

Multilingual Web Applications Using PHP

M.R.K. Murthy Raju

Nida Comm, Anna Nagar, Madurai-625 020, India

Abstract

With a healthy growth in the number of web sites with content in Tamil and with greater penetration of Internet, the number of users who would prefer to receive web content in their language is likely to increase. With increase in the scale of the content and the need for keeping the data up-to-date, we need to think about creating dynamic and data driven web applications in Tamil.

Most of such applications are likely to have the content and interfaces in at least 2 languages (Tamil and English/other local language). There are issues related to developing internationalized web applications and also issues specific to Tamil. This paper attempts to discuss these issues and possible tools and approaches.

Issues related to Multilingual web applications:

1. Planning for multilingual interface, multilingual content and multilingual interactivity.
2. Dynamically identifying the most suitable language for a particular user and generating the page in the right language on the fly or redirecting the visitor to the right segment of the site
3. Minimizing the code duplication and make the projects maintainable.
4. Providing for plugging in additional languages in future.

Issues Specific to Tamil:

1. Any popular database driven application results in a huge amount of data and it may be advisable to use Unicode for storing the data for future compatibility and to avoid the need for batch conversions in future when the application has to move to Unicode completely.

2. As Unicode for Tamil is supported at OS level only in Windows 2000/NT and in some Unix platforms, we may need to go with 8-bit encoding for some more time for rendering content in the browser and for getting the user inputs. The choice will be between TAB/TSCII.

3. Whatever encoding is used for storing the data, there will be problems when we need to serve the same data to different web sites (Ex. News Feeds or shared classifieds databases etc. or other Application Services), because Tamil web sites use many different encodings today.

General Approaches to the above Issues:

1. Use Content negotiation routines provided by the web servers like Apache or write custom content negotiation routines to identify and serve the content in the right language and right encoding.

2. Separate display logic from the application logic in the scripts on the server side to reduce code duplication and to ensure maintainability.

3. Write web based convertors from/to Unicode and all currently used 8-bit encodings. This can be done as an open project so that all developers are benefited and we have the solutions quickly. The conversion algorithms readily available now can be used/adopted to reduce the development time. With web based convertors it will be possible for the same data to be served to different sites in different languages.

4. XML and RDF (Resource Definition Framework) can be used when data needs to be served to different web sites. This will make customization at the target site very easy.

5. Migration to Unicode even for rendering content in the browser may be planned when Unicode compliant Operating systems become more widely available.

PHP - A Useful tool:

PHP is a very powerful open source server side scripting language, which is gaining popularity because of its power, steep learning curve for programmers with C or PERL background and ease of coding. It is available for almost all major platforms and web servers. A large body of shared knowledge and experience is available to help us deal with various issues and support is quickly available from the helpful developer community.

It works well with most databases like text databases, Mysql, Oracle, Sybase and its XML processing capabilities are also very good. Many benchmarking studies have shown it to be much faster than ASP and other scripting alternatives.

PHP has a framework called PEAR (PHP Extensions and Applications Repository) which is an initiative similar to CPAN for PERL. This has been started with the objective of building a large archive of components written using certain standard coding practices. These components can be readily integrated by anybody into his applications.

We can make use of the above framework and work as a group using the open source model to build the components and tools required.

Case study:

A case study regarding implementation of a classifieds site in Tamil and English will be presented.

