Localization in the Context of a Third World Country

Dawit Bekele

Department of computer science
Addis Ababa University
P. O. Box 3479
Addis Ababa
Ethiopia
dawitb@ethiolink.com

Mots-clés : Localisation, Amharique, Ethiopie, Normalisation

Keywords: Localization, Amharic, Ethiopia, Standardization

Résumé Après avoir défini la localisation, cet article discute son importance pour les nations du tiers-monde. Ensuite, il présente les raisons essentielles pour lesquelles la localisation est de plus en plus importante pour les pays du tiers-monde à savoir: pour les permettre de travailler dans leurs langues officielles, pour permettre l'accès des NTIC aux jeunes et pour limiter la faille numérique. Cet article continue par présenter l'histoire de la localisation en Ethiopie qui consiste essentiellement d'efforts avec les trois buts suivants: 1- permettre des entrées sorties en Ethiopic 2- Localisation de la date et de l'heure 3- Localisation des interfaces de logiciels. L'article termine par présenter les questions essentielles que les pays en voie de développement doivent considérer. Ces questions sont: Coordination, Normalisation, Vision courageuse, Partenariat et Solution logiciel libre.

Abstract After defining localization, the paper discusses its importance for third world nations. Thereafter, it presents the main reasons why localization is increasingly important for third world countries: to be able to work in official languages, to provide access to the youth and to limit the digital divide. The paper then discusses the history of localization in Ethiopia that concerned mainly efforts with the following three goals: 1- enabling Ethiopic input and output, 2- localization of the date and time, 3- localization of software interfaces. Finally, it proposes the main issues that developing countries seeking localization should consider, which are: Coordination, Standardization, Bold vision but realistic approach, Partnership and Open source option.

1 Introduction

Until recently, most third world countries seemed resigned to using software products in the languages of more prosperous countries. This is especially true in African countries where local languages were rarely used in computers.

However, recently, more and more localization efforts are being made for languages that didn't get this opportunity in the past. Even Microsoft announced that it will localize its Windows operating system as well as its Microsoft Office Suite to 40 languages, many of them spoken in the third world, including Swahili and Amharic in Africa (Microsoft Press Passes, March 16 and March, 2004). The open source community is also contributing in localizing an increasing number of software products to languages in the developing world. Linux, for instance, is localized for than 80 languages, many of which are from the developing world.

Unfortunately, the increasing number of localization projects of the third world are facing unique problems that are inherent to the nature of localization and realities of these countries. Unless these problems are properly studied and adequate solutions provided, the efforts put on these projects may not provide as much fruit as expected.

This paper tries to identify the major challenges of localization by focusing on developing countries' environment. The decades long localization attempt of Ethiopia will guide the reflection. The paper will also propose some important localization strategy issues that developing nations should consider.

2 Localization

Localization can be defined as "the transfer of cultural consciousness into a computer system" (Daniel Yacob, 2004). The localization Industry Standards Association (LISA) also defines localization as follows: "Localization involves taking a product and making it linguistically and culturally appropriate to the target locale (country/region and language) where it will be used and sold". In reality, since the great majority of today's software products are produced in the United States of America and/or in the English language, most of the localization activities consist of transferring a software product developed in English/US culture to other cultures and languages.

Prior to the 1980s, software producers gave little importance for localization. In the early days of computing, even the largest computer companies ignored the importance of localization. For example, most products developed prior to the 1980s used the ASCII character set [1]. As a result, they did not even support characters such as é and è that are compulsory for French, a language used by tens of countries and hundreds of millions of people around the world. Later on, ISO-8559/1-8559/9[2] defined character sets for "major" European languages, Arabic and Hebrew which facilitated localization to these languages. However, it's only after Unicode[3]

---

[1] ANSI X3.4 American National Standard code for Information Interchange

[2] ISO-8559/1 - 8559/9 8-bit single Byte Code Graphic character sets

was released that most alphabets and scripts of the world have had the chance to be used by major software products (ISO, 1998).

In the late 1980s, software companies started to realize the importance of localization to capture more markets. More and more products started to be shipped with an interface in the language of the country of destination. Localization industry started to emerge with Ireland as its hub. In 1990, the Localization Industry Standards Association (LISA) was founded (Esselink Bert, 2000) to coordinate the localization efforts.

Today, localization is taken very seriously by all major software producers. Most software producers have internationalization[4] departments that insure that all the software products they produce are easily localizable[5]. Programming environments are also friendlier to localization: Java, Visual Studio as well as .Net platforms have increasing localization capabilities (Daniel Brandon, 2001).

As indicated in the definition, localization is more than translation from one language to another. It is true that translating the interface, the help system as well as the manuals is probably the most time consuming activity. However, localization also includes making sure that the software uses the particular locale's cultural conventions such as the date and time system, measurement system, number formats[6] colors and icons. Some conventions that are well established in some country may be completely unknown in some other country. For example the mail box image often used in American software products has no meaning in Ethiopian context, where mail is not delivered at home and such boxes do not exist.

3   Why is localization so important for developing countries?

Developing countries in general and African countries in particular have been using computers for at least the last four decades without necessarily benefiting from localization to their own languages. It is therefore a legitimate question to ask why not continue to use computers having western language interfaces. One could also say that localization is by no means an easy task that poor countries can readily afford, especially when these countries have hundreds of languages and different cultures. It also requires sustained effort; for example, Ge'ez Frontier Foundation translated into Amharic more than 40,000 thousands of words of the Linux operating system, which amounted to 40% of the vocabulary. However, just a year later the percentage decreased to 22% due to new words included in Linux software and the slowdown of the translation effort[7].

---

[3] Unicode/ISO 10646 32 bit character set

[4] According to LISA, Internationalization is the process of generalizing a product so that it can handle multiple languages   and cultural conventions without the need for redesign

[5] Excerpt from the talk of Menassie, Zaudou, Software developer at Sun Micro Systems.

[6] French speaking countries use the comma (,) as a decimal separator while most of the remaining world uses the dot (.). This after creates confusions that can have severe consequences.

[7] http://l10n-status.gnome.org/gnome-2.10/index.html gives the level of real-time progress of all GNOME localization projects.

However, there are an increasing number of reasons why these countries have to invest in localization, in spite of the high cost that it may incur. Some of these reasons are presented below.

One of the most important reasons, at least for some of the countries, is that the local languages are the working language of the offices where the computers are used. In Ethiopia for instance, the official working language of the Federal Government is Amharic and the regional governments use their official working language. Until now, even though English is not the official language of the country, government and private organization employees who need to work with computers were required to know English. This requirement was acceptable when computers were used only in large federal organizations that have employees with, almost always, university degrees. Today, thanks to some projects such as the WeredaNet that connects each Wereda (district) administration of the country, computers are used by people who do not even speak the federal working language let alone English. This new bread of users cannot use the computers that they have available unless they are localized to their languages and cultures. Therefore, localization is no more a luxury but a necessity.

Another reason follows: on one hand computers are being increasingly used by younger people since, for example, SchoolNet programs are being launched all over the world, even in the developing world. On the other hand more and more school systems use local languages as teaching medium. In Ethiopia, the multi-billion Birr[8] SchoolNet program is installing and connecting computers in all high schools of the country. This gives to the millions of schoolchildren access to computers that they did not have before. It is expected that primary schools will have similar settings in the near future. At the same time, political leadership and education specialists are pushing for the use of local languages, at least in primary schools (African Ministers, 2002), (Kwesi Kwaa Prah, 2002). As a result starting from mid-1990s, all Ethiopian public primary schools teach in the regional languages. Consequently, since the young schoolchildren do not yet master languages other than their own, providing them with localized computers becomes a necessity.

The third reason is related to the digital divide. Most governments are now convinced that they have to avoid or at least limit the digital divide that exists between rich and poor countries but also between rich and poor within the same country, in order to avoid or at least limit some serious economic and social consequences. There are studies that show that in African countries, the overwhelming majority of the population does not speak the western languages that their countries have adopted as national languages. In Ethiopia, even though English is used in higher education and correspondences with foreign organizations, it is not an official language and few people use it on regular basis. Without localization, the use of computers is automatically restricted to the very few that speak western languages and it is therefore impossible to talk about bridging the digital gap.

The above three reasons, and probably others not given here, make localization a necessity, for an increasing number of countries. Even those that do not yet feel the need, will be confronted to it in the near future, because of the desirable and unavoidable popularization of computers.

---

[8] Birr: Ethiopian Currency equivalent to approximately 0.09 Euros in March 2005

4 History of localization in Ethiopia

Ever since the introduction of computers in Ethiopia in the 1960s, users and government highly desired localization. This is because, as already stated, the official working language of the central government is Amharic and it is only natural that users want to use the computer in their working language.

Amharic is a semetic language that derive from Ge'ez, a language now extinct except in the Ethiopian orthodox church, where it remains the liturgical language. Amharic is one of the more than eighty languages spoken in Ethiopia. It is the official language of the Ethiopian government for several centuries and the lingua franca of the country.

Amharic uses Ethiopic (Ge'ez) syllabary for its writing system. The syllabary has 461 symbols (QSAE, 2002). The number of symbols of Ethiopic has been increasing throughout the years to accommodate sounds that are not supported and that are needed for new languages that have started using the syllabary. In addition, all other Ethiopia's and the now independent Eritrea's languages used this script before the government change of 1991. In particular, Afan Oromo, the language of the largest ethnic group of the country, has been using Ethiopic until 1991.

During the last four decades, at least the following localization efforts have been vigorously pursued:

1. Enabling existing software products to accept Ethiopic inputs and outputs. This has been given highest priority by users and developers. The main problems were associated with the high number of symbols in Ethiopic compared with the Latin script overwhelmingly used in computer software products and the lack of relevant standards (Abass B. Alamnehe, 1994).

   Developers used mainly the following four strategies to solve the encoding problem: using a subset of the character set rather than the whole set; using two or more fonts to cover all Ethiopic symbols, separating the diacritic marks from the symbols; or using a two-bytes character coding system. Unfortunately, none of the above strategies were fully satisfactory. Nonetheless, today, Ethiopic is included in Unicode, an encoding system developed to accommodate all alphabets of the world and not just Latin as ASCII does. Since Unicode is supported by increasing number of software products, it can therefore be considered that the encoding problem is now solved.

2. Localization of the date and time. Ethiopia differs from rest of the world for having its own calendar and time system. The official Ethiopian calendar is 7 years, 8 months and 11 days behind the Gregorian calendar. It also differs from the Gregorian calendar for having 13 months; 12 of which have 30 days and the 13th has 5 days or 6 days on leap years. In addition, the official time system is shifted by 6 hours compared to the system used in most of the world. Therefore, for an Ethiopian, it is 1 o'clock in the morning when it is 7 o'clock according to the western time system. Of course, none of the software products developed in the developed world comes with these date and time systems integrated. Some Ethiopian software developers have produced software products that display the Ethiopian date and time. However, the

most transparent localization for Ethiopian date and time systems has been on open source software such as Linux and OpenCMS that avail the source code hence make complete and transparent replacement of the Gregorian and western systems by the Ethiopian date and time systems.

3. Localization of software interfaces. Localization of software interfaces has been given low importance until recently since most users had some understanding of English and were expected to be able to work on computers using English interface. It is also because the most commonly used software products were Microsoft products and Microsoft did not allow localization by third party developers. Besides, the original software developers had no interest to invest on localization for a country with a GDP per capital of just over 100 USD (UNDP, 2003)!

4. The only software products that have localized interface were a small part of the few products developed locally, such as CMS used in courts and PIS used by the Ministry of justice (Dawit Bekele, 2003). Fortunately, in recent years, localization of software interfaces has been facilitated by open source software. Very recently, Microsoft has also shown interest to localize Windows and the Office suite to Amharic.

5 Localization strategy issues for third world countries

Very little is done in the area of localization, especially in third world countries. However, it is very probable that localization projects will increase in the coming years as they have during the past few years. These projects can learn from the decades-long localization attempts in Ethiopia and save precious time and money. This section proposes localization strategy issues that third world nations should consider in order to reach their final objectives with minimum effort and time.

The localization strategy of third world countries should consider at least the following major issues:

• Coordination

• Standardization

• Bold vision but realistic approach

• Partnership

• Open source option

It is important that all localization efforts for a specific locale are coordinated in order to avoid redundant works that only create confusion and additional cost, things that developing countries cannot afford. The coordination can be ideally done by an ICT authority or ministry. Otherwise, organizations interested in localization can form associations in the same way developed nations' localization companies established associations such as LISA.

Standardization greatly facilitates localization work and increases the quality of the localization as well as its acceptability by the general public. Adequate standards should be

developed at the appropriate time. Setting a standard before the area that is standardized is properly explored limits its acceptability since better options may be discovered once the standard is adopted. Setting a standard well after all techniques of the standardized area are explored will create the proliferation of incompatible products that will not disappear easily even after the standardization (Cargill Carl F., 1997). For example, in Ethiopia, even though an encoding standard is now available, most people continue to use the incompatible non-standard fonts they have been using for a long time. Similarly, since all sorts of incompatible keyboard layouts are being used, it is almost certain that many people will prefer to continue to use the non-standard keyboard layouts that they are used to even after a keyboard layout standard is established.

Third world countries should and can have bold visions such as localizing all major software products to all their major languages. It is important that they have this kind of vision in order not to succumb to the digital-divide. With proper strategy and with convinced leadership this is not an unrealistic vision. However, they should also realize that localization is not just translation of words and it requires a lot of sustained effort. Therefore, they should have realistic and progressive approach that enables them to see the fruit of their efforts as soon as possible. For example, localizing to all languages of the country at the same time is generally a bad approach since problems in the localization of one language can delay that of the others and also because it will drain heavily the resources of the country.

Localization will benefit if there is increasing partnership among localization project. There are a lot of intersections between localization works even between those targeting completely different cultures and languages. For example, in the Amharic localization of Open CMS performed by Addis Ababa University, many problems were encountered and most of them were solved by people involved in other localization projects around the world. Very surprisingly, the most useful assistances came from Chinese professionals who faced similar problems due to their use of a writing system different from Latin.

Partnership can also save a lot of cost for developing countries that cannot afford to spend a lot of money on localization. It also enables countries newly embarked in localization to learn from the experiences of others and reach their goals faster.

Last but not least, developing countries should seriously explore the localization of free and open source software. There are many reasons why developing nations benefit from using free and open source software products. This paper will only focus on those reasons that concern localization. Free and open source software products are most of them internationalized probably because they are developed by international teams with members from all around the world. Proprietary software products are less often internationalized. Free and open source software products are also interesting for localization since their source code is freely available. This makes any localization possible even those that were not planned by the original developers.

6 Conclusion

Localization is becoming an important issue for third world countries. It will continue to be an even bigger issue when more and more citizens of third world countries have access to computers and when their leaders are more aware of its importance.

Nonetheless, even though localization has become easier in the last few years there are still a high number of challenges and new countries should learn from others' experiences in order not to repeat the same mistakes. In this regards, the Ethiopian localization experience can give many lessons to African countries in particular and third world nations in general.

Bibliography

ABASS B. ALAMNEHE (1994), The need for a standardization of Ethiopian Script, Proceedings of EthCITA E-mail Conference, Volume III paper 3

AFRICAN MINISTERS (2002), Eighth conference of Ministers of Education of African members States (MINEDAF VIII)

CARGILL CARL F. (1997), Open Systems Standardization, A business Approach, Prentice Hall

DANIEL BRANDON (2001), Localization of Web Contents, Journal of Computing Sciences in Colleges, Volume 17 Issue 2

DANIEL YACOB (2004), "Localize or be Localized: An Assessment of Localization Frameworks", International Symposium on ICT Education and Application in Developing Countries, Addis Ababa, October 19-21, 2004.

DAWIT BEKELE (2003), Computerization Study Report, Ministry of Justice of the Federal democratic Republic of Ethiopia, Branch Office for Addis Ababa Administration

ESSELINK BERT (2000), "A practical Guide to Localization", John Benjamin's Publishing Company.

ISO (1998), ISO/IEC 10646-1:1993/Amd 10:1998(E), Amendment 10: Ethiopic Architecture and Basic Multilingual plane

KWESI KWAA PRAH (2002), NEO-COLONIALISM AND the African development challenge, TRIcontinental, Havana, Cuba, No. 150, 2002, Language

MICROSOFT PRESS PASS (March 16, 2004), Microsoft Enables Millions More to Experience Personal Computing Through Local Language Program

MICROSOFT PRESS PASS (March 2004), Microsoft Local Language Program, Quote Sheet

QSAE (2002), QSAE (Quality and Standards Authority of Ethiopia), Ethiopian Standard Ethiopic character set ICS:01.140.10, ES 781:2002

UNDP (2003), Human Development Report, www.undp.org