

## 6. Thinking Outside The Localization Box

Robert Adkins, Technetra, *E-mail : radkins@technetra.com*

**After more than twenty years, the non-English computing world of India is poised to explode in an exciting flurry of Open Source Software projects. What standards are needed? And how can progress be sustained?**

### Unicode or Many Codes?

The multi-language localization community in India is split into 3 camps. The original ISCII (Indian Script Code for Information Interchange) camp, was developed, promoted and commercialized by the Indian government. However, ISCII is now being overshadowed by the adoption of a second-generation, globally coordinated Unicode approach. A third, almost splinter camp, the Akshara camp, has advocated tackling the phonetic basis of Indian languages directly. Instead of relying on ISCII's and Unicode's complex and potentially inaccurate combination of vowels, consonants, matras and context markers to generate syllables, "Aksharas" or syllables are encoded directly.

Many contributors to the localization efforts of Indian language information processing believe that Unicode should become the single standard for the future. However, a blind acceptance of Unicode may only result in the adoption of a premature standard. The limitations of the basic Unicode strategy are well known. The real problem for both ISCII and Unicode is the necessity to have context markers (made up of multi-byte meta-sequences) which identify the "correct" way to render a glyph. This, of course, makes parsing and sorting a nightmare. The Akshara approach avoids all this by individually encoding 13000 phonetic syllables across various Indian languages. The limitations of Unicode alert us to the fact that the domain of localization is still an important R&D subject and should be nurtured through a variety of theoretical, investigative, and implementation efforts.

Open Source Software provides the framework for the development of multiple approaches to the complex requirements of language processing. ISCII, while being an open standard, has suffered from closed implementation efforts. In contrast, Unicode is both an open standard as well as a relatively open implementation environment. Consequently, Unicode products and projects tend, even at the implementation level, to promote interoperability and transliteration. The Akshara approach, similar in concept to PC-ISCII which was established for

simple embedded processors, is open but has never been adopted as a standard and has not garnered a broad set of collaborators.

### The Right Approach

Instead of being ends in themselves, standards for localization should be regarded as milestones from which further innovation will be developed. The Indian computing community should promote a range of viable projects, as long as they are open source, that seeks to preserve and advance Indian language information processing. Then, encourage all R&D communities as well as implementers and educators to participate. OpenOffice.org, for example, might implement an ISCII layer in order to process the many legacy documents and datasets that have been proliferated in that format. Mozilla has a subproject for ISCII support already. The Unicode layer of course will continue as the basis for a majority of Indianization projects. However, important projects such as Mozilla and OpenOffice.org should also build an Akshara layer for the unique benefits syllabilization offers. Then, just as the user can select a particular language to work in, he or she can select the kind of encoding and lexical processing algorithm desired.

### Local Computing Creates The New Economy

In the past, countless resources of the Indian government have gone into narrowly parochial localization standards which were then implemented through carrot and stick incentives. Few vendors rushed in to Indianize their products. Today, the local Indian computing landscape offers more potential. OSS promises the alternative of adopting and adapting a worldclass set of applications and tools. Already projects to localize desktops (Gnome, KDE, XFCE) and major applications (OpenOffice.org, Mozilla) are producing impressive results.

It is time for the Indian government to set an example by supporting collaborative projects using the practical wisdom from all worthwhile localization approaches across all of India's languages. This demands an inclusive embrace of multiple innovative efforts. The classical approach of exclusionary tunnel vision forced by premature and potentially flawed standards must be avoided. The real boon, however, is that best-of-breed open source applications and tools can be harnessed to build cost-effectively a new services industry around the emerging localized computing environments. Then the growing services industry will provide all the resources needed to sustain continued localization.