

XML-Unicode environment for the creation and accessing of Indian Language theses:
Vidyanidhi experiences

By Shalini R. Urs

Given the legacy and reality of linguistic diversity in India, the challenges of addressing the issues of Indian Languages and scripts are quite daunting. Meeting the challenges of 'language interoperability' becomes a priority for any digital library of Indian language materials. One of the critical design considerations for managing multilingual content is identifying and implementing a suitable format for encoding and archiving. Unicode and XML offer enormous potential for meeting the goals of 'language interoperability' admirably. The drive to make Indian language materials web enabled led us to choose and experiment implementing XML-Unicode environment for the Vidyanidhi Project.

Despite the robustness, technical soundness and the practical viability of Unicode, its adoption for Indic scripts has not been widespread-almost non-existent. We believe that the reasons for the non-implementation of Unicode are largely due to the misconceptions and misconstructions rather than the limitations of Unicode. Though there are a few problem areas, which are not satisfactorily addressed, Unicode is perhaps the best available option for a multi language multi-script environment such as the Indic scripts. The advantages of enhanced content management abilities of XML have made it a leading development platform for both document creation and archiving. XML offers itself as a very viable and excellent candidate for creating, accessing and managing Indian language content and collection. With the development and availability of software platforms and tools XML –Unicode environment appears to be the 'best solution' for a digital library of Indian doctoral theses.

Vidyanidhi is an initiative, which aims to evolve into a national effort to develop an information infrastructure for facilitating the creation, archiving and accessing of doctoral theses submitted to Indian Universities. Vidyanidhi's strategies are centred on developing appropriate policy frameworks; evolving suitable resources and tools for facilitating the creation of doctoral theses; evolve into an archive and a repository for Indian doctoral theses; and provide education and training. One of our current development efforts is to test the feasibility of taking the Indian language theses to the web. In this endeavor, we have tested and implemented the 'MS Word to XML' strategy.

In this paper, we present the complexities and intricacies of Indian languages and scripts; the enormous advantages of XML-Unicode environment and Vidyanidhi approach and our experiment in successfully taking Indian Language theses to the Web.