewing Unicode macron characters it could be because of a number of reasons:

1. your browser
2. the operating system that needs a minor adjustment
3. your system does not support Unicode
4. you may need to download a Unicode font.

First thing to do is check the browser by viewing the relevant browser section below.

You should make sure that you are using the most recent version of your browser. This can be achieved by checking the following URL's:

www.microsoft.com/ie

www.netscape.com

www.opera.com

www.mozilla.org

The following describes how to configure browsers for different fonts.

Internet Explorer

Select the following menu's Tools / Internet Options / Fonts, then select the desired fonts.

Netscape Navigator

Edit / Preferences / Fonts. It is particularly important to set default fonts for Unicode. These fonts will be used when NN encounters documents encoded in a Unicode variant.

Multilingual and Unicode Support for browsers

Setting up Windows Internet Explorer 5 for Multilingual and Unicode Support
<http://www.hclrss.demon.co.uk/unicode/explorer.html>

Setting up Windows Netscape Web Browsers for Multilingual and Unicode Support
<http://www.hclrss.demon.co.uk/unicode/netscape.html>

Setting up Macintosh Web Browsers for Multilingual and Unicode Support
<http://www.hclrss.demon.co.uk/unicode/macbrowsers.html>

Operating Systems Windows | Macintosh | Unix | Internet Explorer | Netscape Navigator | Multilingual and Unicode Support for browsers
Windows

For Windows 2000, Start / Settings / Control Panel / Regional Options / General tab, set desired language(s) This adds fonts as well as system support for these languages.

Tuned fonts: Go to Start / Settings / Control Panel / Add/Remove Programs / Windows Setup / Components / Multilanguage Support / Details and check the languages you want.

Also useful is to go to <http://windowsupdate.microsoft.com> and install the Explorer language packs, which add the fonts to your system.

Full fonts: Microsoft Office 2000 CD-Rom has Arial Unicode font, which is the most complete. Obtainable by performing a custom install.

Choose Add or Remove Features. Click the (+) next to Office Tools, then International Support, then the Universal Font icon, and choose the installation option you want.

Macintosh

Language Kits are installed using Mac O/S 9.0 installation CD. Launch the Mac OS Install application. Proceed through the initial screens, selecting the appropriate boot disk. When you reach the "Install Software" screen, click on the "Customize" button. This opens up the Custom Installation and Removal dialog box. Scroll down to "Language Kits." Click on the check box, and then select "Customized Installation" from the installation popup to the right. (It will say "None selected" at first.)

This brings up a dialog box with a list of all the available language kits. Select the ones you want, or use the menu at the top of the dialog box to select all of them. Proceed with the installation.

If you already have Mac OS 9.0 installed, you will be asked if you want to add or remove software after you select the installation disk. Click on the "Add/Remove" button. This will bring you to the Custom Installation and Removal dialog box.

The installation procedure is the same for Mac OS 8.6, except that you will be installing "Multilingual Internet Access" instead of Language Kits. For Mac OS 8.5.5 and earlier, it will be necessary to purchase the individual language kits.

Unix

A range of quite comprehensive fixed-width Unicode on-screen pixel fonts for X11/Unix users can be downloaded from <http://www.cl.cam.ac.uk/~mgk25/ucs-fonts.html> or directly as <http://www.cl.cam.ac.uk/~mgk25/download/ucs-fonts.tar.gz>

Check the included README file for detailed installation instructions.

An earlier version of these fonts is already automatically installed when you use the XFree86 4.0 X server, which is the one commonly used under Linux.

The Mozilla and Netscape 6 web browsers can make use of these fonts directly and are highly recommended for anyone interested in utilizing Unicode web pages (make sure to use the very latest version though). Just select the "-misc-fixed-iso10646-1" font for the "Unicode" category in the "Edit|Preferences|Fonts" setup menu.

The Netscape 4.x browsers cannot handle 16-bit Unicode fonts at all. However, the above package contains scripts to generate 8-bit fonts in all ISO 8859 variants, which Netscape 4.x will then use to display those Unicode characters that are also found in ISO 8859.

Printing Unicode characters

Although your computer can read Unicode, some older printers are not capable of displaying Unicode characters. The reason is the printer (PostScript) not your computer.

If initially you can not print the Unicode try the following:

1. Download a patch from the Printers web site.
2. Change the Printer Properties as described in a, b and c below. Please not that different printers have different properties. But a little guessing and initiative you should be OK. If not, maybe a post to the macron discussion group will help.

a) Select File/Print

b) Select - Never "substitute" printer fonts in place of the real fonts (or)

c) Select - Download fonts (or print TrueType as Graphics)

If trying to print from a browser (a web site) and the above does not work, try copying and pasting into Microsoft Word 97 or higher and printing from there.

Unicode Character Description Tag

In order for a browser to recognise that you have used Unicode in your page/site you must include the Character set Description tag at the top of each page. By default if you are using an editor it usually puts it there by default.

You should check it before publishing. IE 5 will recognise the Unicode without the tag but many browsers will not.

The tag required looks like this:

```
<meta http-equiv="Content-Type" content="text/html; charset=utf-8">
```

Unicode in Meta Tags

With the release of a number of Unicode keyboard drivers, Unicode should be used in the Meta Tags as secondary text as some search engines do not recognize Unicode.

If using Unicode in the Meta you should not use numeric code but the raw code.

Unicode code chart

In HTML view, insert the code into where the macron should go , for example for the word Māori in html would look like this Māori .

Ā
Ē
Ī
Ō
Ū
ā
ē
ī
ō
ū

ISO Latin1/Unicode

ISO Latin 1 is no longer recommended. Please read about UTF-8 first.

Type the code in place of a macron i.e. MÄori to give you Māori. Note that the a has two dots over it instead a straight bar.

If the person reading your Web page has a Māori font as their default font they will see a true macron i.e. “M□□ori”. If not they will see the two dots, which is arguably the next best thing to a macron. Many people who use Māori fonts in their word processors etc will notice that the two dots means they used a Maori keyboard but not a Māori font.

CAUTION: Use font sets that are common and not custom made by you that no one else owns. If you use custom fonts then others reading your document with you're your custom font will more than likely see two dots.

Character Code

Ä Ä

Ë Ë

Ï Ï

Ö Ö

Ü Ü

ä ä

ë ë

ï ï

ö ö

ü ü

HTML Code for Umlaut

Macrons are displayed correctly in Web browsers when the macron text has been created using ANSI code as seen in Table 2 below. The persons Web browser also must have a macron font installed as the default font.

Character code

ä ä

ë ë

ï ï

ö ö

ü ü

Ä Ä

Ë Ë

Ï Ï

Ö Ö

Ü Ü

Unicode in Browser Comparisons

Browser	Version	NZ %	Overseas %	All %
Microsoft Internet Explorer	4.X	40.6%	39.7%	40.1%
	3.X	14.5%	11.5%	13.4%
	5.X	14.8%	8.6%	12.5%
	2.X	2.1%	1.9%	2.0%
		72.1%	61.8%	68.1%
Netscape	4.X	17.4%	23.4%	19.4%
	3.X	8.5%	5.9%	7.5%
	2.X	0.9%	0.9%	0.9%
	1.X	0.2%	0.1%	0.2%
	5.X	0.0%		0.0%
		27.1%	30.2%	28.0%
Other		0.9%	8.0%	3.9%
		0.9%	8.0%	3.9%

OS	Version	NZ %	Overseas %	All %
Windows	95	51.3%	47.4%	49.6%
	98	31.6%	26.3%	29.5%
	NT	5.7%	8.7%	6.8%
	3.X	3.2%	2.3%	2.8%
	CE	0.1%	0.2%	0.1%
		91.8%	84.9%	88.8%
Macintosh		6.5%	4.6%	5.7%
		6.5%	4.6%	5.7%
Unix	Linux	0.2%	0.2%	0.2%
	SunOS	0.1%	0.2%	0.1%
		0.3%	0.4%	0.3%
Other		1.4%	10.1%	5.1%
		1.4%	10.1%	5.1%
OS/2		0.1%	0.1%	0.1%
		0.1%	0.1%	0.1%

Browser Information

Internet Explorer Fact Sheet

Internet Explorer 5.x

Reads Unicode in numeric formats. If you add the Unicode meta tag Internet Explorer 5 will also read raw Unicode.

Internet Explorer 4.0

Reads Unicode in numeric formats. If you add the Unicode meta tag Internet Explorer 5 will also read raw Unicode.

Internet Explorer Does not read Unicode Troubleshooting

If you have a Unicode compatible browser and are still having trouble viewing macrons. In IE 5 go to Tools/Internet Options/General/fonts and select the Unicode option. IE5.0 treats UTF-8 as a character set, but doesn't have a provision for assigning a specific font for it. This basically requires that the default encoding should be assigned a Unicode-compatible font; in most cases this will not be a problem.

Netscape Browser Fact Sheet

Netscape Navigator 6.x

Reads Unicode. But it is advisable to add the Unicode meta character tag.

Netscape Navigator 4.x

Reads Unicode. But it is advisable to add the Unicode meta character tag.

Netscape Navigator 3.x

Unless a registry fix is applied then no, Netscape Navigator 3 is not capable of reading Unicode. not capable of interpreting UTF-8 encoding.

The following registry key could be added to rectify this. We make no assurances as to its efficacy; however, should it be found useful, it could be distributed to Windows users as a *.reg file; allowing them to download it and make the change with simply by double clicking on the file.

Implementing Unicode in Navigator 3.x:

HKEY_CURRENT_USER

Software

Netscape

Netscape Navigator

INTL

UseUnicodeFont "1"

To make this change, it is necessary to navigate to the Netscape Navigator key in the registry identified above, then add a new key name INTL, then to that key add a new string: UseUnicodeFont and give it the value 1.

Navigator versions 1.x and 2.x

Is not be capable of reading Unicode.

Troubleshooting

As discussed previously, the user will need to have appropriate fonts assigned to the Unicode or UTF-8 encodings.

If you see a question mark (?) instead of a macron vowel you need to follow the steps below:

- a.) Select the View menu
- b.) Select Character set
- c.) Select Unicode (UTF - 8)

By default Netscape comes with the "system" default font (eg: Times Roman on Windows) assigned to all encodings. This may be sufficient as it includes the required characters. If this is not the case, an appropriate font will have to be assigned to this encoding.

These font assignments are made via the Edit menu, specifically:

Edit -> Preferences -> Appearance -> Fonts

There is also a mechanism for setting the "default" encoding, for those occasions when none is forthcoming from the server or the document itself.

This is accessible from the view menu:

View -> Character Set [!] or in early models: View -> Encoding

This encoding selector keeps a selection "current" only until the browser is shut down or another page is parsed which does include a charset directive.

To maintain this "default" encoding permanently, select Set Default Character Set at the bottom after selecting your preferred encoding.

Lynx Fact Sheet

Lynx is standards-compliant browser. It will read raw and numeric Unicode pursuant to the operating system.

Opera Browser Fact Sheet

Version 6 of Opera supports Unicode. All previous versions prior to Version 6 do not support Unicode.

Opera for Windows

Opera for Linux (6.0 TP3)

Opera for Mac

Opera for Symbian OS

Opera for QNX

Opera for OS/2

Changing your browser options to read macrons correctly

All you need is some Māori fonts like the ones for download on this site or get your fonts macronised for free.

You have to change your default font to a Māori font or another macronised font.

Below are instructions for changing the default font for Microsoft Internet Explorer 4 and 5. These same instructions may work for earlier versions i am not sure.

1. Open Microsoft Internet Explorer
2. Go to Tools/Internet Options (version 4 - View/Internet Options)
3. Click General
4. Click Fonts (This will open a dialogue box see Diagram A below).
5. Under the heading Web page font, click on Aerial Māori. (If your default font is different select its equivalent Māori font.)
6. Under the heading Plain text font, scroll down to Courier Māori.
7. Press OK

Embedded macrons for older computers

Since Unicode is not supported by older computers you can use embedded Unicode/ASCII Māori fonts in you web pages.

More information can be found Britstream <http://www.truedoc.com/webpages/intro/>

An example of a page that has used embedded macron fonts is <http://www.olelo.hawaii.edu/OP/resources/webfonts2.html> .

Some web sites are not designed for older software. So at the end of the day it is a decision up to the web writer whether or not to use Unicode, embedded fonts or the less desired ASCII.

Macrons in Web addresses/URL's

At the time of writing this section there is/was one main factor governing the reason why you can not include macrons in web addresses.

The domain naming standards of (world wide) state that an URL/Web address/domain name must contain the following: character set is lower case a-z, 0-9 and an hyphen. This excludes any possibility of including a macron in your URL/domain.

Other problems that could occur is people not knowing where a macron is supposed to be placed. It has also been suggested that people may not have the right software to create a macron. If you have looked at this page you will see there is plenty of free software available and easy code to type macrons.

If it were possible to have an URL with macrons it would also be advisable to register that same URL without the macrons and have it re-direct to your macronised address. This would almost be perfect. But it can be very expensive, especially if you are an home based not commercial web author. The cost i think would not be worth the hassle other than for some businesses and organisations. At the time of writing it is about \$85 plus GST (NZ) to register a domain name for one year and then there is the charge from your ISP for monthly rentals and set-up fee. The cheapest i could find in a quick survey was \$30 set up fee and \$30 a month rental.

If you have a Māori web that is written only in Māori, there should be options for you to have a macron in your URL. The intended audiences for these pages would know where the macron (s) are placed. There is probably a 99% chance if a person does not know where the macron is placed they will not know how to read the page either.

Macrons in Meta tags.

In brief, Yes! use macrons in meta tags ! .

Search engines can and do interpret macrons as macrons (have not tested Unicode in meta). It must also be remembered that some search engines ignore some of your tag if it is too long. So, if your meta tag is getting too long (one search engine maximum is 20 words) you will have to decide how likely is it that people searching for your site will use macrons or not.

I personally perform a search using macrons when applicable. This means that people who have used macrons get preference with me. I suggest we all, when applicable use macrons in Māori words when we use them in search engines like Yahoo etc. If we do this we are helping to promote correct orthography of the Māori language. I prefer to visit a site where the author has taken the time and respects the Māori language enough to insert a macron in their web site.

One example is <http://www.piperpat.co.nz/nz/maori.html> where they use macrons in their very large Māori link site.

The best suggestion is to add macronised and non macronised words in the Meta tag if it will fit.

Search engines are beginning to impose tough criteria regarding how they search your site. Presently this imposes no problems to macronised and non macronised meta tags. Just follow the simple rules of the Search engines. I can not imagine that it could affect the Māori language because of the way a macron can change the meaning of a word. If you are curious about what people have typed into the Search engine to access your page some web counters tell you where the hit to your site came from and if it was a search engine, what the user typed in etc.

How do Search Engines interpret macrons

The more common search engines such as <http://www.google.com> will find your web site whether a person searches for the word with or without macrons.

For example i wanted to find my test web page on the net, which was written in Māori using macrons. I first searched for a specific word with a macron and found my page, i then searched for that same word without the macron and it found my macronised Māori web page. This was without using a Meta tag.

In the Meta use non macron word first then the macron version if you have room, if you are short on space in your meta do not use macrons. Because macron usage in the IT/Computer/Internet is still a new thing most people will probably search for keywords without macrons. I do not support this way of searching but it is inevitable in these early days of Māori IT.

Using Graphics for macrons.

This is a very bad idea. It will make your page or site un-searchable and extremely large. With Unicode there is no longer any reason at all to even contemplate using graphics.

To ignore the latest code and software technology and use graphics for macrons will include laborious hours of wasted time and difficulties. For most people it is desired as a sign of stupidity.

Obviously you would need to insert the graphics and then you would need to upload your graphics to the server. If you ever wanted or needed to change the font for your site you would have to create new graphics, delete the existing graphics on the server, then upload the new graphics.

Some internet users have an option turned off to view graphics. This would make a page using graphics for macrons eligible.

If you have say a 10 page web site with a page and an half per page it will take for ever to download the site or individual pages. Most people would not bother to wait. Personally I have only heard of the odd web page that utilises graphics for macrons. Then the authors became so fed up with the download time and editing problems they changed their graphics back to text.

Cascading Style Sheets Font Family (CSS2)

Cascading Style Sheets gives you the opportunity to select what font(s) you want the end users of your web site to use regardless of what their default font is. If your users computer does not have the font you specify, the computer will select the font that it thinks is the nearest match. So it is advisable to add more than one font to the code. This particular code is fully supported by Internet Explorer 3 and above, Netscape 4 and above, Opera 3.5 and above.

The following code will do the above task for Times New Roman Māori:

```
<STYLE TYPE="text/css" }  
BODY { font-family: Times New Roman Māori }  
</STYLE }
```

To add other fonts:

```
<STYLE TYPE="text/css" }  
BODY { font-family: Times New Roman Māori, Arial Māori }  
</STYLE }
```

Weft32

At the time of writing, the above program was still in a Beta version and had been for some time. This little utility enables you to embed fonts into your HTML page. This means that your audience can see your fonts as you do even if you have created a new font.

Full information and a download can be found at

<http://www.microsoft.com/typography/web/embedding/weft2/default.htm?fname=%20&fsize=>

References

1. Te Taka Keegan - Lecturer, School of Māori and Pacific Development, School Of Computing and Mathematical Sciences, University of Waikato, Hamilton, New Zealand <http://www2.waikato.ac.nz/smpd/taka/tetaka1.html>
2. Mark R. Laws - Maaka. Knowledge Engineering Laboratory - KEL. Information Science Dept - University of Otago - Rorohiko Whai Ki Aho-Putaiiao - Te Whare Wananga o Otakou for his co-operation and Macintosh advice, Also for use of his software "Aotearoa". URL <http://kel.otago.ac.nz/maaka/>
3. Keola Donaghy for his permission to use one of his Hawaiian fonts for download. URL <http://www.olelo.hawaii.edu/>
4. Andrew Cunningham, Multilingual Technical Project Officer, Library Network Unit, State Library of Victoria. Email: andrewc@vicnet.net.au or andjc@ozemail.com.au. <http://www.ozemail.com.au/~andjc/papers/> Currently managing the redevelopment of the Open Road <http://www.openroad.vic.gov.au/> project. For all his great ideas and support regarding Unicode.
5. Te Taura Whiri i Te Reo Māori , (The Māori Language Commission) Māori Orthographic Conventions document. <http://www.tetaurawhiri.govt.nz>