

Linux Localization in India

Venkatesh Hariharan

Abstract

The long-stagnant localization industry in India is entering an exciting phase due to a confluence of factors. The most significant development in recent years has been the growth of voluntary organizations that are localizing open source technologies to Indian languages. By harnessing the power of open-source technologies like the Linux operating system, the Mozilla browser etc, voluntary groups like IndLinux.org, Bengalinux.org etc are creating a completely localized environment that enables Indians to use computers in their own languages. This is in contrast to the past where localization was limited to the development of fonts, word processors etc that essentially operated as islands within an English language environment. Since a mere 10 percent of India speaks English, the localization of Linux to Indian languages promises to make information technology widely available to the masses.

Text

Open source technologies have emerged into the mainstream due to the popularity of the Linux operating system, the Apache web server and many other software systems. The impetus for the growth of open source software came from the rise of the Internet, which allowed individuals scattered across the globe to work together on projects of common interest. Open source has confounded critics who predicted that open source products would never match the sophistication of proprietary software. Much has been written about the battle between proprietary and open source software and I shall not delve into these in this position paper. This paper focuses on a little-known aspect that's slowly gaining media attention—the use of open source technologies as a platform for localization.

Open source is emerging as a powerful tool for linguistic groups across the world in their attempt to bring computing technology to the masses. A recent article in the Economist¹ points out that while Windows 2000 operating system is available in 24 languages, and Windows XP in 33, interfaces for Linux are available in twice as many languages. The article points out that KDE, one of the leading interfaces for Linux, has already been localized for 42 languages, with a *further 46* (emphasis mine) in the pipeline. GNOME, another popular interface for Linux lists teams working on over 80 languages on its web site.² The site lists little known languages like Tetum, national language of the recently liberated East Timor, Dzongkha, the national language of the tiny mountain kingdom of Bhutan, Fijian, Khmer, Nepali and many others.

¹ Open Source's Local Heroes, The Economist, December 4th, 2003. Online at http://www.economist.com/printedition/displayStory.cfm?Story_ID=2246308 (subscription required)

² See The Gnome Translation Project page at <http://developer.gnome.org/projects/gtp/teams.html>

Commercial software vendors ignore many of these languages because software in these languages is not commercially viable. They are probably right from their standpoint. However, from a societal standpoint, this deprives people of a voice in the digital age and exacerbates the digital divide.

Open source technology gives a voice to these languages in the digital age and is probably the greatest development in the history of localization of software. Indeed, when we review the history of open source technology itself, 50 years hence, we may find that localization of open source software is the single-biggest factor that spread the power of digital technology across the world.

It is for these reasons that we at IndLinux.org call Linux and open source software, “God's gift to India.” To my mind, open source represents one of the finest opportunities for taking the benefits of this wonderful technology to the masses. From the standpoint of cultural, political and economic freedom, there are enormous reasons why open source is relevant to India's future in the digital age. That's the reason why my friend Prakash Advani and I started IndLinux.org to localize open source software to Indian languages.

To my mind, there are two great advantages that open source provides for localizers, especially to those who are motivated by the mission of bringing technology to their fellow language speakers:

- 1) Leveraging of “software economics” and the open source model
- 2) A complete stack of software
- 3) The “bazaar” model versus the “cathedral” model

Leveraging of “software economics” and the open source model: Software is fundamentally different from other manufactured goods in that the marginal cost of making another copy of the software is close to zero. Making the first copy is the hardest task. Once the first copy is made, making a further million copies (or letting users copy it themselves) needs comparatively less effort. Development groups like IndLinux.org (Hindi), Bengalinux.org (Bengali), Swatantra Malayalam Computing (Malayalam) are leveraging this to develop Linux user interfaces localized to Indian languages. The software thus developed is licensed under the General Public License so that anyone modifying the software has to contribute the changes back to the community. The localized software is almost always, given out free so as to speed up the adoption of language computing.

The GPL license ensures that many tiny tributaries of individual contributions merge into a mighty flowing river. This is important because, in the past, many individual localization efforts—font development, for instance—have been fragmented and often at odds with the larger objectives of Indian language computing. Font developers used to develop their fonts in proprietary formats in order to maintain a “lock” over their customers. This meant that to retrieve and process that information, users had to ensure that the font is installed on their system. Since many of the open source localization

projects in India are motivated by the goal of information access, they are developing their own fonts and breaking the hold of proprietary font vendors in the process.

Cost is also an issue that will affect the popularity of localized software, piracy notwithstanding. In developing countries like India, the per capita income is around \$410. If the cost of operating systems and application software itself add up to this number, it is impossible for the vast majority of India to afford this technology. In countries like the US, where per-capita incomes are around \$30,000, it may be affordable but not in India. What India therefore needs is software that's priced in rupees and not in dollars. Linux fits the bill since it is available free and gives you the freedom to localize it to the language of your choice.

A complete stack of software: We wanted to create an operating system for India and when we looked around, there was only one choice—Linux--because we could not modify proprietary operating systems. If I wanted to translate "file" into the Hindi equivalent, I had no freedom to do that with proprietary software. The Linux operating system was a natural choice because it gave us the freedom to add interfaces in any language we chose.

The open source world has also been growing in maturity. From a graphical user interface to small applications like Instant Messaging, to complex programs like word processors, the complete stack of software is now available in open source, making it a viable alternative to proprietary software. Rather than developing complex applications from the ground up, we simply had to take existing applications and localize them. As mentioned above, many open source groups have also been developing missing components like fonts, spell checkers etc that are then used as a common community resource.

The “bazaar” model versus the “cathedral” model: Anecdotal evidence that I gathered from friends at Microsoft indicates that Microsoft had plans to localize Windows to Indian languages as far back as 1995. However, the Internet took everyone by surprise and Microsoft’s resources were diverted to warding off the threat posed by Netscape. It’s only now; eight years later that Microsoft has released a Hindi language interface to computers.

Linguistic groups, however, may prefer to march to the beat of a different drummer. The Economist article mentioned above describes how Janos Noll, founder of the Hungarian Foundation for Free Software accelerated the Hungarian translation of OpenOffice by throwing a pizza party. He got a dozen volunteers working locally while around 100 volunteers collaborated over the web and completed 21,000 text strings in three days. Anecdotal evidence suggests that the translation of the GNOME interface to the Mongolian language is the work of one man who was passionate about localization and kept plugging away for several months to complete the task.

The open source model may seem chaotic and ad-hoc but it works due to the reasons described by Eric Raymond in his paper, "The Cathedral and the Bazaar."³ In his paper, Raymond contrasts the loosely coupled collaborations over the Internet (the Bazaar) with the project management approach (the Cathedral) within software companies.

The advantage of the Bazaar model is that anyone can take the initiative for localization and the digital destiny of a language need not be dictated by the commercial calendars of proprietary software vendors.

For us at IndLinux.org, open source was an attractive alternative because when linguistic groups come together to localize Linux in a transparent manner, localization can be done in a manner that is far more culturally sensitive than any centrally controlled process. For example, should "file" be called a "file" in Hindi because the word is now part of the popular lexicon among Hindi speakers or should it be called something else? Who should decide this? Should a software company decide this or should local linguistic groups decide this? At IndLinux.org, we have volunteers from all over India who are localizing Linux to Hindi and the terminologies are decided in a consensus driven process rather than presented as a fait accompli. Anyone who is interested in helping with localization can always sign up on one of our mailing lists. This makes open source localization a democratic and inclusive process.

Apart from this, there are even deeper "cultural localization" issues that are so deeply embedded in computers that we are not even aware of it. For example, the whole metaphor of "desktops" and "files and folders" evolved from the work of Ben Schneiderman who argued that the interface of the computer must reflect the real world around us. In rural India, where most people have never owned a desk and are used to squatting on the floor, "desktops" and "files and folders" clearly do not reflect their reality. Yet, rural India is where 70 percent of India lives. Therefore, an enormous amount of research needs to be done to invent appropriate user-interface metaphors for India and Linux is an ideal platform to do this.

³ The Cathedral and the Bazaar by Eric Raymond. See www.catb.org/~esr/writings/cathedral-bazaar/