

Fast Facts

Project: Enhancing Communication and Co-operation across South Asia: An ICT Solution to Script Barriers

Lead Organization: Advanced Centre for Technical Development of Punjabi Language, Literature & Culture, Punjabi University

Country: India

Budget: AUD 29,819.00



Situation

South Asia is one of those unique parts of the world where single languages are written in different scripts. This situation resides with the Sindhi language, which is written in a Perso-Arabic script in Pakistan and in Devanagiri script in India and a growing numbers of Sindhis in the EU and US who use Roman script. Sindhi is recognised as the language of a regional state in Pakistan Sindh with a fairly developed print culture in the form of newspapers, magazines and other publications and University departments researching into the language. It has its own True type fonts that are cheap and easily accessible. However, literature or electronic Sindhi produced in Sindh is not available to the Sindhi population in India, which is geographically scattered and without a local state to support language development. Two barriers to communication exist, firstly the lack of state support and secondly the issue of mutually incomprehensible scripts: Arabic and Devnagri cannot reach the population in the West and vice versa.

Solution

The aim of this project was to facilitate electronic and written communication between Sindhi people living in India and Pakistan through the development of a bi-directional web based Sindhi Language Transliteration Tool. The target groups were Media organizations (such as magazines/newspapers), literary and literacy promotional organizations, writers and NGOs involved in dissemination activities amongst the urban and rural poor, virtual Sindhi speaking communities, schools, and colleges. This project developed a complete machine transliteration system for Sindhi scripts to facilitate the use of these technologies on the web, thus enhancing networking between India and Pakistan. The main modules developed were:

1. Sindhi (Perso/Arabic) to Sindhi (Devanagiri) electronic Dictionary.
2. Sindhi (Devanagiri) to Sindhi (Perso/Arabic) electronic Dictionary.
3. Transliteration rules and mapping tables.
4. Sindhi Corpus for Devnagri and Arabic scripts.
5. Language models for Sindhi.



Workshop in Progress, MSU Baroda, Gujarat, India

The project team also provided the user option to convert a complete website of Sindhi from Devanagiri to Perso-Arabic script and vice-versa including all the links, as has already been done before in their previous projects for Shahmukhi (Urdu) to Gurmukhi (<http://s2g.learnpunjabi.org>) and Urdu/Hindi (<http://uh.learnpunjabi.org>) script transliteration. Hence, the final output of the project was a website for

online transliteration of any Sindhi document into Sindhi script or Devanagari script and an ICT tool to convert any Unicode based Sindhi website into other script also developed.

Outcomes

- The project has been published in newspapers in India and Pakistan. The information has been spread in English as well as Hindi/Sindhi (Dev), Sindhi (Perso-Arabic) and Punjabi.
- Contact has been established with user groups in Pakistan and the Sindhi Diaspora in Europe and North America. In particular contact with Dalit led NGOs in Pakistan has been made in light of future dissemination of the software.



SANGAM Online System <http://sangam.learnpunjabi.org>

- As an impact of media coverage, the team has received emails from Sindhi community. They are very excited and eagerly waiting for the system to use. Mr Rakesh Lakhani an IT Professional, Sindhi Blogger & Activist and VimmiSadarangani, Associate Professor of Sindhi, Tolani College of Arts & Science, Adipur. Kutch Gujarat has offered their services or helps to the project team and participated in the 5 days' 2nd Sindhi workshop at MSU Baroda, India.
- The transliteration accuracy of the system ranges between 92-97%, which is a bit lower than expected. In order to achieve higher transliteration accuracy, the project team employed both rule-based as well as statistical approaches to Sindhi script transliteration.
- Presented and published research work:
 - ICON2014, "Sangam: A Perso-Arabic to Indic Script Machine Transliteration Model", 18-21 December, 2014, Goa University, Goa, India;
 - SANGAM system demonstration has been successfully presented during ICTD 2015 International Conference at Singapore, 15- 18, May 2015.

Project Contact

Professor Gurpreet Singh Lehal (gslehal@gmail.com)
 Advanced Centre for Technical Development of Punjabi Language
 Literature & Culture, Punjabi University
 Patiala, India
 Tel: 0091-175-3046171
 Fax: 0091-175-3046313
www.learnpunjabi.org

ISIF Asia Secretariat
 Asia Pacific Network Information Centre (APNIC)
 6 Cordelia Street, South Brisbane, QLD, 4101 Australia | PO Box 3646
 Tel: +61 7 3858 3100 | Fax: +61 7 3858 3199
<http://www.apnic.net>