Full paper

Localisation in the Web: Speaking your Audience's Language

Pablo Matamoros
Bay of Plenty Polytechnic
New Zealand
pablo.matamoros@boppoly.ac.nz

Abstract

Most people, including technology professionals, don't know the meaning of the terms internationalisation and localisation. How many people know what a localisation specialist does? How many organisations localise their products?

An underrated field for most of the IT industry, often considered a "high tech translation" process (LISA, 2007), internationalisation and localisation are one of many career paths in IT.

This paper intends to promote awareness of the importance of the field by reviewing the technologies, standards and issues involved. It also discusses definitions provided by organisations like LISA (the Localization Industry Standards Association), GALA (The Globalization and Localization Association), TILP (The Institute of Localisation Professionals) and others.

A series of surveys are part of this research. The first (preliminary results are discussed in this paper) has been designed to study the apparent lack of interest many organisations have (even in today's global economy) in localising their websites. Following surveys will analyse methodologies and technologies used to internationalise and localise software and websites.

Keywords: localisation, internalisation, global content management systems.

Introduction

With a estimated population of 4.29 million inhabitants, New Zealand is a small market (National Population Estimates: December 2008 quarter - Statistics New Zealand, 2009). The local economy relies heavily on exports and tourism. Although often trade delegations are organised, to be part of them is simply not possible for many small businesses. Therefore their main contact with the international market is their websites.

Considering that the English language is used only by 29.4% of the internet users worldwide (Top Ten Internet Languages - World Internet Statistics, 2008), it would make sense that many businesses provide versions of their websites that cater for some of the remaining user groups.

This quality assured paper appeared at the *22nd Annual Conference of the National Advisory Committee on Computing Qualifications* (NACCQ 2009), Napier, New Zealand. Samuel Mann and Michael Verhaart (Eds).

For example, having a website translated to Mandarin Chinese and Spanish will make it accessible to 56.8% (8.5% Spanish, 18.9% Mandarin Chinese and 29.4% English) of the internet users. Of course, the relevance of translating content to one or several languages will depend on the product or service provided and the target market.

However, how many websites offering products and services, which could be sold with ease in non-English speaking markets, are actually translated? Are businesses aware of the costs and benefits?

We shouldn't forget that in New Zealand there are three official languages: English, Māori and Sign Language. Therefore many official websites might require to be available in Māori.

Translation is only a small part of the process of making available a product for different ethnic groups. There are other technical and cultural issues to solve. This process actually consists of two steps: internationalisation and localisation. Power software houses like Microsoft understand this field very well, having specific career paths for internationalisation and localisation specialists (Microsoft Careers — Technical — Localization).

1. Definitions

Many fields in IT have a clear cut definition. A network administrator deals with the administration of computing networks. A programmer writes the code for an application. Of course these are simplifications of the two more popular roles in New Zealand at the moment.

On the other hand the difference between internationalisation and localisation is not clear - at least for the general public, and even IT professionals. They are often used in the wrong context or not known. If there is a need to make software available for a non-English speaking country, would the development company hire a translator or a localisation specialist? Further, is there a need for an internationalisation specialist?

During the GALA Vendor Roundtable of 2004, Dan DePalma asked the audience: "How many of you use the word localization or the word translation for the services you provide?" The answer was evenly split. The conclusion was that specialists in this field should speak the same common language. Thus the efforts of organisations like GALA, LISA, TILP and many others to promote standards and certifications on the field. Let's review the formal definitions provided by them.

Globalisation involves the enterprise efforts that are necessary to launch a product or service internationally. It goes beyond the product itself, often including the revision of several business processes. It is abbreviated as

G11N which derives from the first and last letter of the world while 11 refers to the number of letters between the L and the N.

Internationalisation is the process of enabling a product at technical level for localization. This allows the product to handle multiple languages and cultural conventions without the need for redesign. In a similar fashion to **G11N**, it is abbreviated **I18N.**

Localisation is the process of modifying products or services to account for differences in distinct markets. It includes translation, but also customs, conventions, standards and other characteristics of a particular culture or region. It is abbreviated **L10N.**

In other words, globalisation includes all the business efforts towards the expansion of the market of a product. Internationalisation involves the creation of a design that makes it easy to adapt a product for different regions or ethnicities (called locales), while localization is the adaptation itself. Internationalisation is done once per product, localisation is done for each locale of the product.

For example, in an imaginary appliances factory, the colours and labels of a microwave could be changed according to the target market (locale) by moving a lever in the production line. The process of designing the mechanism behind the lever is called internationalisation. Localisation is the identification of the colours and labels to be used to suit each locale. The globalisation process involves all the efforts made by the factory to facilitate the sale of the microwave.

In the context of web development, our lever is a combination of programming code, data files (often XML) and database tables, while localisation is usually the actual translation of the content (text, video, graphics, etc.) of those data files and tables. The globalisation process would provide the necessary funds to hire internationalisation and localisation specialists.

2. Technologies and standards

While globalisation is more related to management and business theories, internationalisation and localisation are purely technical. They involve a set of standards and technologies that have evolved for many years, in some cases providing room for the development of very successful companies highly specialised in the field like SDL, GlobalSight and, in New Zealand, Straker Interactive.

The technology used is divided into two main groups: language technologies and tools to manage the globalisation process (Localization Industry Standards Association, 2007). We will focus on the first group only.

Language technologies allow translators to deliver a translated text in a shorter period of time. They don't replace the translator. Here lies one of the main issues: often these tools are thought as a replacement of the translator.

Language technologies are divided into three sub classifications: Terminology Management Systems (TMS), Translation Memory (TM) and Machine Translation (TM).

A TMS is a sort of dictionary. These applications usually consist of a base dictionary to which users add specific terminology for later use and to keep consistency among translation projects (often involving more than one translator). More complex TMSs allow the inclusion of synonyms, notes, context, etc.

TM (not to be confused with TMS) is a technology that follows the same philosophy as version control systems used in the software development industry. A text is divided into small pieces, usually sentences; every time this is changed the system searches for those sentences that have being modified and reports them. The idea is that it does not make sense to ask a translator to retranslate the whole content of a text when it has only a few changes, instead the software will tell the translator the sentences that need revision.

The best known tool of this group is Machine Translation. Unlike TM, MT actually translates the text by a combination of complex algorithms, dictionaries and TMS components. There are three main paradigms for MT (Jurafsky & Martin, 2008): direct, transfer and interlingua.

The direct approach translates text word by word. Transfer performs a syntactic and semantic analysis of each sentence, and produces an output by following a set of rules. Interlingua is an improvement on the transfer model. In a traditional transfer model there is a set of rules per each pair of languages, but this is not very practical in a many-to-many languages environment. In an interlingua system a sentence is first translated to a language-independent representation of the meaning. This "meaning" is then translated to one or many target languages.

The latest developments in MT incorporate statistical analysis in conjunction with a combination of the direct, transfer and interlingua model.

These language technologies are present today in two main software groups: Localisation Workbenches and Global Content Management Systems (GCMS).

Localisation Workbenches combine all three language technologies in a single desktop application. They are optimised for specific tasks. For example, some of them focus only on the translation of text documents (Word, PDF, txt files, etc.), others are designed specifically to translate software interfaces (Figure 1 - QT Linguist screenshot).

GCMSs are engineered to facilitate the publication and maintenance of the content of websites. Although there are many commercial GCMS, there is still a lack of them in the Open Source world. Most Open Source CMSs allow the integration of GCMS functionalities into the system, but at a cost, these are normally plugins or extensions that add an extra layer to the system. Recently GlobalSight launched an open source version of their system that runs under a Java platform (GlobalSight, 2009).

Figure 2 illustrates an implementation of a GCMS, a custom made web application designed for PetLinq[lq] (an educational website on the use of PET, a type of computed tomography). This system was developed on Django (a Python framework) adapting some of the localisation tools included with the framework. The left column of the screen shows the previous version of the text, the right column the new text and on the bottom is the current translation of the text. In this implementation, when a text is published for the very first time, the system uses machine translation to provide the translator with a rough translation of the text. Later the translator proofreads or very often rewrites the whole translation. This process is referred as Computer-Aid Human Translation (CAT or CAHT) (Jurafsky & Martin, 2008).

For all these technologies, LISA has developed standards through the OSCAR (Open Standards for Container/content Allowing Reuse) special interest group. These standards are: Term Base eXchange (TBX), TBX-Basic, Translation Memory eXchange (TMX), Segmentation Rules eXchange (SRX), Global information management Metrics eXchange (GMX), XML text memory (xml:tm) and Term Link. In August 2008 TBX was approved as an ISO standard (ISO 30042:2008).

Despite the complexity of the technologies and methodologies mentioned in previous paragraphs, the most common methodology for the creation of multilingual websites is the manual translation of the source code or the content stored in databases. According to the initial results of our survey, 55% of the organisations chose this methodology to localise their products (Figure 5). This implies use of the services of either a programmer who also acts as translator, a translator with some knowledge on web programming or both a translator and a programmer.

3. Business attitude towards localisation

The previous paragraphs illustrate that localising a website is not as simple as it might seem for most people. Then, is it worth all the effort? Should New Zealand organisations localise their websites?

In the case of official websites, many public organisations should localise to te reo Māori. In the 2006 Census of Population and Dwellings, there were 565,329 (14% of the New Zealand population) people who identified with the Māori ethnic group, 23.7% of them could hold a conversation about everyday things in te reo Māori. Nevertheless, not many government websites are localised to te reo Māori. The New Zealand government portal certainly is not nor the websites of many tertiary institutions with a strong Māori background like the Bay of Plenty Polytechnic or Waiariki Institute of Technology.

Businesses trading at international level should also localise their websites. As stated in the introduction, English speaking users account for only 29.4% of the total of internet users worldwide. Therefore it would make sense for many New Zealand businesses to have versions of their websites localised for different regions and cultures. However my personal experience (partially confirmed with the initial results of the survey) tells me that internationalisation/localisation is not considered; rather short term solutions like online machine translations are used.

In this last case, online products like Yahoo Babelfish, WorldLingo, Google translate, and other machine translations are thought of as the panacea. Unfortunately these applications, known as online automated translation, are far from accurate. A fluent and precise computer translation is still science fiction as it is a combination of Linguistics and Artificial Intelligence. Progress has been made in this field, but human translation is still the best solution. The day software can accurately translate a text we will be seeing a real Artificial Intelligence breakthrough.

4. Survey

4.1. Methodology

The objective of the survey was to study the level of "awareness" of the internationalisation and internationalisation fields. The target audience was anyone involved in the development of websites: developers, graphic designers, copywriters, marketers, business owners, etc. Follow up surveys will focus on the technologies used by those who indicated an interest in further participation in this study.

The survey itself was a localisation exercise. Forms were distributed in Spanish and English. The data had to be combined, therefore having tables matching the answers on both languages. Ideally the survey should have being translated to as many languages as possible due to the nature of the topic. Unfortunately, it has being limited to Spanish and English speakers for the moment.

The surveys (English and Spanish) were written with Google Docs and distributed among my personal contacts through email and Web 2.0 applications like Linkedin and Facebook. The survey was also published in my personal website (www.CodeSpanish.com) to increase the number of responses. Later in the research, it was distributed among local businesses throughout the Bay of Plenty ICT Cluster.

The collected data included quantitative and qualitative information and was analysed using Ms Excel.

At the time of writing this paper, 34 responses were collected: 13 for the English version and 21 for the Spanish version.

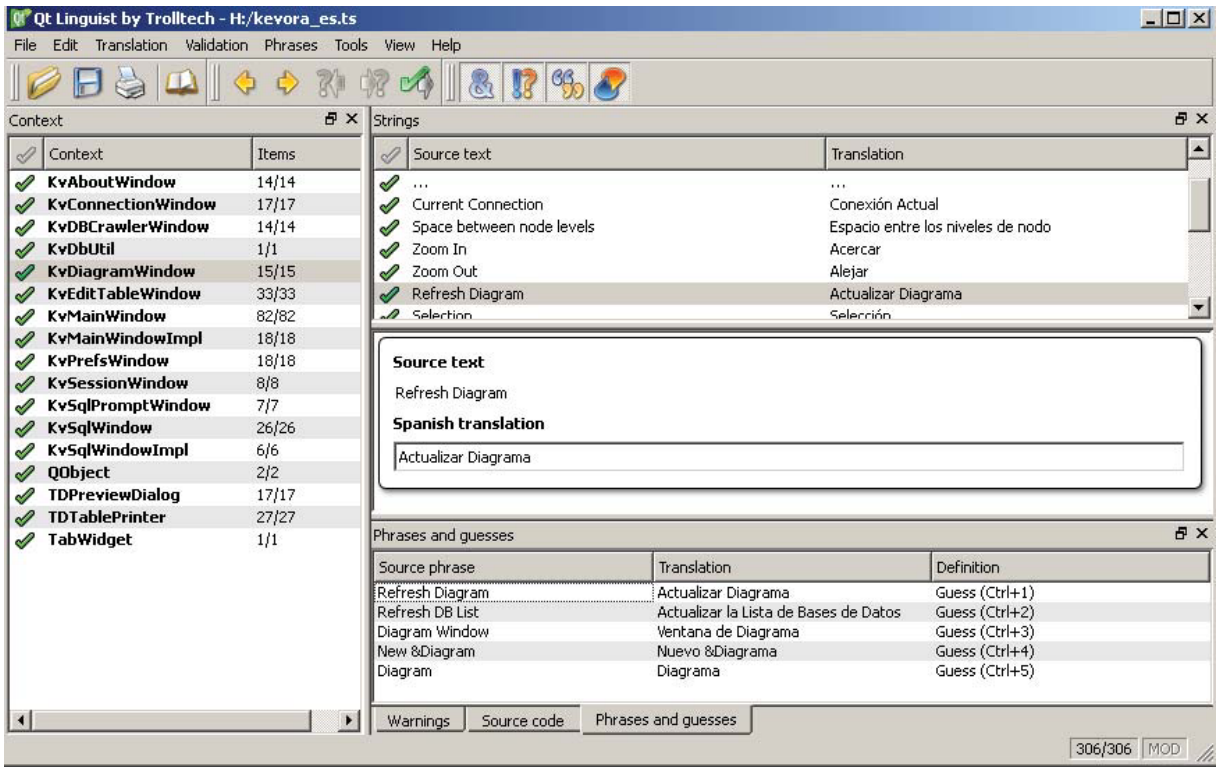

Figure 1 – Screenshot of a Localisation Workbench (QT Linguist)

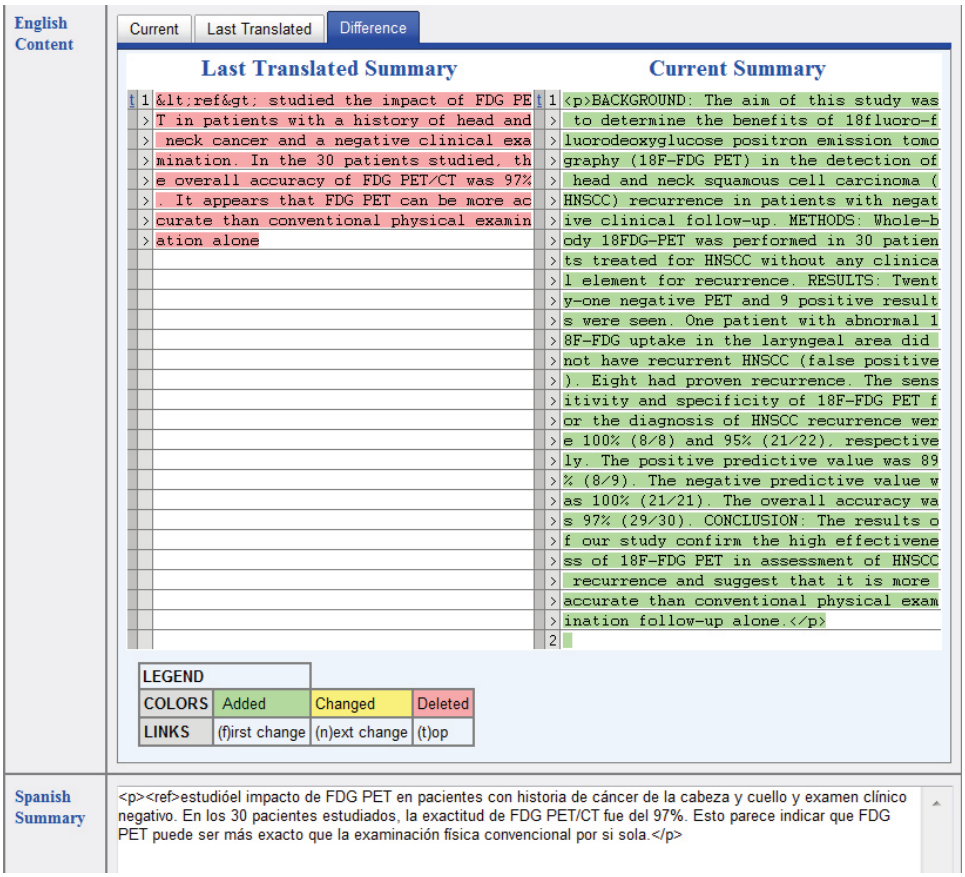

Figure 2 – Screenshot of GCMS (PETLinQ^{Iq})

4.2. Questionnaire

The following is a list of the questions used in the questionnaire:

1. How would you classify your organisation?
   a. Freelance / Self-employed
   b. Small business (2-99 employees)
   c. Medium-sized business (100-500 employees)
   d. Large-sized business (> 500 employees)

2. Which of the following categories best describes the activities of your organisation?
   a. Agriculture, forestry and fishing
   b. Mining
   c. Manufacturing
   d. Electricity, gas, water and waste services
   e. Construction
   f. Wholesale trade
   g. Retail trade
   h. Accommodation
   i. Transport, postal and warehousing
   j. Information media and telecommunications
   k. Financial and insurance services
   l. Rental, hiring and real estate services
   m. Professional, scientific and technical services
   n. Administrative and support services
   o. Public administration and safety
   p. Education and training
   q. Health care and social assistance
   r. Arts and recreation services
   s. Other

3. Where are your organisation's headquarters located?

4. Before reading the introductory text, did you know the meaning of the terms “internationalisation” and “localisation”?

5. What is the main language used in your organisation’s website and/or websites your organization develops?
   a. Arabic
   b. Chinese
   c. English
   d. French
   e. German
   f. Italian
   g. Japanese
   h. Korean
   i. Portuguese
   j. Spanish
   k. Other

6. What other languages have being used in your organisation's website and/or websites your organisation develops? If none, skip this question.
   a. Arabic
   b. Chinese
   c. English
   d. French
   e. German
   f. Italian
   g. Japanese
   h. Korean
   i. Portuguese
   j. Spanish
   k. Other

7. If your organisation doesn't localise content, please select one or more of the following reasons
   a. Costs
   b. Complexity
   c. Not needed
   d. Never thought about it
   e. Not interested
   f. Other

8. If your organisation doesn't localise content. Would it consider localising its websites and/or websites it develops?

9. If your organisation does localise content. What impact has this had?
   a. Increased market
   b. Increased customer experience
   c. Improved branding
   d. Reduced fines due to failed compliance
   e. Increased my costs
   f. Increased complexity
   g. Delayed launching of websites
   h. Haven't benefited from it
   i. Other

10. If your organisation does localise content, which methodologies or technologies are used to localise?
    a. Ad hoc human translation and manual modification of code/data
    b. Third party online machine translation (Babel Fish, WorldLingo, Google Translate, etc.)
    c. Global Content Management Systems
    d. Other

11. Would you like to participate in future surveys related with internationalisation and localisation?

12. Please feel free to add any comments or suggestions.

Due to the limitation of the platform used for the survey the results underwent a control prior to the final analysis. Google forms, although economical and easy to implement, don't allow the use of embedded logic in the survey. Therefore in some cases the answers to questions 5 and 6 contradict those of questions 7 to 10. For example, somebody who according to question 6 doesn't localise content could have answered question 9 (impact of localisation).

In these contradictory cases the answers were considering valid using the following logic:

- answers to questions 7 and 8 were valid if answers to questions 5 and 6 showed the use of only one language. This meant the absence of internationalisation/localisation.
- answers to questions 9 and 10 were valid if answers to questions 5 and 6 showed the use of more than one language. This meant the presence of internationalisation/localisation.

4.3. Preliminary Results

Although at the time of writing this paper only 34 people have completed the questionnaire, some interesting trends can be highlighted.

The most important tendency is that almost half of those consulted ignored what localisation and internationalisation means. It is noteworthy that all of the respondents had being involved in Web projects in one way or another, and some of them in traditional software development. This highlights the importance of promoting the message that adjusting software or a website to make it available for a specific region, culture or market, involves more than "translating" the content. Translation is just part of the internationalisation and localisation process.

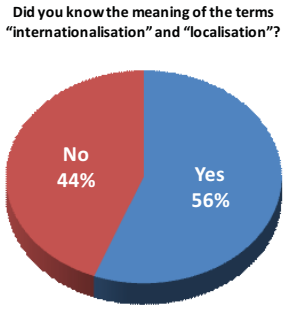

Figure 3: Knowledge of the terms internationalisation and localisation

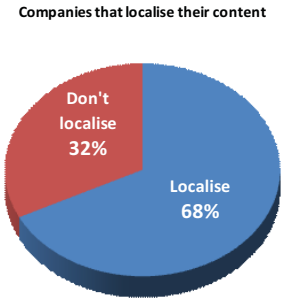

Figure 4: Companies that localise their content

While many of the participants ignored the terminology, 23 (68%) of them actually localise their content. Of the remaining 11 (32%), 4 of them stated that they never thought about it and only 3 (8.82% of all respondents) didn't need it. Here we can see again that the lack of knowledge of the importance of the field plays an important role as only 8.82% of the respondents stated that they don't need to localise their content.

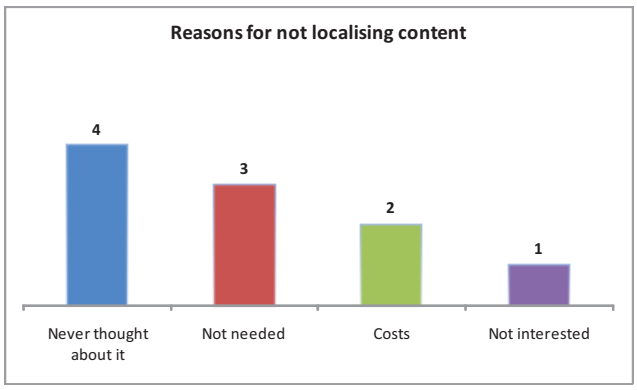

Figure 5: Reasons for not localising content

Those who localise confirmed that despite an increase in complexity and costs, localisation improves customer experience and increases the market. Only 1 (4.35%) of 23 businesses that localise their websites stated that they haven't benefited (Figure 6).

As noted, an increase in costs is one of the main concerns when deciding to localise. This could be minimized if GCMSs were used instead of ad-hoc localisations like 55% of the respondents resort to (Figure 7). Ongoing development costs are reduced by using a GCMS, as only a translator is hired on regular basis.

As mentioned in the methodology description, it would be necessary to localise this survey into more languages than English and Spanish to obtain more objective data, in particular to know the preferred languages when localising. However, we can see similar results to those obtained in a study performed by SDL on the preferred target languages of English speaking companies when using automated translation (SDL, 2008). Although the results of this study are not very clear, it showed that FIGS languages (French, Italian, German and Spanish) were the preferred languages, followed by Asian languages.

By adding the answers of organisations where English is not their main language, our results also show the predominance of FIGS and English as preferred localisation languages (Figure 8).

5. Relevance for our students

From the results obtained in our study and the bibliography reviewed, it can be said that localisation and internationalisation are definitely a career path in the same way than other subfields in the IT industry are (usability and testing are examples of it). Internationalisation and localisation are a great opportunity for our international and Maori students to have a role in the IT industry. Maori students could help with the translation and proofreading of a website or software that has been localised to Te Reo. Our Chinese students could help to localise to Chinese, a Korean student to Korean, a Brazilian student to Portuguese and so on.

Microsoft and Google provide great examples of localisation projects in New Zealand. In November 2005,

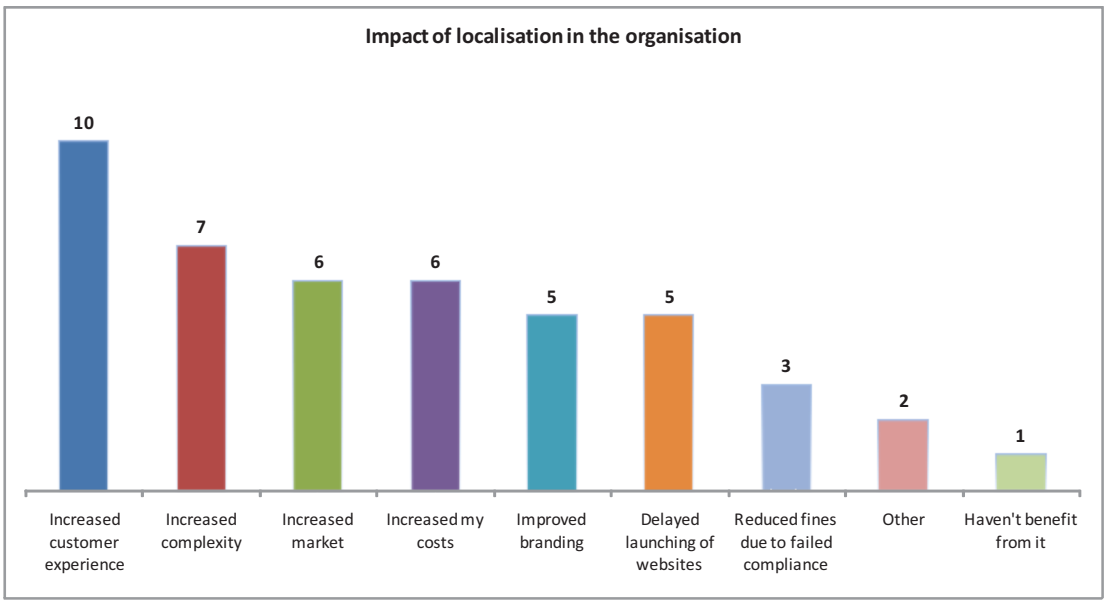

Figure 6: Impact of localisation in the organisation

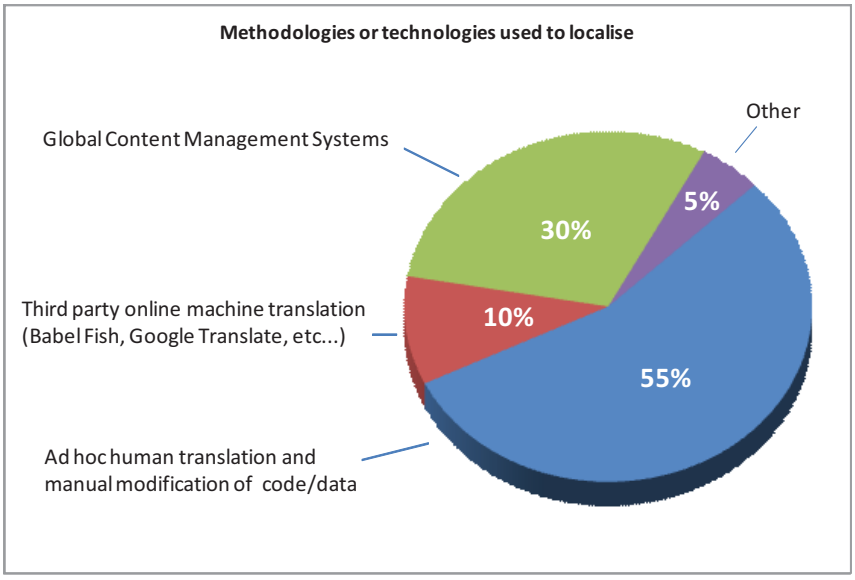

Figure 7: Methodologies used to localise

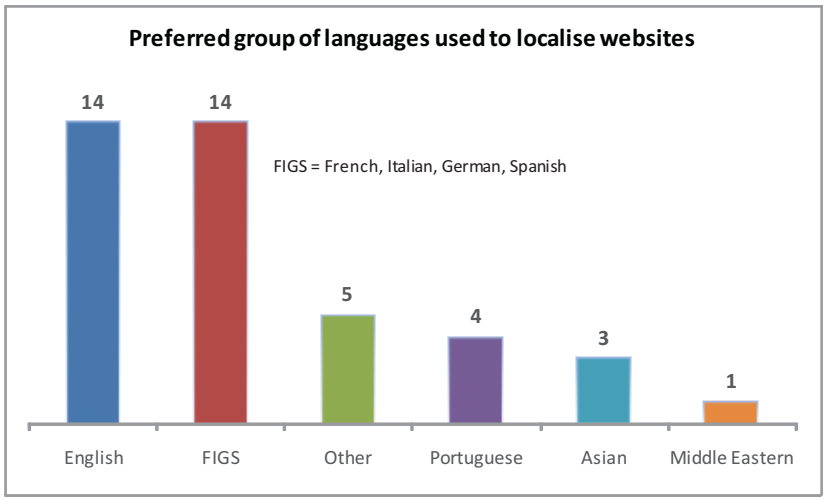

Figure 8: Preferred group of languages used to lo localise

after a year of collaboration between Microsoft New Zealand, Te Taura Whiri i te Reo Maori (Maori Language Commission) and Waikato University, Microsoft launched Microsoft Windows XP and Microsoft Office System 2003 in Te Reo. In July 2008, with the assistance of Te Taura Whiri i te Reo Maori, Google launched a te Reo version of their search engine. This effort required the work of a team of volunteers in New York and New Zealand.

Te Reo is not the only opportunity to work in the field in New Zealand. Companies like Staker Interactive (ShadoCMS) base their business in localisation while other firms (mainly web design companies) provide it as part of their services.

At international level the opportunities in the internationalisation and localisation field are big. At the time of writing this paper Google and Microsoft were searching for localisation project managers in Silicon Valley. Of course, the key for these positions is the ability to speak fluently more than one language.

6. Conclusion

In today's economic environment, more than ever before companies that promote and sell their products over the internet need to expand their markets. The internationalisation and localisations of websites is a powerful tool to accomplish this by reaching non-English speaking audiences.

However, despite the efforts of organisations like LISA, GALA and TILP, internationalisation and localisation are fields barely known by more than half of the professionals involved in web development. Thus it is very difficult to promote these types of services.

Internationalisation and localisation require the use of technologies and standards in combination with the knowledge of more than one language. This provides an excellent opportunity for our Maori and international students to be involved in IT by assisting on the localisation of software or websites. The use of students to provide these services will help both parties; businesses can reach bigger audiences and new markets at a low cost, and students will gain some real exposure to the IT industry.

Although the field is not a core competence in our programs, it is relevant in today's global economy in the same way as other technical skills that we teach like usability or testing. It is not necessary to create a paper on the topic; it could form part of an industry project or even a simple class activity to have students of different ethnicities working together.

7. References

Codespanish » Blog Archive » Speaking your Audience's Language - Localisation in the World Wide Web. Awareness Survey. (2009, January 21). . Retrieved March 5, 2009, from http://www.codespanish.com/softwaredevelopment/localisation-in-the-world-wide-web-%e2%80%93-survey.

ColdFusion CMS Multilingual Content Management System. (n.d.). . Retrieved March 1, 2009, from http://www.shadocms.com/.

Collaboration launches Microsoft Office and Windows in Reo Maori. (n.d.). . Retrieved March 1, 2009, from http://www.waikato.ac.nz/news/index.shtml?article=497.

Django | Internationalization | Django Documentation. Available at: http://docs.djangoproject.com/en/dev/topics/i18n/#topics-i18n [Accessed February 18, 2009].

GALA: The Globalization and Localization Association. Available at: http://www.gala-global.org/ [Accessed February 11, 2009].

Global Information Solutions - Translation - Localization - Globalization - SDL. Available at: http://www.sdl.com/en/ [Accessed February 18, 2009].

GlobalSight - Home. (n.d.). . Retrieved March 4, 2009, from http://www.globalsight.com/index.php.

Google launches Maori version - Technology - NZ Herald News. (n.d.). . Retrieved March 1, 2009, from http://www.nzherald.co.nz/technology/news/article.cfm?c_id=5&objectid=10523259.

Google Translate. (n.d.). . Retrieved March 9, 2009, from http://translate.google.com/.

How to run a successful localisation business. (n.d.). . Retrieved March 3, 2009, from http://www.gala-global.org/en/resources/downloads/040629%20LocWorld%20GALA%20Vendor%20Roundtable.pdf.

ISO 30042:2008 - Systems to manage terminology, knowledge and content -- TermBase eXchange (TBX). (n.d.). . Retrieved March 4, 2009, from http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=45797.

Jacqueline Smith: Websites speak the right language - Small business - NZ Herald News. (n.d.). . Retrieved March 10, 2009, from http://www.nzherald.co.nz/small-business/news/article.cfm?c_id=85&objectid=10544260.

Jurafsky, D., & Martin, J. H. (2008). Chapter 25. Machine Translation. In D. Jurafsky, & J. H. Martin, *Speech and Language Processing* (pp. 859-908). New Jersey: Prentice Hall.

LISA: Homepage. Available at: http://www.lisa.org/ [Accessed January 8, 2009].

LISA: OSCAR, LISA's Standards Committee. Available at: http://www.lisa.org/OSCAR-LISA-s-Standa.79.0.html#c119 [Accessed January 8, 2009].

Localisation in the World Wide Web - Awareness Survey. (n.d.). . Retrieved March 5, 2009, from http://spreadsheets.google.com/viewform?key=prYdC1TwYW-HS6q5p2lMKeg.

Localización en la World Wide Web - Encuesta de concientización. (n.d.). . Retrieved March 5, 2009, from http://spreadsheets.google.com/viewform?key=prYdC1TwYW-EurOd4SIXeVg.

Localization Industry Standards Association. (2007). *LISA: Globalization Industry Primer.* Retrieved January 2009, from www.lisa.org: http://www.lisa.org/Globalization-indust.468.0.html

Microsoft Careers — Technical — Localization. (n.d.). Retrieved 03 03, 2009, from Microsoft Careers - United States: http://members.microsoft.com/careers/careerpath/technical/localization.mspx

National Population Estimates: December 2008 quarter - Statistics New Zealand. (2009, 02 13). Retrieved 03 03, 2009, from Statistics New Zealand: http://www.stats.govt.nz/products-and-services/hot-off-the-press/national-population-estimates/national-population-estimates-dec08qtr-hotp.htm

Open Source Migration. (n.d.). . Retrieved March 4, 2009, from http://www.globalsight.com/index.php?option=com_content&view=article&id=68:migration&catid=2:learn&Itemid=72.

PETLinQ IQ Login. (n.d.). . Retrieved March 9, 2009, from http://petlinqiq.com/ptlq_auth/new_login/start/?next=/.

Qt – A cross-platform application and UI framework — Qt – a cross-platform application and UI framework. (n.d.). . Retrieved March 4, 2009, from http://www.qtsoftware.com/.

QuickStats About Māori. (2007, March 27). . Retrieved March 5, 2009, from http://www.stats.govt.nz/NR/rdonlyres/095030F8-BD62-4745-836D-0EF185619C37/0/2006censusquickstatsaboutmaorirevised.pdf.

Scoop: Google Māori goes live. (n.d.). . Retrieved March 1, 2009, from http://www.scoop.co.nz/stories/ED0807/S00073.htm.

SDL. (2008). *Trends in automated translation in today's global businesses.* SDL.

The Localization Industry Standards Association. (2008, 10). Retrieved 3 4, 2009, from LISA: Standards News: http://www.lisa.org/Standards-News.466.0.html?&no_cache=1&sword_list[]=iso&sword_list[]=tbx

TILP - The Institute of Localisation Professionals. Available at: http://www.tilponline.org/ [Accessed January 13, 2009].

Top Ten Internet Languages - World Internet Statistics. (2008, 12 31). Retrieved 03 03, 2009, from Internet World Stats: http://www.internetworldstats.com/stats7.htm

Welocalize. (2009, 1 5). *Open Source Migration.* Retrieved 3 4, 2009, from GlobalSight: http://www.globalsight.com/index.php?option=com_content&view=article&id=68:migration&catid=2:learn&Itemid=72

WorldLingo Language Translation Services. Free Online translations. Translation Software. (n.d.). . Retrieved March 9, 2009, from http://worldlingo.com/.

Yahoo! Babel Fish - Text Translation and Web Page Translation. (n.d.). . Retrieved March 9, 2009, from http://babelfish.yahoo.com/.