

Indigenous Language Presence on the Web – the Māori Example

Keegan, Te Taka, Cunningham, Dr. Sally Jo

University of Waikato,
Hamilton, New Zealand

tetaka@cs.waikato.ac.nz
sallyjo@waikato.ac.nz

He Mihi

He hōnore he korōria ki runga. He maungārongo ki te mata o te whenua.
He whakaaro pai, tētehi ki tētehi. Rire rire hau. Pai mārire. [1]

1. Introduction

In 1998 a survey was undertaken to determine the Māori language presence on the World Wide Web. Web sites with a significant Māori language presence were searched for and when found certain characteristics were recorded. It was disappointing to note there were relatively few Web pages written in Māori available at that time. 41 Web sites were found, with a total of 304 Web pages written either partially or totally in Māori. A similar survey was undertaken 4½ years later in 2002 [2]. The number of Web sites using the Māori language had little more than doubled (from 41 to 100). However the number of Web pages had increased 100 fold (from 304 to 30,346).

It is important to note that while the researchers were confident that they had found the majority of Web sites, it is possible that the strategies that they used did not detect all of the Web sites written in Māori. Furthermore the surveys are a snapshot in time; the transient nature of the Web means that the picture that was seen in November 2002 will be different from the current picture at the time of this conference, October 2003.

This paper will highlight some issues that arise from the Māori language Web surveys. Section 2 will look at some social issues that have become apparent and Section 3 will discuss some technical issues that have arisen. To complete the analysis Section 3, will also present a case study of one of the Māori language Web sites focusing on the earlier issues that have been mentioned and commenting on the usage of the Web site.

2. Social Issues

Purpose of Web sites in the Māori language

In the 1998 survey there were three predominant reasons for creating Web sites in the Māori language. The first was to provide information about the Māori language which included word lists, dictionaries and references to written material (19.5% of sites, 17.1% of pages). The second purpose was personal home pages where individuals provide personal information in the Māori language (19.5% of sites, 5.9% of pages). The third purpose was to make available historic information or historic texts (14.6% of sites, 41.1% of pages). There were a further nine purposes identified which included supplying information about Health & Development, General Māori issues, Songs & Video, Religion, Media, Nature, Education Resources, Tertiary and Finances. These accounted for the remaining 46.3% of Web sites and the remaining 36.6% of Web pages.

Over the 4½ year time period a shift was noted in the purpose of the Web sites. A new leading purpose emerged, that of education: Web sites providing information about education and/or supplying educational resources (26.0% of sites, 16.3% of pages). The second purpose was to provide information about the Māori language (12.0% of sites, 6.5% of pages) followed by the third purpose of making available historic information or historic texts (8.0% of sites, 61.8% of pages). There were also 8 school Web sites found with pages written in Māori (8.0% of sites, 0.1% of pages). Perhaps the emergence and increasing presence of the education sector in Māori language Web sites is a reflection of the changing political attitudes in New Zealand towards the place of Māori language in education. There were a further eleven purposes identified covering a wide range of topics and accounting for the remaining 46% of Web sites and the remaining 15.3% of Web pages.

Provider of Web sites in the Māori language

The New Zealand government was responsible for one in six of all Māori language Web sites in the 1998 survey. Four and a half years later central and local government were responsible for one in three sites; the National Library, a major provider of Māori language Web materials, is included in this grouping. This shows that the government of New Zealand is becoming more supportive of Māori language Web sites.

Coupled with this is the rise in support from education providers. In 1998 tertiary providers, education initiatives and schools (both primary and secondary) together account for 1 in 13 of all Web sites, where in 2002 this was 1 in 4. Since this is where some of the large Māori language collections are located, this grouping accounts for 63% of the total Web pages.

It was disappointing to note in 1998, that apart from some personal Web sites, Māori tribes and Māori organisations were not providing any Web material in Māori. Although the 2002 survey found 3 tribal groups and 4 Māori organisations/ interest groups that were prepared to make information available in Māori, this increase from 0 to 7% in terms of Web sites and from 0 to 0.4% in terms of Web pages is not a significant increase.

Funding

Funding is a significant issue in the creation and ongoing maintenance of Māori language Web sites. Some sites stated that they would like to make more information available and translate more texts from English into Māori for the Web but were prevented from doing so by funding constraints. Results from the 2002 survey show that the Ministry of Education provided funding assistance to 21% of the Web sites and 78% of the Web pages. These results are somewhat skewed as 3 sites providing very large document collections account for almost 75% of the total Web pages.

While government support for large collections of indigenous language documents is to be commended, a report developed by the Digital Collectives in Indigenous Cultures and Communities Meeting[3] stated that there was a concern that funding could reflect the Eurocentric values of the funding provider. Consequently sites seeking funding for the purpose of making Māori language material on the Web may find that the focus of the site is altered to satisfy the restraints placed on the funding.

Personnel

Although only stated on one site, another key reason why there are not more Web sites created in the Māori language is that there is a paucity of people who have a Māori language knowledge and a sufficient knowledge of Web development to be able to successfully build a Web site in the Māori language.

3. Technical Issues

Orthography

An important concern confronting writers of the Māori language in the Web environment is the ability to type and display the correct orthography of the Māori language. All Māori language characters can be displayed using a standard ASCII character set, apart from the 5 lengthened vowels which use a macron or bar symbol over the vowel to indicate that the sounding of that particular vowel is lengthened (for example the ā in Māori).

The surveys found four methods that were used to represent the lengthened vowel. The first method was to simply ignore the problem and not signify when a vowel is lengthened (for example, 'Maori'). The second method was to type a second vowel whenever a vowel needed to be lengthened (for example, 'Maaori'). The third method was to use *Māori fonts*—that is fonts that had the umlaut character altered so that it appeared as a macron character. The fourth method was to use a macron character defined in a Unicode compliant character set like UTF8.

In both the 1998 and 2002 Web site surveys the older the site the more likely it was to use an older method of representation, e.g no representation at all or a double vowel method.[4] Second, the larger the site the more it was committed to using the Unicode character set. In terms of Web pages the 2002 survey showed 62% of Web pages used the Unicode macron character and 30% uses no representation at all. In terms of Web sites 54% used no representation whereas 24% used the umlaut character with the altered font and 21% used the Unicode character.

In 2000 Te Puni Kōkiri (*Ministry of Māori Development*) produced a report titled, "The Macronisation of Web Content" [5]. The report recommends full compliance with the Unicode Character Set UTF8 to represent the macron character. In recognition that operating systems and application software are not completely compliant with the UTF8 character set, the report suggests using 7 bit ASCII with 'numeric character references' to define characters outside of the standard character set. Te Puni Kōkiri then proceeded to use its recommendations to good effect on its own Web site.

Navigation

An important consideration when analysing the Web sites was the ability to navigate in the indigenous language— that is, checking to see if the hyperlink labels or buttons were written in Māori and if the scroll over texts were displayed in Māori. It was noted that navigation throughout the site in Māori had increased significantly from the 1998 to the 2002 survey. In 1998 only 7 sites (17%) supported Māori language navigation while in 2002 this had increased to 42 (42%). In general the larger the site (in terms of Māori language Web pages) the higher the chances were that the site could be navigated completely in Māori.

The increased use of Māori in navigation throughout a Web site could be due to two factors. As Web literacy increases over time, and creators of Web sites become more skilled, then it follows that it must be easier for them to create more Māori language pathways, including buttons and hyperlinks in Māori. Also, with larger Web sites perhaps there is more

funding and time allocated to the Māori language component and consequently a greater likelihood that matters such as navigation in Māori are likely to be addressed.

Although in 2002 there was an increased usage of Māori in navigation it was actually quite rare to find a site completely written in Māori. There was invariably a word, a phrase, a button or a copyright notice written somewhere in the more dominant language in New Zealand, English. Many of the Web sites were bilingual, written in Māori and English, and most times when this was the case the Māori language version of the site was not as extensive as the English language version. Further, when navigating a bilingual site in Māori, the reader is often transported into the English version of the Web site, sometimes due to incorrect coding of hyperlinks, others times due to the text only being available in English.

Multimedia

Traditionally the transfer of knowledge in Māori society occurred both orally (prayers, songs, chants, genealogy recitals, stories and speech making) and visually through artwork such as painting, carving, weaving, embroidery and tattooing. The increasing ability of the Web to make available multimedia resources has not gone unnoticed by the creators of Māori Web sites. In the 1998 survey the sites made very little usage of graphics and there were few sound files available. However in 2002 most sites contained imagery that was distinctively Māori, and many sites had obviously put a lot of time and effort into creating attractive Māori artwork. Further, the 2002 survey noted 704 Māori sound files at 15 Web sites containing speech, song or chant, and almost all were recorded from a native speaker. As revitalization efforts stress the importance of mastering the oral language, the ability to hear it over the Internet is particular important.

Interaction

A new aspect that was noted in the 2002 Web survey was the ability of users of sites to be interactive in the medium of the indigenous language. Facilities to allow site users to communicate with others in the Māori language were not noted in the 1998 survey although the facility was available early in the decade through bulletin board systems like Te Wahapū[6]. However in the 2002 survey Web sites were seen to offer the ability to interact by email in the Māori language, one site offered the ability to send greeting cards in Māori and another site had two bulletin boards that only used Māori language.

4. Case Study: The Niupepa Collection

The Niupepa Collection is a collection of historic Māori newspapers published between 1842 and 1933. It is a large source of historic legacy texts, almost 18,000 newspapers pages, that were predominantly written in the Māori language. The newspapers have been captured both in facsimile and full text form, and both forms are currently served by the Greenstone software[7] of the New Zealand Digital Library (NZDL) at: www.nzdl.org/niupepa. The collection is a valuable and desperately needed text source for scholars, teachers and students of the Māori language. For a comprehensive explanation of the process of delivering the Niupepa on the Web see [8].

The creators of the Niupepa Collection had three primary goals for the Web site: to make the collection freely available by placing it on the internet, to make the information in the collection more accessible by incorporating a full text search facility, and to encourage and promote the use of the Māori language. Previously the collection had been stored on microfiche where access was restricted by the physical restraints of the microfiche reader and research undertaken on the collection was very laborious due to the sequential access restraints of the microfiche media.

The provider of the resource is a tertiary institute, the Computer Science Department at the University of Waikato, who were interested in the task due to the many research aspects that it involved and due the significance that this collection could have to the Māori and indeed greater research community if the legacy texts were made more widely and freely available. The Computer Science Department and the NZDL were responsible for funding the initial and ongoing costs of the project, the Ministry of Education providing two significant grants to assist with the OCR process, and donations were also received from various libraries throughout New Zealand.

Two forms of orthography can be seen in the collection. The Web site itself through the usage of the Greenstone software uses the Unicode character sequence to represent the lengthened vowel. However when the original newspapers were written, approximately 100 years ago, there was no distinction used to signify the vowel length. This is perhaps due to the fact that at that time the majority of the readers were fluent in the Māori language and thus did not need to rely on these types of indicators to determine correct pronunciation. The decision was made by the site creators to retain as much of the original format of the newspapers as possible; thus the facsimiles of the newspapers do not show any indication of the lengthened vowel nor do the text files that were extracted from these facsimile files.

It is interesting to note that the orthography issue it not a new problem, the newspapers themselves discussed orthography. Disillusioned with the confusion regarding the pronunciations of the consonants *ng* and *wh* the editors of the newspaper *Te Pipiwharauoa*[9] decided in June 1908 to use the *ŋ* symbol to represent the *ng* consonant and they merged the *w* and *h* character to create a new symbol to represent the *wh* consonant. They continued using these symbols for the following 12 publications until May 1909 when they returned to using the *ng* and *wh* characters[10].

Navigation in the Niupepa Collection can occur in the Māori language, supported by the Greenstone software's facility for supporting multilingual content and interface elements. At this stage the Niupepa collection does not offer any of the resources in a multimedia format other than displaying the facsimile of the newspaper pages and a graphic showing the

mast head or logo of each newspaper. There is however the ability to seek feedback in the medium of the Māori language. An email query to the collection written in Māori is responded to in Māori by Māori literate members of the Niupepa project.

Usage

Site usage statistics were recorded as the site was being built, in the years 2000-2002, and since the site has been complete (2003). While there are a few gaps in the record they can still give preliminary insights into the usage of this Māori language Web site. They showed the average monthly hits for the site in 2000 were 6,823, in 2001 were 10,412, in 2002 were 15,720 and so far in 2003 have been 19,483 –a hit being defined as every click on the Web site. The Web site is obviously being used at an increasing rate and the initial goal of making the collection freely available to users seems to have come to fruition.

The newspaper pages are available through three different methods on the Niupepa Web site; the full text search facility, browsing through a list of titles, and accessing by selecting a time period. Since the completion of the Web site, the search facility has been the most popular method of accessing the pages accounting for 69.8% of the methods used to access the newspaper pages, followed by searching for title, 23.9%, and finally access by selecting a time period at 6.3%.

The third statistics that was of interest to the site creators was the usage of the site in the medium of the Māori language. The newspapers are primarily in Māori (70.4%), though 26.9% are written in both Māori and English, with the remaining 6.7% being in English. The default language of the Web site is in Māori, so that if users want to access the newspapers in English, or one of the other languages, then they have to change the language of the interface. However, the majority of the Māori population cannot speak Māori with only 22% claiming that they may have a reasonable conversation in Māori[11]. What then did our site usage statistics show? It showed that in results to date for 2003 that Māori language was the preferred language of usage for 46.1% of the hits. Clearly the site is encouraging and promoting the use of the Māori language.

It is important to note that usage statistics are only preliminary results and to get a good understanding of the usage of the collection user surveys need to be undertaken. Also Duncker[12] suggests that a standard Digital Library interface is counterintuitive to Māori users. However the early indications do suggest that design goals of the Niupepa site creators are being realised by the users of the site.

5. Summary

This paper has reported results of surveys and analysis that was undertaken to determine whether, using the Māori language as an example, indigenous languages have a presence on the Web. While there are some issues, both social and technical, it appears that possibilities do exist for indigenous languages on the internet though involvement from indigenous peoples themselves needs to be encouraged.

[1] Honour and glory above. Peace on earth. Goodwill to mankind.

[2] A comprehensive report on the surveys is due to be published in the 2003 autumn edition of *He Pukenga Korero*. Published by Te Putahi-a-Toi, School of Maori Studies, Massey University. See <http://maori.massey.ac.nz/He-Pukenga-Korero/>

[3] *Digital Collectives in Indigenous Cultures and Communities*, page 9, report on meeting held in Hilo, Hawaii, August 10-12, 2001. See <http://www.si.umich.edu/pep/dc/meeting/indigenous.pdf>

[4] Generally speaking there has been a historical progression in orthographical methods used. Early Māori language printing did not signify vowel length, perhaps due to a high Māori language fluency among the population of that time. The double vowel method was introduced because the need was realised to signify vowel length but computers of that time could not easily type and display the macron character. This led to frustrations with word length and sorting, consequently a third method was tried, that of altering fonts. However most computers released in the last 2-3 years are Unicode compliant and are consequently able to easily type and display the macron character without the need to use altered fonts.

[5] Ryniker, B. (2000). *Report on the Macronisation of Web Content*. Te Puni Kōkiri. October 2000. www.tpk.govt.nz/using/macron_paper/index.html sighted 26 Sep. 2003

[6] Benton, Richard. (1996). Making the Medium the Message: Using an Electronic Bulletin Board System for Promoting and Revitalizing Māori, in Mark Warschauer (ed.), *Telecollaboration in Foreign Language Learning*, pp 187-204. Honolulu: National Foreign Language Resource Center, University of Hawaii at Manoa.

[7] for information on the Greenstone software see: Witten, I H & Bainbridge, D. (2002). *How to Build a Digital Library*. Morgan Kaufmann. San Francisco, CA.

[8] Apperley M D, Keegan T, Cunningham S J, Witten I H (2002). Delivering The Māori Newspapers on the Internet in Curnow J, Hopa N, McRae J (ed.s), *Rere Atu Taku Manu! Discovering History Language and Politics In The Māori Language Newspapers*. Auckland University Press. Pages 211-36.

[9] *Te Pipiwharauoa He Kupu Whakamarama* Newspaper edited by Korohere, R T. Te Rau Press Gisborne. June 1908, issue 123, page 1.

[10] *Te Pipiwharauoa He Kupu Whakamarama* Newspaper edited by Korohere, R T. Te Rau Press Gisborne. May 1909, issue 134, page 11.

[11] Benton, R & Benton, N. (2001). Reversing Language Shift in Aotearoa/New Zealand, in Fishman, J (ed.), *Can Threatened Languages Be Saved?* Multilingual Matters Ltd. Clevedon.

[12] Duncker, E (2002) *Cross-Cultural Usability of the Library Metaphor*. JCDL conference proceedings. July 13-17 2002. Portland Oregon.