

Project factsheet information

Project title	Enhancing Communication and Co-operation across South Asia: An ICT Solution to Script Barriers
Grant recipient	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
Dates covered by this report	20 – 02 – 2014 / 31 – 10 – 2014
Report submission date	30 – 05 – 2014
Country where project was implemented	India
Project leader name	Professor Gurpreet Singh Lehal gslehal@gmail.com
Team members (list)	Dr. Tejinder Singh Saini, tej74i@gmail.com ; tej@pbi.ac.in Dr Virinder S Kalra, virinder.kalra@manchester.ac.uk Project Staff Mr. AnkurRana (Project Consultant) Mrs. RenuGoyal (Lexical Data Entry Operator) Mr. Ashish Kumar (Lexical Data Entry Operator)
Partner organizations	N/A
Total budget approved	AUD 29,819
Project summary	Sindhi in Arabic script is recognized as the language of Sindh in Pakistan 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 scattered and unable to read the Arabic script, writing in Devanagri script. Thus the aim of the project is 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 will be 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 will develop 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.



Table of Contents

Project factsheet information	1
Table of Contents	2
Project Summary	3
Background and Justification.....	4
Project objectives.....	5
Users and uses.....	5
Indicators	6
Project implementation: understanding the chain that leads to results	8
Project outputs, communication and dissemination activities	15
Project outcomes	19
Project management and sustainability.....	29
Impact.....	30
Overall Assessment.....	30
Recommendations.....	32
Bibliography	32

Project Summary

Tips: It is recommended to **complete this section once you have finalized the text of the report**. It will be easier to go back through to build the summary based on the highlights of the report the project team just put together.

The Project Summary can be up to **one page long**.

It should include a brief justification; an outline of the project objectives to be achieved; the project real timeline and the main activities conducted.

The abstract of the project written when ISIF Asia initially approved the project and the objectives listed in the Grants Agreement signed by APNIC and your organization should be useful inputs when preparing this section of the report.

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 Devanagri 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. Thus the aim of the project is 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 will be 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 will develop 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.

Sindhi: سنڌي (Perso-Arabic), सिन्धी (Devanagri)



The main modules to be developed are:

1. Sindhi (Perso/Arabic) to Sindhi (Devanagri) electronic Dictionary
2. Sindhi (Devanagri) 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

In addition, we shall also provide the user option to convert a complete website of Sindhi from Devanagri to Perso-Arabic script and vice-versa including all the links, as has already been done before in our 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 will be a website for online transliteration of any Sindhi document into Sindhi script or Devanagri script and an ICT tool to convert any Unicode based Sindhi website into other script will also be developed.



Background and Justification

Tips: The reader should be reminded of the **context** your organization is working, and where the project has been developed in.

This section provides a window to **understand the challenges** faced by the community you are working with.

Include a detailed description about the situation **before the project start**, describing any relevant aspects that make the project relevant in such a particular scenario.

The reader should be provided with a clear description about the problem(s) to be addressed through this project and the motivation from your organization and team members to get involved and offer a solution.

South Asia is one of those unique parts of the world where single languages are written in different scripts. Thus far, with the aid of Pan Asia ICT and ISIF grants we have resolved the communication issues by software that offers Urdu/Hindi translation and Punjabi transliteration. A similar problem resides with the Sindhi language, which is written in a Perso-Arabic script in Pakistan and in Devanagri in India and a growing numbers of Sindhis in the EU and US who use Roman script. Whilst in speech, Sindhi spoken in India and Pakistan is mutually comprehensible in the written form it is not. We aim to provide a tool that will help Sindhi people to link across a hostile geographical divide. In so doing we will provide an ITC solution to a social problem that had seemed insurmountable for centuries.

Since the independence of India and Pakistan in 1947 the populations of each country has developed a set of negative stereotypes that has resulted in three wars and continuing hostility. One of the barriers to people to people communication has been the lack of access to each other’s literature due to script divides. Thus our work thus far has aimed to lessen hostility between the two countries by aiding communication.

There are existing Sindhi online newspapers in Perso-Arabic modified script and thus the material available to a wider audience is substantial. Potentially all members of the substantial Sindhi speaking population around the world will benefit vastly from the Transliteration Tool. Email facilities would cross the geographical divide as well as allow for online chatting. Business could communicate easier with each other, same as media, educational and literary organizations. By enabling greater access to the print and news media in Pakistan to Indians, many of the myths about each other’s countries will be countered, which feed into negative portrayals.

There are wider implications for the development of literacy in both India and Pakistan. Given that in Pakistan there is a wealth of educational materials, from basic to university levels, and in Pakistan a massive population, literate in basic Arabic script (due to being taught the Koran) but has no access to materials. The transliteration tool will help in providing a facility where there can be an instantaneous conversion of materials.

The project is clearly in line with the overall objective of the program to in that it will develop a practical solution using existing technologies to provide Internet based transliteration for Sindhi. This will increase networking and provide access to ICT for a range of communities, thus enhancing learning through innovation and localization.

Year	Funding Agency	Project	Budget
2014-2015	The Information Society Innovation Fund (ISIF) Grants, Australia	Enhancing Communication and Co-operation across South Asia: An ICT Solution to Script Barriers	AU\$ 29,819
2006-2008	Pan Asia Grants, Singapore	Shahmukhi to Gurmukhi Transliteration Solution for Networking	US\$ 30,000



Project objectives

Tips: Please include here the **original objectives** as listed on the Grant Agreement.

If any objectives were modified, added or removed during the reported period this should be explained/justified.

Original Objectives

Web based software package developed for automatic machine Sindhi transliteration with following features:

1. Conversion of scripts with the click of a mouse.
2. Transliteration accuracy to be more than 95% at word level.

As part of this process the following tools were also developed:

1. Dual script Sindhi electronic Dictionaries.
2. Mapping tables and rules for transliteration.
3. Sentence aligned parallel script Sindhi corpus.
4. Language models for Sindhi.
5. Module for converting any Sindhi (Urdu script) website to Devanagri script and reverse.



Dissemination via social media and the active NGOs and websites engaged with Sindhi language and literature. A launch Event held in Patiala and in collaboration with Sindhi language centre, Sindhi media in India and Pakistan were engaged in the dissemination process.

Modification

The objectives were modified in terms of their ambition because there was a lack of soft data for Sindhi in the Devnagri script. This meant that resources had to be diverted to the manual creation of a corpus, which was a very time consuming task. Thus the overall objectives remained the same their scope and ambition was restrained.

Users and uses

Tips: Discuss with your project team who would be the future users and how they would use the findings throughout the project lifecycle. The uses identified should relate to the theory of change that you have discussed with your project team. The discussion about theory of change, users and uses, will be a very important input to your communication strategy: depending on who the user is and of what use will be the findings, a communication strategy can be developed. For example, if the users of the findings are policy makers and the use is to influence a change in the regulatory framework, which communication approach will work the best?

Who will be the user of these findings?

What are the more relevant things the project team wants to learn about or evaluate through the lifecycle of this project?

There are a number of different users who could benefit from this report:

1. NGOs involved in dialogue between India and Pakistan who can understand the process by which cross border work is possible.
2. Academics and researchers involved in the issue of transliteration at any juncture where the issue arises, for example in situations where Persian and another script are used (there are many examples of this across the world, especially in the Central Asian republics).
3. Academics and researchers in South Asia who are engaged in computational linguistics and who may not have access to the research papers generated from the project.
4. The Indian Government department for languages engaged in researching, promoting and publishing literature in regional languages could benefit where they desire to publish the text in multi/bi scripts but lack adequate resources so far.
5. The international availability of the report can be of specific use to the foreign ministries of India and Pakistan engaged in peace making for carrying out communication through internet where they can transliterate the content into their own scripts; and who are otherwise unaware of such a technological tool available at hand

Lesson Learnt by the Project Team:

The most important lesson learnt by the project team was that the transliteration system could be extended with ease to incorporate a host of other Indian languages sharing similar speech but having different scripts for writing, for example, like Kashmiri and Konkani languages of India.

Indicators

Tips: Indicators help to *measure project's progress*.

Indicators help the objectives that were set by the project team to be affordable, tangible, and measurable.

They help to verify the success and rewrite the course in case we are not achieving it.

An indicator could be quantitative (percentage, amount) or qualitative (perception, opinion).

The ISIF Asia secretariat suggests the SMART approach to indicators:

- S** **Specific**
- M** **Measurable**
- A** **Achievable (acceptable, applicable, appropriate, attainable or agreed upon)**
- R** **Relevant (reliable, realistic)**
- T** **Time-bound**

Baseline	Indicators	Progress	Assessment	Course of action
Not exist	Sindhi (Devnagri) Typing Tool	Completed	We have enhanced Hindi Typing system by incorporating additional ग, ड, ज, ब graphemes of Sindhi Script in Unicode.	This tool is used for creating Sindhi (Devnagri) electronic Dictionary and Corpus data.
Not exist	Development of a Sindhi rule based on Transliteration Engine	Completed	Like Urdu/Hindi and Shahmukhi/Gurmukhi, Sindhi Transliteration is achieved by extending the existing Machine Transliteration Model. Transliteration Accuracy ranges between 40-50% at word level.	System had improved through users Feedbacks. Electronic Dictionaries are planned/created to attain higher level of Transliteration Accuracies.

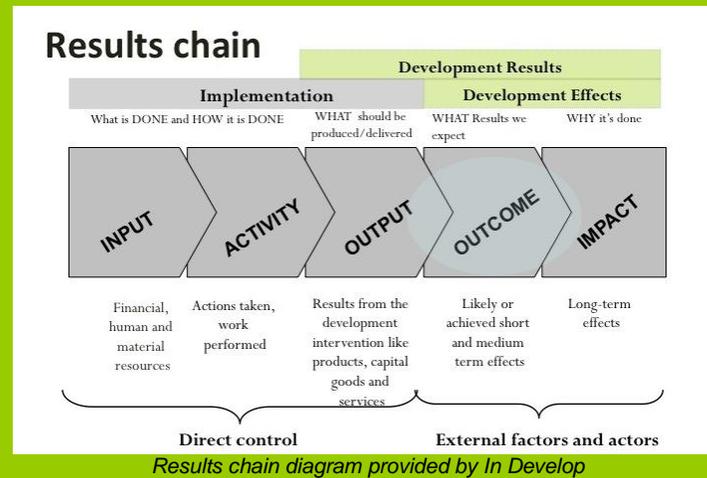


Baseline	Indicators	Progress	Assessment	Course of action
No electronic Dictionaries exist	Dual script Sindhi electronic Dictionaries	Completed (at least 25000 words in Devnagri and Arabic scripts)	It is necessary to continuously improve this dictionary through feedback from users of the tool.	Integration of Dictionaries into Rule based system: Transliteration Accuracy rose up to 80-85% at word level.
Not exist	Create 2.5 million Word corpuses of Sindhi (dev). 5 million-word corpuses of Sindhi (Arabic).	Completed	The status of Sindhi in Devanagri has hindered progress and reflects the need for the tool.	Using NLP Tool for Corpus Analysis to produce Sindhi Script Bi-gram and tri-gram language model.
Not exist	Creation of Offline transliteration tool having transliteration accuracy more than 95% at word level	Completed	Transliteration Accuracy ranges between 92-97% at word level	Planning "S A N G A M:: Perso Arabic - Indic Script Transliteration System" (Transliterate among Urdu-Hindi, Shahmukhi-Gurmukhi and Sindhi Scripts at a single UI)
Not Exist	Web based "S A N G A M::Perso Arabic - Indic Script Transliteration System"	Completed	Transliteration Accuracy ranges between 92-97% at word level	N/A

Project implementation: understanding the chain that leads to results

Tips: This is the most important section of the report. Here, the reader will **understand the processes and operational issues** of your project and how they contribute to the achievement of the objectives and the theory of change behind the project implementation.

Is possible that the project team's understanding of the development problems to be addressed with this project will have evolved or **changed** from those described when the project was originally submitted and approved. If that is the case, please share what motivated the change and what course of action has the project team identified.



The aim of the project is 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. At each stage of the project a number of organisations, individuals and language activists have been involved in achieving the project's objectives. The communication mechanism has been via personal visits, workshops and online discussion.

Initial System Testing:

The following Sindhi Language experts were involved in the initial testing of the proposed system:

1. Dr. Bharat Rattanpal, Faculty of Technology & Engineering, MSU Baroda.
2. Ms. MadhuriWardey, Faculty of Technology & Engineering, MSU Baroda.

The transliteration tables given below were tested and improved upon. The initial testing support was an important input and integral part of the system to examine whether the rule based engine output achieved as intended or not. It was analogous to conducting pilot study for any research.

Transliteration tables

Sindhi	Unicode	X Sindhi Unicode	Hindi	Hindi Unicode	X Hindi Unicode
ا	0627	0x27	अ	0905	0x05
ب	0628	0x28	ब	092C	0x2C
پ	067B	0x7B	ब्र	097F	0x7F
ڀ	0680	0x80	भ	092D	0x2D
ت	062A	0x2A	त	0924	0x24
ٿ	067F	0x7F	थ	0925	0x25
ڙ	067D	0x7D	ट	091F	0x1F
ڻ	067A	0x7A	ठ	0920	0x20
س	062B	0x2B	स	0938	0x38
پ	067E	0x7E	प	092A	0x2A
ج	062C	0x2C	ज	091C	0x1C
چ	0684	0x84	ज्ञ	097C	0x7C
ڙ	062C 06BE	0x2C 0xBE	झ	091D	0x1D
ڙ	0683	0x83	ञ	091E	0x1E
ڙ	0686	0x86	च	091A	0x1A
ڙ	0687	0x87	छ	091B	0x1B
ح	062D	0x2D	ह	0939	0x39
خ	062E	0x2E	ख	0959	0x59
د	062F	0x2F	द	0926	0x26
ڌ	068C	0x8C	ध	0927	0x27
ڌ	068F	0x8F	झ	097E	0x7E
ڌ	068A	0x8A	ड	0921	0x21
ڌ	068D	0x8D	ढ	0922	0x22
ڌ	0630	0x30	ज्ञ	095B	0x5B
ر	0631	0x31	र	0930	0x30
ڙ	0699	0x99	ड	095C	0x5C

Sindhi	Unicode	X Sindhi Unicode	Hindi	Hindi Unicode	X Hindi Unicode
ز	0632	0x32	ज़	095B	0x5B
س	0633	0x33	स	0938	0x38
ش	0634	0x34	श	0936	0x36
ص	0635	0x35	स	0938	0x38
ض	0636	0x36	ज़	095B	0x5B
ط	0637	0x37	त	0924	0x24
ظ	0638	0x38	ज़	095B	0x5B
ع	0639	0x39	-		
غ	063A	0x3A	ग	095A	0x5A
ف	0641	0x41	फ़	095E	0x5E
ق	06A6	0xA6	फ	092B	0x2B
ك	0642	0x42	क	0958	0x58
گ	06AA	0xAA	क	0915	0x15
خ	06A9	0xA9	ख	0916	0x16
گ	06AF	0xAF	ग	0917	0x17
گ	06B3	0xB3	गु	097B	0x7B
ڄ	06AF 06BE	0xAF 0xBE	घ	0918	0x18
ڄ	06B1	0xB1	ङ	0919	0x19
ل	0644	0x44	ल	0932	0x32
م	0645	0x45	म	092E	0x2E
ن	0646	0x46	न, ँ	0928, 0902	0x28, 0x02
پ	06BB	0xBB	ण	0923	0x23
و	0648	0x48	व	0935	0x35
ھ	06BE	0xBE	ह	0939	0x39
ي	064A	0x4A	य	092F	0x2F
ء	0621	0x21	-		
لا	FEFB	FEFB	ला	0932 093E	0x32 0x3E
لا	0627 06F1	0x27 0xF1	ला	0932 093E	0x32 0x3E

Online Discussion:

- Dr. Baldev Matlani, ex-head of Sindhi Department at Mumbai University.
- Mr. Rakesh Lakhani <http://rakeshlakhani.in/> Running Sindhi Blog.

As the tool was technically developed, a language expert and community activist helped in improving the outputs, as given below. They suggested what resources were lacking in the Sindhi language, such as data

dictionaries. They also highlighted important grammar and syntax rules that need to be carefully looked into for the development of an accurate rule based transliteration engine. Their motivation yielded enough encouragement to carry out the project at a rapid pace.

Meetings in Hyderabad, Sindh, Pakistan:

Between September 6-9th 2014, Dr Virinder Kalra made a trip to Hyderabad in Sindh in Pakistan to form links and inform activists and academics about the transliteration software. Meetings were held with the following people:

1. Anwar Figar Hakro chairman Dept of Sindhi, University of Sindh and staff members from the department.
2. Aijaz Wassan, Chairman, Department of Sociology, University of Sindh and colleagues.
3. Sindhi Language Authority, with Dr. Fehmida Hussain and Dr Taj Joyo.
4. Sindhi language activists: Zulfiqar Halepoto, Masood Lohar , ManzoorJokhio , Naseer Mirza.



Sindh University, Sindh, Pakistan



Sindhi Language Authority (SLA), Hyderabad, Sindh, Pakistan

First Sindhi Workshop, Punjabi University, Patiala, Punjab, India:

5 Days workshop of Sindhi Language Experts was organised at Punjabi University, Patiala from October 27-31, 2014. The following language experts participated and provided their valuable expert feedback for the development of the system.

1. Dr. M. K. Jetley, Vice Chairman, Sindhi Academy, Delhi.
2. Dr. Ravi ParkashTekchandani, Head, Sindhi Department, Modern Indian Languages, Delhi.
3. Dr. Baldev Matlani, Former Head, Mumbai University.
4. Dr. Murlidhar Bhavnani, Delhi.
5. Dr. Bharat Rattanpal, Faculty of Technology & Engineering, MSU Baroda.



Punjabi University Campus, Patiala, Punjab, India



Workshop in Progress, Punjabi University Campus, Patiala, Punjab, India

Transliteration Output

Input Sindi (Devanagri)

सिंधी साहित जे हिन डूख्ये दौर में, जंहिं में माण्हू सिंधी लिखणु पढ़णु त परे, गाल्हाइण में पिया हबिकनि. पर अजू बि कूझु माण्हूनि खे पंहिंजी भाषा लाइ मानु, सिक ऐं प्रेम आहे। अण्हनि मां हिक आहे शाइरु मुरली गोविन्दाणी. खेसि ज्ञाण आहे त हू जेको लिखे थो कूझु थोरा माण्हू ई पढंदा. पोइ बि हू लिखे थो, शायद इन लाइ त पंहिंजीअ भाषा में लिखण सां संदसि मन खे शांती मिले थी. शाइरी हून खे वरिसे में मिलियल आहे। संदसि पिता श्री नाराइणदास 'नारद' सुठा सूफी शाइर ऐं रागाई पिणि हूआ। शाइर मुरलीअ लिखण जी शुरुआत 1966 में हिंदी शाइरीअ सां ऐं 1990 खां पोइ सिंधीअ में कई। हेल् ताई हिन बिन्ही बोलियुनि में 500 कविताऊं लिखियूं आहिनि। रेडियो भुज तां 55 भेरा पंहिंजुं रचनाऊं पेश करे चुको आहे। बु किताब, हिक हिंदीअ में "करुणा उपहार" ऐं सिंधीअ में "रिश्तनि जो मरम्" (गज़ल) शायी थी चुका अथसि। हिन अदीब सिंधी साहित जे हर सिन्फ ते लिखियो आहे, चाहे जदीद शाइरी हूजे या कदीम शाइरी, पेश किताबु "महकियल मुर्क" जेको तव्हां जे हथनि में आहे इन में पढंदा त शाइर हर सिन्फ खे पेश कयो आहे, चाहे अहो वज़न ऐं मात्राअुनि वारो गज़लु हूजे या आज़ादु नज़्म। मां, शाइर मुरलीअ खे चडनि सालनि खां सुत्राणां। नौकरीअ जे रिटाइरमेंट खां पोइ गज़रियल टिनि वर्हयनि में सिंधी शाइरीअ में पंहिंजा परे ज़माए चुको आहे। जंहिंजो मिसालु आहे त घणनई रिसालनि में हिन जा गज़ल छपिजणु ऐं वक्त बि वक्त मुशाइरनि में शर्कत करणु। शाइर मुरलीअ जूं कवीताऊं दिलि खे छुहंदइ, ऐं साफ़ सुथरनि ख्यालनि वारियूं आहिनि।

Output Sindhi (Person/Arabic)

سنڌي ساهت جي هن ڏکي دور ۾، جنهن ۾ ماڻهو سنڌي لکڻ پڙهڻ ته پري، ڳالهائڻ ۾ پيا هڪن۔ پر اڄ به ڪجهه ماڻهن کي پنهنجي پاشا لاءِ مان، سک ۽ پريم آهي۔ انهن مان هڪ آهي شاعر مرلي گوونداڻي۔ کيس ڄاڻ آهي ته هو جيڪو لکي ٿو ڪجهه ٿورا ماڻهو ئي پڙهندا۔ پوءِ به هو لکي ٿو، شايد ان لاءِ ته پنهنجيءَ پاشا ۾ لکڻ سان سندس من کي شائتي ملي ٿي۔ شائري هن کي ورتي ۾ مليل آهي۔ سندس پتا شري نارائڻداس 'نارڊ' سنا صوفي شاعر ۽ راڳائي پڻ هئا۔ شاعر مرليءَ لکڻ جي شروعات 1966 ۾ هندي شائريءَ سان ۽ 1990 کان پوءِ سنڌيءَ ۾ ڪئي۔ هيل تائين هن بنهي بولين ۾ 500 ڪوتائون لکيون آهن۔ ريڊيو پيچ تان 55 ڀيرا پنهنجون رچنائون پيش ڪري چڪو آهي۔ ٻه ڪتاب، هڪ هنديءَ ۾ "ڪروڻا اڀار" ۽ سنڌيءَ ۾ "رستن جو مرم" ("گزل) شاپا ٿي چڪا اٿس۔

هن اديب سنڌي ساهت جي هر صنف تي لکيو آهي، چاهي جديد شائري هجي يا قديم شائري، پيش ڪتاب "مهڪيل مرڪ" جيڪو توهان جي هٿن ۾ آهي ان ۾ پڙهندا ته شاعر هر صنف کي پيش ڪيو آهي، چاهي اهو وزن ۽ ماترائن وارو غزل هجي يا آزاد نجم۔

مان، شاعر مرليءَ کي چڱن سالن کان سڃاڻان۔ نوڪريءَ جي رٽائرمينٽ کان پوءِ گذريل ٽن ورهين ۾ سنڌي شائريءَ ۾ پنهنجا ڀير ڄمائي چڪو آهي۔ جنهنجو مثال آهي ته گهڻئي رسالن ۾ هن جا غزل ڇپجڻ ۽ وڪت به وڪت مشائرن ۾ شرڪت ڪرڻ۔ شاعر مرليءَ جون ڪويتائون دل کي چهندڙ، ۽ صاف سٽرن ڪيالن واريون آهن۔

Second Sindhi Workshop, MSU Baroda, Gujarat, India:

A 5 days' workshop was organized on February 16-20, 2015 at MSU Baroda, India. The overall objectives of this workshop were:

1. Information Sharing and Knowledge Creation.
2. To get final feedback from Sindhi speaking Community.
3. To bring more social impact of this project.
4. Creating more awareness among Sindhi speaking regions outside Punjab.
5. To publicize or disseminate the results of the project.

The following language experts participated and provided their valuable expert feedback for the developing system.

1. Dr. Varsha Udhodacani, Lecturer, Sindhi Department, Faculty of Arts, MSU Baroda.
2. Tara C. Madhani, Secondary Teacher (Retd.), H.S.G.S. School, Vadodara.
3. Bharti R. Kewalramani, Principal (Retd.), Guest House, MSU Baroda.
4. Dr. Hundraj Balwani, Academic Secretary (Retd.), Text Book Board, Sector 10-A, Gandhinagar, Gujarat.
5. Subhash C. Sharma, Principal (Retd.), kubernagar, Ahmadabad.
6. H.K. Israni, Principal (Retd.), Gandhidham, Maitri Mandal, Adipur, kutch.
7. Seema Gurnani, Teacher (Retd.), Indian Institute of Sindhology, Adipur.
8. Vimmi Sadarangani, Associate Professor, Department of Sindhi, Tolani college of Arts & Science Adipur, kutch.



Workshop in Progress, MSU Baroda, Gujarat, India



Guest House, MSU Baroda, Gujarat, India

Activities that we struggle to implement:

1. Prior to the research project there was no coordinated mechanism for Sindhi language activists and experts to communicate with each other, as they have no state in India, which can support such activities. This was overcome through electronic facilitation. We thus created a network as part of the project to develop the software. We were careful to take into account a mix of gender and generation when involving people in this network, but were still not able to achieve a balance. This reflected the community of activists, academics and experts.
2. The online soft data for Sindhi (Dev. Script) is not available. We have to manually generate this soft data from printed published work like books magazines, newspapers etc. This aspect has affected the project objectives.
3. There is lack of ready to use computing resources for Sindhi language. Unfortunately, there is no any electronic dictionary between the two scrip and further, non-standardization of Sindhi (Devnagri) spellings has created a big challenge in terms of attaining higher transliteration accuracy at word level.



Input	Project activities	Outputs	Outcomes	Timeline	Status	Assessment
N/A	System Analysis and Design Detailed Analysis of Sindhi Language, Scripts and vocabulary for transliteration point of view	Sindhi Transliteration Tables Generated.	Sindhi Transliteration Tables Generated.	March, 2014	Completed	Challenges: Missing Diacritical marks and short Vowels Filling the Missing Script Maps Multiple Mappings for Perso-Arabic Characters Transliteration Ambiguity at Word level Word-Segmentation Issues
One Consultant and two Lexical Data Entry Operators Hired for 12 months and Three Laptops Purchased	Development of Natural Language Processing (NLP) Tools for Transliteration between Shindi Scripts	Developed Sindhi (dev) Typing tool Sindhi Electronic Dictionaries 3 million Word corpus of Sindhi (dev) 5 million Word corpus of Sindhi (Arabic)	N/A	Appointments: March 2014(2), and June 2014 April 30 2015	Completed	N/A
N/A	Offline Sindhi (dev) to Sindhi(Arabic) transliteration and Reverse	Offline Beta Version developed	Research Publication in ICON2014 Sangam: A Perso-Arabic to Indic Script Machine Transliteration Model	Feb 2015	Completed	The system has been designed using dictionaries database, mapping rules and language model
N/A	Online Perso Arabic-Indic script Transliteration System	Free Online Version developed	http://sangam.learnpunjabi.org Transliteration Scripts Urdu-Hindi Sindhi-Sindhi(dev) Shahmukhi-Gurmukhi	April 2015	Completed	AJAX & Dot Net web Service Technology

Project outputs, communication and dissemination activities

Tips: Take into account that the reader of your report has not been involved in project implementation, so readers do not have any further knowledge besides the information you are providing here.

This section of the report will allow you document the communication and dissemination efforts that the project team has conducted, which might be part of a specific communication strategy design as part of the project, or in place for the organization as a whole. When possible, please provide information about strategies in place and the rationale behind them.

Lessons can be learned from many aspects of project implementation, covering a wide variety of aspects such as technical, social, cultural and economical. Taking the rationale behind the project and its objectives can serve as a framework to draw your conclusions. Lessons can be identified by project partners, beneficiaries and general staff from the organization. A project diary and other activity records can serve as a tool to reflect during project team meetings and immediately after project activities are conducted.

Outputs are immediate, visible, concrete developmental change that is the tangible consequence of project activities, under direct control of the project team.

Example of possible outputs to report are:

- New products and Services (software, online platforms, applications);
- Information sharing and dissemination (publications, conferences, multimedia, social media);
- Knowledge creation (new knowledge embodied in forms other than publications or reports, such as new technologies, new methodologies, new curricula, new policies);
- Training (short-term training, internships or fellowships, training seminars and workshops) and
- Research Capacity (research skills; research management capacity and capacity to link research to utilization of research results).

Project outputs	Status	Assessment	Dissemination efforts
Sindhi Transliteration Tables	Completed, March, 2014	This is base of our Rule based Transliteration System	N/A
Sindhi (Devnagri) Typing Tool	Completed, April, 2014	Helpful for creating Sindhi (Devnagri) electronic Dictionary and Corpus data	N/A
Online Perso Arabic-Indic script Transliteration System <u>Demonstrated at</u> 	Completed, April, 2015	Transliteration Accuracy ranges between 92-97% at word level	 <p>Research Publication in ICON2014, Sangam: A Perso-Arabic to Indic Script Machine Transliteration Model, 18-21 December 2014, Goa University, Goa, India.</p> <p>'Sangam: A Perso Arabic-Indic Script Machine Transliteration System' has been accepted as a demonstration for the ICTD 2015 International Conference at Singapore, 15-18, May 2015.</p>

Table 1: Character level ambiguity

Unicode	Sindhi (Perso Arabic)	Sindhi(Devnagri)
0627(begin)	ا	अ
0627 (mid or end)	ا	अ, ा
0646 (begin)	ن	न
0646 (mid or end)	ن	न, ं, ॅ
0648 (begin)	و	व, ऊ, ओ, औ
0648 (mid or end)	و	व, ो, ौ, ु, ू, ऊ, ओ, औ

2. Missing diacritic symbols in Sindhi.

Diacritical marks are critical for correct pronunciation and sometimes even for disambiguation of certain words. The diacritical marks are also used for germination (doubling of a consonant) and mark the absence of a vowel following a base consonant. But the diacritical marks and short vowels are sparingly used in Perso-Arabic script writings. These missing diacritical marks and short vowels create substantial difficulties for transliteration systems, as the missing diacritic marks and vowels have to be guessed by the system and added for correct transliteration.

Table 2: Sindhi words with missing diacritics

Sindhi Word (without diacritic)	Equivalent words in Devnagri script
اندر	अंदरि, अंदरु
اڙي	अड़ी, अड़े
گسڪڻ	गिसिकणु, गिसिकणि
ڪٽ	खट, खटि

3. Ambiguity at word level

The causes of word ambiguity in Sindhi text are due to character level ambiguity and missing diacritics in the Sindhi text. The solution to this problem will be to develop a system for word sense disambiguation that is widely discussed area of NLP. Statistical language modeling has been widely used for such type of problems and higher-level language information will be needed to choose the most relevant word in target script.

Table 3: Ambiguity at word level in Sindhi

Sindhi Word	Equivalent words in Devnagri script
وکیل	वकील, विकेल
قسم	किस्मु, कसमु

4. Handling Typical Spellings

Transliteration is not trivial to automate, but transliteration of Perso-Arabic script to Devnagri script is even more challenging problem. Since the language does not change, so it becomes important the correct spellings and context of the (foreign/loan) words is maintained in target script.

Table 4: Words with Typical Spellings

اسکول	स्कूल
اسٹوڈیو	स्टूडीयो
انویسٹمنٹ	इनवैसटमेंट

Project outcomes

Tips: This section should be completed **ONLY** for the final report.

ISIF Asia expects you to report about the **outcomes** of the project as defined in the table below, based on the project implementation section of this report. Project team is encouraged to discuss the questions provided below to guide the reflection:

Can you identify and describe the relationships between the activities implemented and the social, economical, cultural and/or political benefits of your project implementation?

Outcomes can be defined as:

- Medium-term effects
- Effect of a series of achieved outputs
- Should capture the changes for the beneficiaries
- Take place during the life of project/strategy
- Influence but not direct control

A) Reaching through Social Media:

1. Twitter (Sindhi Voices)



The screenshot shows a tweet from the account 'Sindhi Voices' (@sindhivoices). The tweet text is: 'Software to melt India, Pakistan's Sindhi script barrier - The Times of India' followed by the URL 'fb.me/3jq9N5QZf'. Below the text are icons for reply, retweet, favorite, and a menu. The tweet is retweeted by 'Times of India' (TOI). The article preview shows the title 'Software to melt India, Pakistan's Sindhi script barrier - The Times...' and a snippet: 'Bringing down the script barrier between 25 lakh Sindhis in India and four crore in Pakistan, a first-of-its-kind software will enable Sindhis settled on both sides of the border to read each others'...'. The article is dated '5:50 PM - 3 Sep 2014'. A reply from 'Bob Biswas @InternetLully' is visible, asking '@sindhivoices what's a good book to learn Sindhi in Devanagari script?'.

2. Facebook (Sindhi Sangat)



India/Pakistan Media Coverage:

1. The project has been publicised in newspapers in India and Pakistan. The information has been spread in English as well as Hindi/Sindhi (Dev), Sindhi (Perso-Arabic) and Punjabi.
 - 1) Times of India, Circulated in India, Language: English
 - 2) Daily Sindh Express, Pakistan, Language: Sindhi (Person-Arabic)
 - 3) Sindhu Garjana, Faizabad, India, Language: Sindhi (Devanagri)
 - 4) Jagaran City, Patiala, Punjab India, Language: Hindi
 - 5) Jagbani, PunjabIndia, Language: Punjabi
 - 6) Jagaran City, Patiala, Punjab India, Language: Punjabi
2. 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.

News from Pakistan:

Language: Sindhi (Person-Arabic); **Date:** 2014/09/11, **Daily Sindh Express, Pakistan**

URL:

http://sindhexpress.com.pk/epaper/PoPupwindow.aspx?newsID=130187801&Issue=NP_HYD&Date=20140911



DrVirinder S. Kalra Oversees (UK) Project Coordinator During his visit to Pakistan

3. News Content published on 11/09/2014 in Daily Sindh Express Pakistan

سنڌي ڪمپيوٽر جي مدد سان ٻنهي اسڪرپٽن ۾ پڙهي سگهبي
ڊاڪٽر ورنڊر

ڪي ڪمپيوٽر جي مدد سان ڊيونو گري ۽ عربي اسڪرپٽ واري سنڌي
تبديل ڪرڻ وارو ڪم سال ۾ مڪمل ٿيندو

عربي اسڪرپٽ واري سنڌي ٻولي شاهوڪار آهي، جيڪا ڊيونو گري لکت
۾ پڙهندڙن کي تمام گهڻو فائدو پهچائيندي

پاڪ-هند ۾ ويزا جي ڏکيائين سبب تحقيقي ڪم مانچسٽر يونيورسٽي
حوالي ٿيو. ڊاڪٽر ورنڊر اديبن کي آگاهي ڏني

حيدرآباد (ريپورٽ: اسحاق مگريو) سنڌي ٻولي سنڌ ۽ هند ۾ رهندڙ
سنڌين وٽ عربي ۽ ڊيونو گري اسڪرپٽ ۾ ورهايل هئڻ سبب پٽيالا
يونيورسٽي جي تحقيق جي نتيجي ۾ ڪمپيوٽر جي مدد سان ڊيونو گري
لکت وارا عربي لکت ۾ ۽ عربي لکت وارا ڊيونو گري لکت ۾ آساني سان
پڙهي سگهندا. ان منصوبي تي ڪم ڪندڙ برطانيا جي مانچسٽر
يونيورسٽي جي پروفيسر ڊاڪٽر ورنڊر ڪالرا حيدرآباد ۾ سنڌي ٻولي
اختياري جي چيئر پرسن ڊاڪٽر فهميده حسين ۽ سيڪريٽري تاج جوڻي،
سنڌ يونيورسٽي ۾ سنڌي شعبي جي پروفيسر ڊاڪٽر انور فگار هڪڙو ۽
سوشالاجي جي ڊاڪٽر اعجاز وساڻ کان سواءِ سنڌي ٻولي جي اديبن
ذوالفقار هاليپوٽو، مسعود لوهار ۽ ريڊيو پاڪستان حيدرآباد جي ڊائريڪٽر
نصير مرزا سان ملاقاتون ڪيون ۽ جائيل منصوبي جي بابت کين ڄاڻ
ڏني. ڊاڪٽر ورنڊر ڪالرا پٽيالا جي پنجابي يونيورسٽي جي منصوبي
تحت ان کان اڳ ۾ ٻن لکتن ۾ ورهايل پنجابي جي گرمڪي ۽ شاه
مڪي، اردو ۽ هندي وارين لکتن کي ڪمپيوٽر جي مدد سان هڪٻئي ۾
بدلائي سگهڻ وارو ڪم ڪري چڪو آهي. ڊاڪٽر ورنڊر ڪمار جا وڏا
ورهاڱي کان اڳ ۾ ساهيوال جا رهواسي هئا، جيڪي بعد ۾ هندستان
جي پٽيالا ۾ هجرت ڪري ويا، جتان سندس وڏا برطانيا منتقل ٿي ويا.
47 سالن جي ڄمار جو ڊاڪٽر ورنڊر ڪالرا برطانيا ۾ ڄائو ۽ اتي پڙهي
مانچسٽر يونيورسٽي ۾ استاد مقرر ٿيو. هو چار ڪتاب لکي چڪو آهي
۽ ٽي عالمي ڪانفرنسون ۽ ٻه تحقيقي رپورٽون مرتب ڪري چڪو آهي.
ان کان سواءِ 23 تحقيقي مقالا پڻ لکي چڪو آهي. انفارميشن سوسائٽي

انويشين فنڊ جي 30 هزار ڊالرز جي مالي سهڪار سان پتيا لا جي پنجابي يونيورسٽي ڄاڻايل منصوبو هٿ ۾ کنيو آهي. ڊاڪٽر ورنڊر ڪالرا موجب ڏکڻ ايشيا ۾ اٽڪل 50 ڪروڙ ماڻهن پاران ڳالهائون ويندڙ چار ٻوليون جن ۾ سنڌي، پنجابي، هندي ۽ اردو شامل آهن. انهن ۾ پنجابي ٻولي گرمڪي ۽ شاهه مڪي اسڪرپٽ ۾ ورهايل آهي. سنڌي عربي ۽ ديوناگري اسڪرپٽ ۾ ورهايل آهي. ڄاڻايل منصوبي تحت ٻن ٻارڙن جي وٽيءَ ۾ ڳالهائي ويندڙ ساڳي ٻولي الڳ الڳ اسڪرپٽ ۾ لکجڻ ڪري ساڳي ٻولي ڳالهائيندڙن جي وچ ۾ سنڌو رهندو آيو آهي. پتيا لا يونيورسٽي پاران سنڌ ۾ سنڌي ٻوليءَ جي ماهرن سان مواد ۽ تحقيق جي مٽا مٽا جي سلسلي ۾ هندستان ۽ پاڪستان جي وچ ۾ ويزا جي ڏکيائي جي ڪري اهو تحقيقي ڪم برطانيا جي مانچسٽر يونيورسٽي جي پروفيسر ڊاڪٽر ورنڊر ڪالرا جي ذمي لڳايو ويو آهي. رپورٽ موجب انگريز ايام ڪاري دوران 1861ع ۾ سنڌ جي ڪمشنر سر ٻارٽل فيريئر انگريز عملدارن کي سنڌي سکڻ لاءِ معياري ٻولي جوڙڻ جو ڪم هٿ ۾ کنيو. ان وقت سنڌ ۾ نٿائي، خدا آبادي، لوهانڪي، ميمڻڪي، ڪوچڪي، ديونا گري، گرمڪي، هٿ وائڪي ٻوليون لکت ۾ رائج هيون. هندو عورتون وري گرمڪي ٻولي پڙهڻ ۽ لکڻ ۾ پڙ هيون. جڏهن ته سنڌ ۾ سرڪاري عملدار عربي اسڪرپٽ واري سنڌي ۾ لکڻ ۽ پڙهڻ جو ڪم ڪندڙ هئا. انگريزن پاران ابتدائي طور تي سڀ کان پهرين 1949 ۾ لکائي ويندڙ سنڌي تو انگريزي ڊڪشنري پڻ ديونا گري ۾ لکائي وئي هئي. ساڳي ريت سنڌي ۾ پهريون دفعو ترجمو ٿيندڙ بائيبل پڻ ديوناگري اسڪرپٽ ۾ لکيو ويو هو. اهڙي اختلاف سبب جڏهن سرڪاري طور تي سنڌي لکت لاءِ ڪهڙي اسڪرپٽ ۾ لکڻ گهرجي واري معاملي تي سنڌ جي عالمن، دانشورن ۽ سياستدانن ۽ انگريز ماهرن وچ ۾ پڻ سخت اختلاف ظاهر ٿيا. جن ۾ ڪيپٽن رچرڊ برٽن عربي لکت ۽ ڪيپٽن سيٽڪ ديونا گري ۾ لکت جو حامي هو. جنهن جي بنياد تي اسسٽنٽ ڪمشنر ايلس جي نگراني ۾ سنڌ جي عالمن جي ڪميٽي جوڙي وئي، جنهن ۾ ديوان

News Content Published in The Express Tribune, September 19th, 2014



HOME PAKISTAN BUSINESS WORLD SPORTS LIFE & STYLE MULTIMEDIA OPINION MAGAZINE BLOGS URDU NEWS LIVE NEWS

SINDH PUNJAB BALOCHISTAN KP & FATA JAMMU & KASHMIR GILGIT BALTISTAN

Google Custom Search

Transcending barriers: Software to break down the wall within the Sindhi language

By Z Ali

Published: September 19, 2014 **194** SHARES [SHARE](#) [TWEET](#) [EMAIL](#)

arabic اعراب						
خ	ح	ج	ث	ت	ب	أ
kha'	h'aa'	jiim	thaa'	taa'	baa'	alif
ص	ش	س	ز	ر	ذ	د
saad	shiin	siin	zaay	raa'	thaa'	daal
ق	ف	غ	ع	ظ	ط	ض
qaaf	feh'	ghayn	'ayn	thaa'	Taa'	daad
ي	و	ه	ن	م	ل	ك
yaa'	waaw	haa	nuun	miim	laam	kaaf

आ	व	व	भ	त	थ	ट	ठ	स	प	ज	झ	ञ	अ
a	v	v	bh	t	th	ṭ	ṭh	s	p	j	jh	ñ	a
[a]	[v]	[v]	[bh]	[t]	[th]	[ṭ]	[ṭh]	[s]	[p]	[j]	[jh]	[ñ]	[a]
च	छ	ह	ख	द	ध	ड	ड	ढ	र	ड	ड	ज	
c	ch	h	kh	d	dh	ḍ	ḍh	ḍh	r	ḍ	ḍ	j	
[c]	[ch]	[h]	[kh]	[d]	[dh]	[ḍ]	[ḍh]	[ḍh]	[r]	[ḍ]	[ḍ]	[j]	
स	श	स	ज	त	ज	ग	फ	फ	क	ख	ग	घ	
s	sh	s	j	t	j	g	f	f	k	kh	g	gh	
[s]	[sh]	[s]	[j]	[t]	[j]	[g]	[f]	[f]	[k]	[kh]	[g]	[gh]	
ग	ङ	ल	म	न	ण	व	ह	य					
g	ṅ	l	m	n	ṇ	v	h	y					
[g]	[ṅ]	[l]	[m]	[n]	[ṇ]	[v]	[h]	[y]					

DrVirenderKalra has embarked upon a one-year project of transliterating the Arabic and Devanagri scripts of Sindhi with the help of an internet-based software. PHOTO: EXPRESS

HYDERABAD:

The Sindhi people, divided by religion and borders, also read their mother tongue in two different scripts – Arabic and Devanagri. While Indian Sindhis do, to some extent, read the former, the present generation of Sindhis in Pakistan find Devanagri illegible.

“Not even a Sindhi language scholar has emerged in Sindh so far who is ‘au fait’ with the two scripts of the language,” writes Sindhi Language Authority (SLA) chairperson Dr. Fehmida Hussain in her booklet ‘Learning



the Sindhi Language'. "There is a treasure trove of literature, past and present, in both the scripts, but few Sindhis can benefit from it."

Literature is not the only realm where the divided people face an obstruction; the Sindhis on the two sides of the Pakistan-India border also have trouble communicating with each other in writing. However, with the help of technological advancement, this barrier is likely to be circumvented before long.

Dr. Virender Kalra, a lecturer from the University of Manchester, has embarked upon a one-year project of transliterating the two scripts with the help of an internet-based software. Visiting Hyderabad for two days, he met Sindhi writers, poets and SLA officials to gather their recommendations.

Meanwhile, Ali Hassan Mallah, a Jamshoro-based software manufacturer with a knack for Sindhi literature and poetry, claims to have already made transliteration software.

Dr. Virender Kalra

Having previously prepared transliteration software for the two Punjabi scripts – Shahmukhi, with Perso-Arabic letters, and Gurmukhi, with Landa letters derived from the Indic scripts – DrKalra has moved on to Sindhi and its scripts. Like the earlier project, this one is also funded by the Punjabi University in Patiala, India.

"The software will transliterate one script into the other without affecting the phonology of the word," he explained. However, the computer programme, which will be available on the Punjabi University's website, will not help the readers of one script decipher the other.

The majority of Sindhis in India live in the states of Rajasthan, Gujarat and Maharashtra. According to DrKalra, while the schools in these states teach Sindhi in both scripts, most of them use Devanagari.

"Literature has a rich vein of Sindhi in the Arabic script," said DrKalra. "A large number of books, newspapers, magazines and other reading materials are published in it. Indian Sindhis would benefit a lot if they were able to read these."

For SLA secretary Taj Joyo, DrKalra's undertaking is the need of the hour if Sindhis are to transcend their own language barrier.

News from India

<http://timesofindia.indiatimes.com/india/Software-to-melt-India-Pakistans-Sindhi-script-barrier/articleshow/41556896.cms>

9/12/2014

The Times of India
Title : Software to melt India-Pak Sindhi script barrier
Author : Amaninder Sharma
Location :
Patiala:
Article Date : 09/03/2014

Software to melt India-Pak Sindhi script barrier

Amaninder.Sharma
@timesgroup.com

Patiala: Bringing down the script barrier between 25 lakh Sindhis in India and four crore in Pakistan, a first-of-its-kind software will enable Sindhis settled on both sides of the border to read each other's literature despite the different scripts. The yet-to-be-launched software has been developed by Punjabi researchers in

Punjab University, Patiala, and Manchester University, England.

Despite having the same language, Sindhis residing on both sides of the border could not read each others' literature since Pakistani Sindhis use Perso-Arabic script and those in India follow the Devnagari script.

The software, which is in trial stage, will remove this barrier as it will transliterate Perso-Arabic Sindhi into

Devnagari and vice-versa.

"Like Punjabis, Sindhis also follow two scripts. Hence, the immense need to remove this language barrier. We had begun work on this project in March, last year. A Punjabi scholar from Manchester University is also collaborating on this," said Dr GS Lehal of Punjab University, coordinator of the project.

For the full report, log on to www.timesofindia.com



Language: Sindhi (Devanagari)

Date: 2014/11/(01-15), Sindhu Garjana, Faizabad, India

email: garjanasindhu@gmail.com

सिंधी व पंजाबी के कम्प्यूटरीकृत लिप्यन्तरण का साफ्टवेयर तैयार

0 अरबी-फारसी की सिंधी लिपि से देवनागरी सिंधी में लिप्यन्तरण का पंजाबी विश्वविद्यालय ने बनाया साफ्टवेयर 0 कम्प्यूटर का एक और क्रांतिकारी चरण

सिंधु गर्जना संवाद, पटियाला

पंजाबी विश्वविद्यालय पटियाला ने एक ऐसा साफ्टवेयर तैयार किया है जो कम्प्यूटर पर एक क्लिक भर करने से देवनागरी सिंधी को अरबी-फारसी की सिंधी लिपि में लिप्यन्तरित कर देगा और अरबी-फारसी की सिंधी लिपि को देवनागरी सिंधी में परिवर्तित कर देगा। इस मशीनी लिप्यन्तरण के मूल्यांकन एवं मानकीकरण के लिए 27 से 31 अक्टूबर तक विश्वविद्यालय में एक कार्यशाला का आयोजन हुआ। इसमें दिल्ली विश्वविद्यालय में सिंधी के एसोसिएट प्रोफेसर डॉ. रविप्रकाश टेकचंदानी, सिंधी के भाषा शास्त्री डॉ. मुरलीधर जेतली, डॉ. मुरली भावनानी, मुम्बई विश्वविद्यालय के पूर्व सिंधी विभागाध्यक्ष डॉ. बलदेव मटलाणी तथा डॉ. भारत रतनपाल शामिल हुए हैं।

यह सम्पूर्ण प्रकल्प पंजाबी विश्वविद्यालय के अन्तर्गत पंजाबी के तकनीकी उन्नतशील विकास केन्द्र के प्रभारी डा. गुरुप्रीत सिंह लहल कर रहे हैं। इनका साथ दे रहे हैं यहीं के डा. तेजेन्द्र सिंह सैनी और मेनचेस्टर विश्वविद्यालय के डॉ. वीरेन्द्र सिंह कालरा। डा. लहल के नेतृत्व में मशीनी लिप्यन्तरण प्रणाली को अभी जल्द ही विकसित किया गया है।

उन्होंने बताया कि इस कार्यशाला का उद्देश्य यह देखना है कि मशीनी लिप्यन्तरण किस स्तर तक उपयुक्त एवं तर्कसंगत है जिससे भविष्य में इसका प्रयोग शुरू किया जा सकेगा। इस सम्पूर्ण प्रकल्प पर आर्थिक सहयोग कर रहे हैं इन्फारमेशन सोसाइटी इनोवेशन फंड एण्ड कनेडियन इन्टरनेशनल डेवलपमेंट रिसर्च सेंटर के साथ स्वीडिश सरकार की स्वीडिश

इन्टर नेशनल डेवलपमेंट एजेंसी और एशिया पेसिफिक नेटवर्क इन्फारमेशन सेंटर।

उल्लेखनीय है कि डॉ. गुरुप्रीत सिंह लहल ने इसके पूर्व स्वीडिश इन्टरनेशनल डेवलपमेंट एजेंसी और एशिया पेसिफिक नेटवर्क इन्फारमेशन सेंटर की ओर से अन्तर्राष्ट्रीय स्तर पर शाहमुखी से गुरुमुखी एवं हिंदी से उर्दू

अनुवाद के प्रोजेक्ट पर सफलतापूर्वक कार्य किया है। इसी आधार पर सिंधी के प्रोजेक्ट पर विश्वविद्यालय को 29 हजार 819 आस्ट्रेलियन डालर की राशि कार्य करने के लिए मंजूर की गई है।

डॉ. लहल ने फोन पर सिंधु गर्जना को बताया कि लम्बे समय से मांग की जा रही थी कि सिंधी भाईचारे में धार्मिक और भौगोलिक.... (शेष पेज 6 पर)



Language: Hindi, Date: 2014/03/12, JagaranCity, Patiala, Punjab India

isif  asia



This work has been developed with the support provided by the Information Society Innovation Fund (ISIF Asia) – 2013, licensed under the Creative Commons Attribution-ShareAlike 3.0 Unported.

जागरण सिटी
पटियाला

लुधियाना/पटियाला, बुधवार, 12 मार्च 2014

अब सिंधी भाषा से अंजान नहीं रहेगी युवा पीढ़ी

▶▶ पंजाबी यूनिवर्सिटी को मिला सिंधी भाषा ट्रांसलेशन इंटरनेशनल प्रोजेक्ट ▶▶ सिंधी परिवारों को होगा लाभ

प्रदीप शाही, पटियाला

भारत और पाकिस्तान में बड़े पैमाने पर बोली जाने वाली सिंधी भाषा से अंजान हो रही युवा पीढ़ी को भाषा के मूल स्वरूप से अवगत करवाने के लिए पंजाबी यूनिवर्सिटी को इंटरनेशनल सिंधी भाषा ट्रांसलेशन प्रोजेक्ट सौंपा गया है। यह प्रोजेक्ट भारत-पाक के सिंधी परिवारों के लिए एक सेतु के रूप में पहचान स्थापित करेगा। यूनिवर्सिटी के कंप्यूटर साइंस विभाग के साथ इस इंटरनेशनल प्रोजेक्ट को एक साल में पूरा करने का करार हुआ है।

भारत-पाक में बड़ी तादाद में हैं सिंधी बोलने वाले

सिंधी भाषा को भारत और पाकिस्तान में बड़ी

विशेष
दैनिक जागरण



डॉ. गुरप्रीत सिंह लहलह

देरा का विभाजन होने के बाद भारत पहुंचे सिंधी परिवारों की युवा पीढ़ी मूल सिंधी भाषा से दूर हो चुकी है। यहां पर सिंधी भाषा में अब देवनागरी के शब्दों ने अधिक स्थान ले लिया

819 आस्ट्रेलियन डॉलर की राशि प्रदान की जाएगी।

कम होंगी धार्मिक व भौगोलिक दूरियां : डॉ. लहलह

विभाग के प्रो. डॉ. गुपुंधर सिंह लहलह ने बताया कि इस प्रोजेक्ट को सफल बनाने के लिए उनके साथ डॉ. तेजिंदर सिंह तथा डॉ. वरिंदर सिंह कालड़ा जुड़े हैं। लंबे समय से मांग की जा रही थी कि सिंधी भाईचारे में (भारत-पाक) धार्मिक और भौगोलिक दूरियां कम करने के लिए ऐसे सांस्कृतिक विकास किया जाए, जो दोनों देशों की लिपि को आपस में बदल सके। अब सिंधी को देवनागरी व देवनागरी की सिंधी में बदला जा सकेगा। इसके बाद युवा पीढ़ी अपनी मूल सिंधी भाषा से अंजान नहीं रहेगी।

Language: Punjabi, Date: 2014/05/04, Jagbani, PunjabIndia,

URL: <http://www.jagbani.com>

पੰजाब

139 देसां ਦੇ ਮੁਕਾਬਲੇ 'ਚ ਪੰਜਾਬੀ ਯੂਨੀਵਰਸਿਟੀ ਨੂੰ ਮਿਲਿਆ ਪ੍ਰੋਜੈਕਟ

☆ ਸਿੰਧੀ ਨੂੰ ਲਿੱਪੀਆਂਤਰ ਕਰਨ ਲਈ ਬੜੇਦਾ ਯੂਨੀਵਰਸਿਟੀ ਦੇ ਵਿਦਵਾਨਾਂ ਵਲੋਂ ਪੀ. ਯੂ. ਨੂੰ ਸੇਵਾਵਾਂ ਪ੍ਰਦਾਨ ☆ ਸਿੰਧੀ ਭਾਈਚਾਰੇ ਨੂੰ ਇਸ ਲਿੱਪੀਆਂਤਰ ਹੋਣ ਨਾਲ ਕਾਫੀ ਲਾਭ ਹੋਵੇਗਾ : ਡਾ. ਲਹਿਲਾ ਪ੍ਰੋਜੈਕਟ ਲਈ 29819 ਆਸਟ੍ਰੇਲੀਅਨ ਡਾਲਰ ਦੀ ਰਾਸ਼ੀ ਮਨਜ਼ੂਰ

ਪਟਿਆਲਾ, 4 ਮਈ (ਜੱਸ-ਨ) - ਸਿੰਧੀ ਬੋਲੀ ਨੂੰ ਲਿੱਪੀਆਂਤਰ ਕਰਨ ਦੇ ਪ੍ਰੋਜੈਕਟ ਨੂੰ ਪੂਰਾ ਕਰਨ ਲਈ ਗੁਜਰਾਤ ਦੀ ਬੜੇਦਾ ਯੂਨੀਵਰਸਿਟੀ ਦੇ ਦੋ ਵਿਦਵਾਨਾਂ ਨੇ ਆਪਣੀਆਂ ਸੇਵਾਵਾਂ ਪੰਜਾਬੀ ਯੂਨੀਵਰਸਿਟੀ ਦੇ ਹਵਾਲੇ ਕੀਤੀਆਂ ਹਨ, ਜਿਸ ਨਾਲ ਇਹ ਪ੍ਰੋਜੈਕਟ ਦੀ ਗੁਣਵੱਤਾ ਨੂੰ ਵੀ ਸਾਬਤ ਮਿਲੇਗਾ ਤੇ ਸਿੰਧੀ ਲੋਕਾਂ ਵਲੋਂ ਵਧੇਰੇ ਜਾਂਦੇ ਹੋਰ ਬਾਦਾਂ ਨੂੰ ਵੀ ਲਿੱਪੀਆਂਤਰ ਕਰਨ ਵਿਚ ਸਹਿਯੋਗ ਮਿਲੇਗਾ।

ਪੰਜਾਬੀ ਯੂਨੀਵਰਸਿਟੀ ਦੇ ਪੰਜਾਬੀ ਡਾ. ਸਾਹਿਤ ਅਤੇ ਸੱਭਿਆਚਾਰ ਦੇ ਤਕਨੀਕੀ ਵਿਕਾਸ ਦੇ ਉਚਿਤ ਮੁਕੱਦਮੇ ਦੇ ਡਾਇਰੈਕਟਰ ਅਤੇ ਕੰਪਿਊਟਰ ਸਾਇੰਸ ਵਿਭਾਗ ਦੇ ਪ੍ਰੋਫੈਸਰ ਡਾ. ਗੁਰਪ੍ਰੀਤ ਸਿੰਘ ਲਹਿਲਾ ਨੇ ਦੱਸਿਆ ਕਿ ਜਦੋਂ ਪਿਛਲੇ ਦਿਨੀਂ ਐਮ. ਐਸ. ਯੂਨੀਵਰਸਿਟੀ ਬੜੇਦਾ ਗਿਆ ਸੀ ਤਾਂ ਉਸ ਸਮੇਂ ਸਿੰਧੀ ਬੋਲੀ ਨੂੰ ਆਨ-ਲਾਈਨ ਸਿੰਧੀ ਲਿੱਪੀਅੰਤਰਨ ਦੇ ਮਿਲ

ਅੰਤਰਰਾਸ਼ਟਰੀ ਪੁੱਜੇ ਕੱਟ ਬਾਬਤ ਵਿਸ਼ਵਪ੍ਰਸਾਰਕ ਗੱਲ ਹੋਈ ਸੀ ਤਿਹ ਤਹਿਤ ਉਥੇ ਅਪਲਾਈਡ ਮੈਥੋਡਿਕਸ ਦੇ ਮਾਸਟਰ ਡਾ. ਡਾਤਰ ਤੋਂ ਕੰਪਿਊਟਰ ਸਾਇੰਸ ਦੇ ਮਾਸਟਰ ਡਾ. ਮਾਧੂਕੀ ਨੇ ਸਾਨੂੰ ਆਪਣੀਆਂ ਸੇਵਾਵਾਂ ਸੌਂਪ ਤੇ ਕਿਹਾ ਹੈ ਕਿ ਉਹ ਸਿੰਧੀ ਬੋਲੀ ਬਾਰੇ ਹਰ ਤਰ੍ਹਾਂ ਦੀਆਂ ਬਾਠੀਕੀਆਂ ਬਾਰੇ ਚਾਹਨਾ ਪਾਉਣਗੇ। ਉਨ੍ਹਾਂ ਕਿਹਾ ਕਿ ਇਸ ਪ੍ਰੋਜੈਕਟ ਬਾਬਤ ਅਸੀਂ ਹਰ ਵੀ ਮਾਹਿਰਾਂ ਤੋਂ ਮਦਦ

ਲਵਾਂਗੇ, ਉਥੇ ਕਿਹਾ ਗਿਆ ਸੀ ਕਿ ਸਾਡੀ ਸ਼ੈਲੀ ਤੇ ਆਨ-ਲਾਈਨ ਕੰਮ ਹੋਣਾ ਸਾਡੇ ਲਈ ਮਾਣ ਦੀ ਗੱਲ ਹੈ।

ਡਾ. ਲਹਿਲਾ ਨੇ ਦੱਸਿਆ ਕਿ ਇਹ ਪ੍ਰੋਜੈਕਟ ਦੋਹਾਂ ਦੇ ਮਿੱਤਰਾਂ ਦੇ 139 ਯੱਕਾਰਾਂ ਅਤੇ ਖੋਜ ਸੰਸਥਾਵਾਂ ਦਰਮਿਆਨ ਕਰਤਬ ਮੁਕਾਬਲੇ ਉਪਰੰਤ ਯੂਨੀਵਰਸਿਟੀ ਦੇ ਇਸ ਕੰਮ ਨੂੰ ਪ੍ਰਾਪਤ ਹੋਇਆ ਹੈ ਇਸ ਵਿੱਚ

ਆਸਟ੍ਰੇਲੀਅਨ ਡਾਲਰ ਦੀ ਰਾਸ਼ੀ ਮਨਜ਼ੂਰ ਕੀਤੀ ਗਈ ਹੈ। ਇਹ ਰਾਸ਼ੀ ਕੈਨੇਡੀਅਨ ਇੰਟਰਨੈਸ਼ਨਲ ਰਿਸਰਚ ਸੈਂਟਰ, ਸਕੀਡਨ ਦੀ ਸਕੈਂਡਿਨੇਵੀਅਨ ਡਿਵੈਲਪਮੈਂਟ ਏਜੰਸੀ ਅਤੇ ਡੇਵੀਡ ਪੈਮੋਰਿਕ ਨੈਟਵਰਕ ਇਨਵੈਸਟਮੈਂਟ ਸੈਂਟਰ ਦੇ ਸਹਿਯੋਗ ਨਾਲ ਕਾਰਜ ਕੀਤੇ ਇਨਵੈਸਟਮੈਂਟ ਸੁਸਾਇਟੀ ਇਨਵੈਸਟਮੈਂਟ ਫੰਡ ਵਿੱਚ ਪ੍ਰਾਪਤ ਹੋਈ ਹੈ। ਕੇਂਦਰ ਵਲੋਂ ਲਿੱਪੀ ਦੀਆਂ ਕੰਪਾਂ ਨੂੰ ਖਰੀਦ ਕਰਨ ਲਈ ਪਹਿਲਾਂ ਹੀ ਗੁਰਮੁਖੀ-ਸਾਧੂਮੁਖੀ ਅਤੇ ਉਰਦੂ-ਦੇਵਨਾਗਰੀ ਲਿੱਪੀਅੰਤਰਨ ਪ੍ਰੋਗਰਾਮ ਵਿਕਾਸ ਕੀਤਾ ਜਾ ਚੁੱਕਾ ਹੈ। ਇਹ ਪ੍ਰੋਜੈਕਟ ਇਸ ਖੇਤਰ ਦੀ ਇਕ ਹੋਰ ਪਹਿਲਕਦਮੀ ਹੈ।

ਜਿਕਰਯੋਗ ਹੈ ਕਿ ਪਾਕਿਸਤਾਨ ਵਿਚ ਸਿੰਧੀ ਸੁਭਾਨ ਸਿੰਧੀ (ਅਰਬੀ-ਫਾਰਸੀ) ਲਿੱਪੀ ਵਿਚ ਅਤੇ ਭਾਰਤ ਵਿਚ ਸਿੰਧੀ ਸੁਭਾਨ ਦੇਵਨਾਗਰੀ ਅਤੇ ਸਿੰਧੀ (ਅਰਬੀ-ਫਾਰਸੀ) ਲਿੱਪੀਆਂ ਵਿਚ ਲਿਖੀ ਜਾਂਦੀ ਹੈ। ਡਾ. ਲਹਿਲਾ



ਡਾ. ਲਹਿਲਾ ਜਾਣਕਾਰੀ ਦਿੰਦੇ ਹੋਏ | (ਜੱਸ-ਨ)

Language: Punjabi

Date: 2014/03/12

JagaranCity, Patiala, Punjab India

ਭਾਰਤ-ਪਾਕਿ ਸਿੰਧੀ ਪਰਿਵਾਰਾਂ ਲਈ 'ਪੁਲ' ਪੰਜਾਬੀ ਯੂਨੀਵਰਸਿਟੀ

ਨਵਦੀਪ ਢੀਗਰਾ, ਪਟਿਆਲਾ

ਭਾਰਤ ਤੇ ਪਾਕਿਸਤਾਨ ਵਿਚ ਵੱਡੀ ਗਿਣਤੀ ਵਿਚ ਬੋਲੀ ਜਾਣ ਵਾਲੀ ਸਿੰਧੀ ਭਾਸ਼ਾ ਤੋਂ ਅਣਜਾਣ ਹੋ ਰਹੀ ਨੌਜਵਾਨ ਪੀੜ੍ਹੀ ਨੂੰ ਭਾਸ਼ਾ ਦੇ ਮੁਲਰੂਪ ਨਾਲ ਜਾਣੂ ਕਰਵਾਉਣ ਲਈ ਹੁਣ ਪੰਜਾਬੀ ਯੂਨੀਵਰਸਿਟੀ ਇਕ 'ਪੁਲ' ਦਾ ਕੰਮ ਕਰੇਗੀ। ਪੰਜਾਬੀ ਯੂਨੀਵਰਸਿਟੀ ਦੇ ਕੰਪਿਊਟਰ ਸਾਇਸ ਵਿਭਾਗ ਨੂੰ ਇਸ ਸਬੰਧੀ ਇਕ ਵਿਸ਼ੇਸ਼ ਪ੍ਰਾਜੈਕਟ ਮਿਲਿਆ ਹੈ ਜਿਸਨੂੰ ਇਕ ਸਾਲ ਦੇ ਅੰਦਰ ਪੂਰਾ ਕੀਤਾ ਜਾਵੇਗਾ।

ਸਿੰਧੀ ਭਾਸ਼ਾ ਦੀ ਭਾਰਤ ਤੇ ਪਾਕਿਸਤਾਨ ਵਿਚ ਰਹਿਣ ਵਾਲੇ ਵੱਡੀ ਗਿਣਤੀ ਲੋਕ ਵਰਤੋਂ ਕਰ ਰਹੇ ਹਨ। ਪਾਕਿਸਤਾਨ ਵਿਚ ਇਸ ਭਾਸ਼ਾ ਨੂੰ ਬੋਲਣ ਵਾਲਿਆਂ ਦੀ ਗਿਣਤੀ ਕਰੀਬ ਸਾਢੇ ਚਾਰ ਤੋਂ ਪੰਜ ਕਰੋੜ ਤੱਕ ਹੈ। ਜਦੋਂਕਿ ਭਾਰਤ ਵਿਚ ਇਸ ਭਾਸ਼ਾ ਨੂੰ ਬੋਲਣ ਵਾਲਿਆਂ ਦੀ ਗਿਣਤੀ 50 ਲੱਖ ਤੋਂ ਵੱਧ ਹੈ। ਭਾਰਤ ਵਿਚ ਸਿੰਧੀ ਭਾਸ਼ਾ ਦੀ ਵਰਤੋਂ ਕਰਨ ਵਾਲੇ ਜ਼ਿਆਦਾਤਰ ਪਰਿਵਾਰ ਮੁੰਬਈ ਤੇ ਕੱਛ ਖੇਤਰਾਂ ਵਿਚ ਰਹਿੰਦੇ ਹਨ।

ਪਾਕਿਸਤਾਨ ਦੇ ਸਿੰਧੀ ਪ੍ਰਦੇਸ਼ ਵਿਚ ਸਿੰਧੀ ਭਾਸ਼ਾ ਵਿਚ ਅਰਬੀ ਤੇ ਫਾਰਸੀ ਦੇ ਅੱਖਰਾਂ ਦੀ ਵਰਤੋਂ ਹੁੰਦੀ ਹੈ। ਭਾਰਤ-ਪਾਕਿ ਵੰਡ ਮਗਰੋਂ ਭਾਰਤ ਪੁੱਜੇ ਸਿੰਧੀ ਪਰਿਵਾਰਾਂ ਦੀ ਨੌਜਵਾਨ ਪੀੜ੍ਹੀ ਆਪਣੀ ਭਾਸ਼ਾ ਤੋਂ ਦੂਰ ਹੋ ਚੁੱਕੀ ਹੈ ਤੇ ਸਿੰਧੀ ਭਾਸ਼ਾ ਵਿਚ ਹੁਣ ਦੇਵਨਾਗਰੀ ਦੇ ਸ਼ਬਦਾਂ ਦੀ ਜ਼ਿਆਦਾ ਵਰਤੋਂ ਹੋਣ ਲੱਗ ਪਈ ਹੈ। ਇਸੇ ਕਾਰਨ ਅੱਜ ਦੋਹਾਂ ਦੇਸ਼ਾਂ ਵਿਚ ਇਸ ਭਾਸ਼ਾ ਦੇ ਲਿਖਣ ਵਿਚ ਕਾਫੀ ਫਰਕ ਆ ਚੁੱਕਾ ਹੈ। ਪੰਜਾਬੀ ਯੂਨੀਵਰਸਿਟੀ ਨੂੰ ਮਿਲੇ



ਇਸ ਨਵੇਂ ਪ੍ਰਾਜੈਕਟ ਤੋਂ ਬਾਅਦ ਦੋਹਾਂ ਦੇਸ਼ਾਂ ਦੀਆਂ ਭਾਸ਼ਾਵਾਂ ਨੂੰ ਇਕ ਬਰਾਬਰ ਪੜ੍ਹਨਾ ਆਸਾਨ ਹੋ ਜਾਵੇਗਾ।

ਏਸ਼ੀਅਨ ਦੇਸ਼ਾਂ ਵਿਚ ਜ਼ਿਆਦਾ ਬੋਲਿਆਂ ਜਾਣ ਵਾਲੀਆਂ ਭਾਸ਼ਾਵਾਂ ਤੋਂ ਨੌਜਵਾਨ ਪੀੜ੍ਹੀ ਨੂੰ ਜੋੜਨ ਲਈ ਭਾਸ਼ਾਵਾਂ ਦਾ ਸਹੀ ਢੰਗ ਨਾਲ ਅਨੁਵਾਦ ਕਰਨ ਨੂੰ ਪਹਿਲ ਦਿੱਤੀ ਜਾ ਰਹੀ ਹੈ ਤਾਂ ਜੋ ਭਾਸ਼ਾਵਾਂ ਨੂੰ ਜਿਊਂਦਾ ਰੱਖਿਆ ਜਾ ਸਕੇ। ਇਸ ਪ੍ਰਾਜੈਕਟ ਤਹਿਤ ਯੂਨੀਵਰਸਿਟੀ ਨੂੰ 29 ਹਜ਼ਾਰ 819 ਅਸਟ੍ਰੇਲੀਅਨ ਡਾਲਰ ਦੀ ਰਾਸ਼ੀ ਪ੍ਰਦਾਨ ਕੀਤੀ ਜਾਵੇਗੀ।

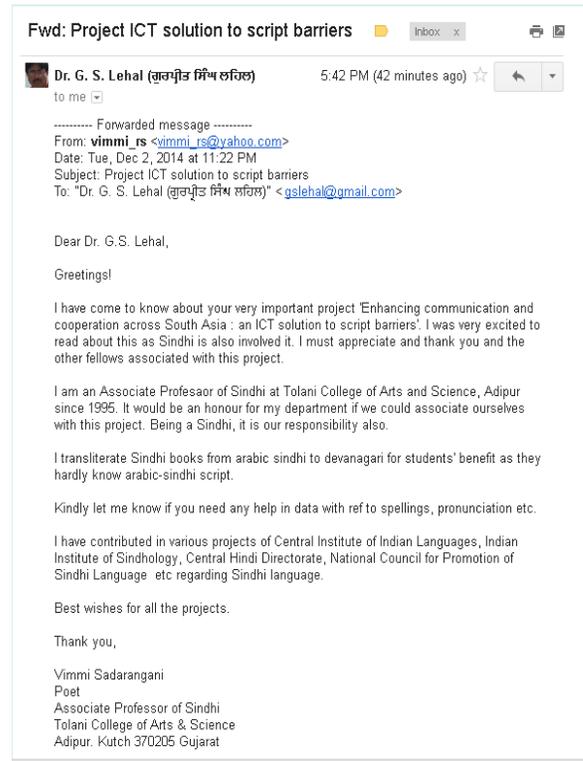
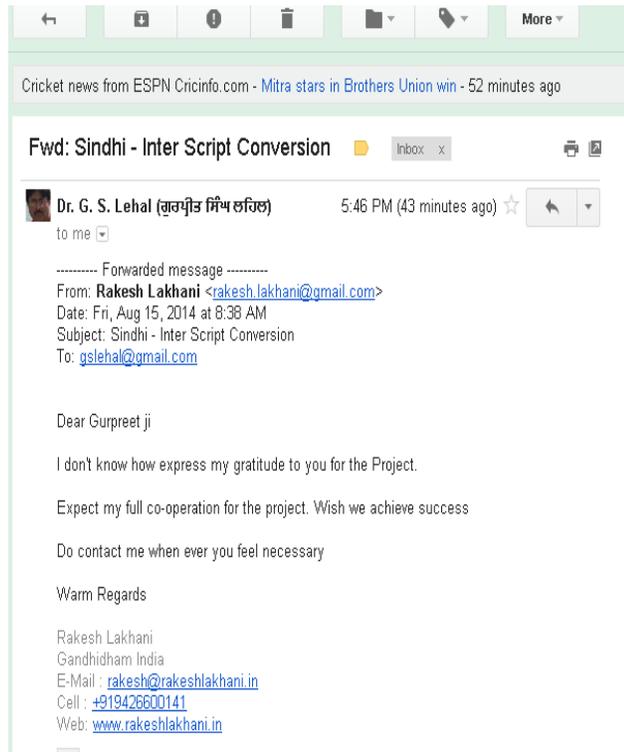
ਪੰਜਾਬੀ ਯੂਨੀਵਰਸਿਟੀ ਕੰਪਿਊਟਰ ਸਾਇਸ ਵਿਭਾਗ ਦੇ ਪ੍ਰੋ.ਡਾ. ਗੁਰਪ੍ਰੀਤ ਸਿੰਘ ਲਹਿਲ ਨੇ ਦੱਸਿਆ ਕਿ ਇਸ ਪ੍ਰਾਜੈਕਟ ਨੂੰ ਸਫਲ ਬਣਾਉਣ ਲਈ ਉਨ੍ਹਾਂ ਦੀ ਟੀਮ ਵਿਚ ਡਾ. ਤਜਿੰਦਰ ਸਿੰਘ, ਡਾ. ਵਰਿੰਦਰ ਸਿੰਘ ਕਾਲੜਾ ਸ਼ਾਮਲ ਹਨ। ਲੰਮੇ ਸਮੇਂ ਤੋਂ ਮੰਗ ਕੀਤੀ ਜਾ ਰਹੀ ਸੀ ਕਿ ਸਿੰਧੀ ਭਾਈਚਾਰੇ ਵਿਚ ਧਾਰਮਿਕ ਤੇ ਭੂਗੋਲਿਕ ਦੂਰੀਆਂ ਨੂੰ ਘਟ ਕਰਨ ਦੇ ਲਈ ਅਜਿਹਾ ਸਾਫਟਵੇਅਰ ਤਿਆਰ ਕੀਤਾ ਜਾਵੇ ਜੋ ਦੋਹਾਂ ਦੇਸ਼ਾਂ ਦੀ ਲਿੱਪੀ ਨੂੰ ਆਪਸ ਵਿਚ ਬਦਲ ਸਕੇ। ਹੁਣ ਸਿੰਧੀ ਨੂੰ ਦੇਵਨਾਗਰੀ ਤੇ ਦੇਵਨਾਗਰੀ ਨੂੰ ਸਿੰਧੀ ਵਿਚ ਬਦਲਿਆ ਜਾ ਸਕੇਗਾ। ਇਸ ਸਾਫਟਵੇਅਰ ਦੇ ਆਉਣ ਤੋਂ ਬਾਅਦ ਸਿੰਧੀ ਪਰਿਵਾਰਾਂ ਦੀ ਨੌਜਵਾਨ ਪੀੜ੍ਹੀ ਆਪਣੀ ਭਾਸ਼ਾ ਨੂੰ ਮੁਢ ਤੋਂ ਜਾਣ ਸਕੇਗੀ।

B) Response from Sindhi Community:

As an impact of media coverage, we have 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 us and participated in the 5 days' 2nd Sindhi workshop at MSU Baroda, India.

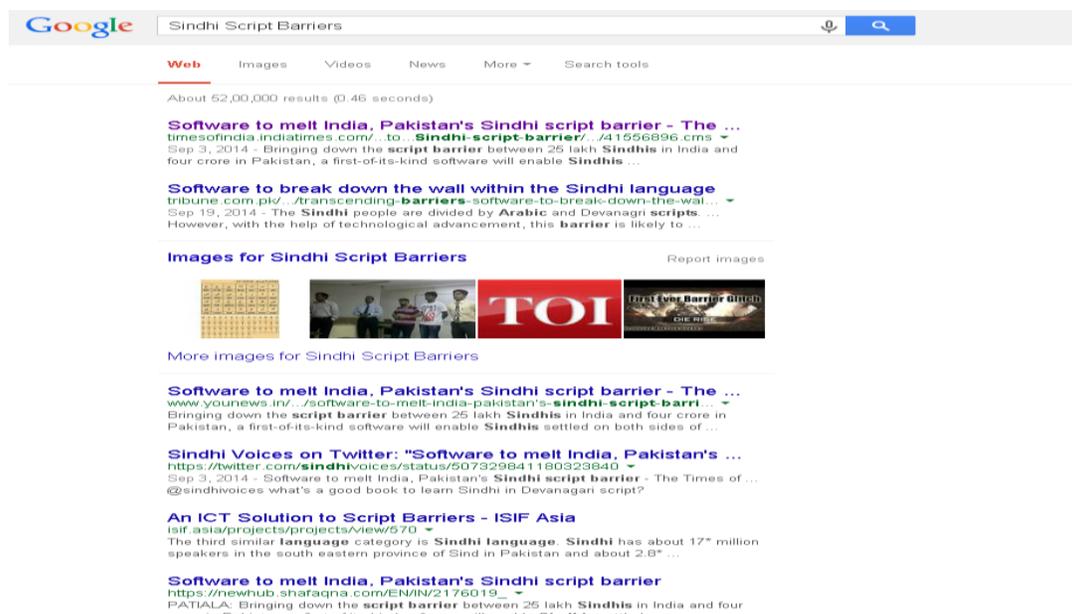


Appreciation from Project Beneficiaries



C) Media Impact through Google Search:

If anyone can search Google with three keywords **Sindhi Script Barriers** more than one dozen links are pointing to our project.





Project management and sustainability

Tips: Please comment on the general project **administration, staffing, procurement**, etc. specially those aspects contributing to the fulfilment of the project objectives as well as those that have delay project implementation.

Indicate **how the project team has strengthened its capacity** and work towards sustainability with the support provided by ISIF Asia? (new equipment, training, improved administrative skills, lessons learned from the project). Has the organization increased its research or administrative skills of the team involved? Has the project allowed for a particular contribution to capacity building of women or marginalized social groups? Special attention should be paid to the expected or unexpected impact on marginalized social groups.

Have you done **anything different** to provide administrative support for this project **besides your “business as usual”** processes and procedures? Has the project inspired change inside your organization?

Sustainability is to be examined not only in terms of staff retention and financial stability of the organization supporting the project but about the communities’ appropriation of benefits perceived from project implementation.

The ISIF Asia Secretariat is very interest to learn if this project has generated opportunities for future development (new funding from partnerships, sponsorships, investment or other funding mechanisms), please provide details.

Please explain if the ISIF Asia grant has helped to consolidate your organization and how. If any of the project activities will continue after the end of the ISIF Asia grant, please describe how your organization is planning to support future developments.

The proposed project needed an expert and committed team to work round the clock for meeting the time based objectives. The existing work force of the Advanced Centre formed the substantial part of the team with additional support from the Department of Computer Science and the University Administration team. However, additional language experts were also needed for the purpose of project implementation and these were accessed through meetings and workshops. They imparted learning and intensive training sessions to the existing team members. This led to retention of existing staff.

The hardware needed was locally procured by exploring the existing internal procurement channels. The software was also suitably upgraded. Another important aspect of project implementation involved reorganizing the department functioning to project based system supported by skilled team members specifically selected and trained for the purpose.

The funding received was invariably intended to enable organization to undertake aforementioned steps and to provide enough financial support for self-sustenance of the department.

This grant has supported the further development of the Advanced Centre, such that it is the leading computational linguistics department in Asia. It has already attracted Indian government investment due to the opportunities created by the ISIF fund and will seek to do more in future.

The existing team of the Advanced Centre at Punjabi University, Patiala will continue to develop and maintain the transliteration software using internal resources as well as to measure and follow the impact of the internet-based software. With the help of this technological advancement, Sindhis on the two sides of the Pakistan-India border now have no trouble in communicating with each other, in writing.

Future developments of the transliteration system would be to incorporate a host of other Indian languages sharing similar speech but having different scripts for writing, for example, Kashmiri and Konkani.

Impact

Tips: This section of the report does not refer to the project activities, but about the “**bigger picture**”. It will be desirable if the project team can reflect on the **impact that the project has contributed to as part of other actions implemented by your organization and/or your partners.**

Impact refers to the influence the project may had on the way people does things through the use or adoption of the project outputs; changes in the context the project was implemented; changes in the community the project has been working with; and/or changes inside the organizations that have participated in the implementation or the relationships established through the project's implementation.

Impact is often impossible to measure in the short term and is rarely attributable to a single activity. Impact can be linked to a vision or long-term development goal that your organization might be working towards.

It can be identified as a logical consequence of achieving a combination of outputs and outcomes.

Impact is usually measurable after the project life and is outside the direct control of the project team and the organization.

It is too early to reflect on the extent to which communication between India, Pakistan and the Sindhi diaspora has of yet improved. However, the methodology of using digital platforms to enhance communication has been demonstrated in our previous work on transliteration. Our methodology of working across the India-Pakistan border (via the UK) is also a fundamental shift in the way in which projects can be implemented. This is in the context of fluctuating political relations between India and Pakistan, which can make research and social impact difficult to engage with.

Overall Assessment

Tips: This section of the report is extremely valuable for the ISIF Asia secretariat as it provides evidence about the role and relevance of ISIF Asia contributions in the Asia Pacific region.

Tips: Briefly provide **your own views** on the value and importance of the project relative to the proposed innovation, investment of time, effort and funding involved. Include the strengths and weaknesses of the project and the steps taken to strengthen the credibility and reliability.

This is your opportunity to conduct a **team reflection about the value of the project for the organization.** The following questions might help you to prepare a substantive overall assessment.

- To what extent the project meet its objectives?
- What were the most important findings and outputs of the project? What will be done with them?
- What contribution to development did the project make?
- Were certain aspects of project design, management and implementation particularly important to the degree of success of the project?
- To what extent the project help build up the research capacity of your institution or of the individuals involved?
- What lessons can be derived that would be useful in improving future performance?

The overall objective of the project has been achieved with the launch of Internet based SANGAM transliteration software. This bi-directional, web based Sindhi Language Transliteration Tool is designed to facilitate written communication among Sindhis on the two sides of the Pakistan-India border and those living anywhere on the planet. As per the project's commitment, the transliteration accuracy of the system ranges between 92-97%, which is a bit lower than hoped, but still we are satisfied with the progress we have achieved so far. The following challenges are faced in Sindhi Script Transliteration:

1. Ambiguity at character level.

2. Missing diacritic symbols in Sindhi.
3. Ambiguity at word level.
4. Handling typical spellings.

In order to achieve high transliteration accuracy, we employed both rule-based as well as statistical approaches to Sindhi script transliteration.

Research Contribution

We have presented and published the research work in ICON2014, “Sangam: A Perso-Arabic to Indic Script Machine Transliteration Model”, 18-21 December, 2014, Goa University, Goa, India and SANGAM system demonstration has been successfully presented during ICTD 2015 International Conference at Singapore, 15-18, May 2015.

Sindhi Literature in bi-scripts for Wider Visibility:

Dr Fehmida Hussain of Pakistan in her book ‘Learning the Sindhi Language’ said, “There is a treasure trove of literature, past and present, in both the scripts, but few Sindhis can benefit from it.”



Thus, this is our first effort to publish Sindhi literature in bi-scripts for wider visibility. Using our typing tool and transliteration software we have digitized and transliterated one book of Dr. M. K. Jetley, Vice Chairman, Sindhi Academy, Delhi, published in India in Sindhi (Dev) script into Sindhi (Perso-Arabic) for people living in Pakistan.

Other contributions are:

1. Bi-directional web based Sindhi Language Transliteration website.
2. Typing Tool for Sindhi (Devanagri).
3. Language models developed to support statistical transliteration.
4. Bi-script electronic dictionaries.

Research Capacity

The center has previously worked on and is currently undertaking three projects, which are of direct relevance to the present application. We have completed the Shahmukhi to Gurmukhi Transliteration Solution for Networking and the Transliteration system from Urdu/Kashmiri to Roman script and Urdu/Hindi Transliteration. The gained experience has contributed significantly to achieve the desired objectives. Earlier, the development of such systems had taken two or more years but this time we have completed the project in one-year duration.

In addition to this the Sindhi language consultant and Sindhi Workshops/Meetings organized in India and Pakistan had played a major role in construction of lexical resources and transliteration rules for the script transliteration engine. We explored the raw script corpus for getting character alignments between the two scripts. To achieve present accuracy, we have implemented statistical and rule based techniques including the usage of n-gram word analysis for transliterating Sindhi text in Urdu script to Devnagri and reverse.

Unlike the existing systems, the SANGAM web based bi-directional transliteration system has been deployed using advanced state of the art technology. Developing such system has definitely impact on the improvement of individual's capability and knowledge.

On the basis of what we learned from past research is that we have developed a common platform for all Perso-Arabic and Indic scripts. As acknowledged earlier, the most important lesson learnt by the project team was that the transliteration system could be extended with ease to incorporate a host of other Indian languages sharing similar speech but having different scripts for writing, for example, like Kashmiri and Konkani languages of India.

Recommendations

Tips: Include any recommendations in this section that you and your project team, the organizations supporting the project and the community you worked with, would like to make to other practitioners or researchers on the field facing similar problems or implementing similar solutions.

Please take a minute to share recommendations with the ISIF Asia secretariat that might help to improve the support provided.

Arguably, Computers have massively aided in the increase in communication via text and speech over the Internet throughout the world; however barring a few number of the world's languages which currently enjoy the privileges/benefits of modern language technologies such as machine translation/Transliteration and speech recognition, the majority of languages of the world have not been able to enjoy measurable technological upliftment and have been marginalized. These under-resourced languages suffer from glaring frailties, both observed and obvious, such as limited presence on the web; lack of linguistic expertise; and most importantly lack of electronic resources for language processing, such as; monolingual corpora; bilingual electronic dictionaries; parallel corpora; pronunciation dictionaries, vocabulary tables, etc.

Therefore, we strongly recommend and urge the researchers to contribute towards this urgent need to develop such lacking resources for all such possible languages around us. The strengthened electronic database will provide ease to other language researchers to do more research in this area.

Bibliography

Tips: Include complete bibliographic references to all sources (printed, on-line, quotes, etc) used to prepare the different sections of this report. The APA style guide offers examples about how to reference a variety of sources. <http://www.apastyle.org/learn/quick-guide-on-references.aspx>(as accessed on 3/7/2013).

1. The Times of India, Software to Melt India-Pak Sindhi Script Barrier, published on 09/03/2014
<http://timesofindia.indiatimes.com/india/Software-to-melt-India-Pakistans-Sindhi-script-barrier/articleshow/41556896.cms>
2. Daily Sindh Express Pakistan, News published on 11/09/2014
http://sindhexpress.com.pk/epaper/PoPupwindow.aspx?newsID=130187801&Issue=NP_HYD&Date=20140911
3. Gurpreet Singh Lehal and Tejinder Singh Saini, "Sangam: A Perso-Arabic to Indic Script Machine Transliteration Model", Proceedings of 10th International Conference on Natural Language Processing, Goa India (2014).
4. 'Sangam: A Perso Arabic-Indic Script Machine Transliteration System' has been presented at the demonstration track of International Conference on Information and Communication Technologies and Development ICTD 2015 at Singapore, 15-18, May 2015
5. Department Information <http://www.learnpunjabi.org>
6. University Information <http://www.punjabiuniversity.ac.in>
7. SANGAM Transliteration System <http://sangam.learnpunjabi.org>
8. Shahmukhi to Gurmukhi Transliteration System <http://s2g.learnpunjabi.org>
9. Urdu-Hindi Transliteration System <http://uh.learnpunjabi.org>