

RESUME'

Dr. TEJINDER SINGH

#127/1, Urban Estate,

Phase-I, Patiala

pin-147 002

0091-9463616269 (M)

email: tej74i@gmail.com; tej@pbi.ac.in

Educational Qualification

1. PhD (Computer Science),
2. Master of Computer Application (Punjabi University, Campus),
3. B. Sc. (Non-Medical)

Degree	Subject	University	Year	Division
1. M.C.A. (Campus)	Computer Application	Punjabi University, Patiala	1999	1st
2. Ph.D.	Computer Application	Punjabi University, Patiala	2012	Degree awarded

Current Job Experience (May 2006 to till date):

Presently I am working as a System Manager (Technical Officer-III) in Research Centre for Punjabi Language Technology, Punjabi University Patiala. We are working as a team on various National/International projects and in-house NLP tools for Indian Languages. I have completed my PhD research in 2012 in computer science under the faculty of physical sciences. The brief description of the research is as follows:

Machine transliteration is an important problem in an increasingly multilingual world as it plays a critical role in many downstream applications such as machine translation or cross-lingual information retrieval systems. The multitude of foreign languages and mutually incomprehensible scripts of the same language pose a barrier to information exchange as we cannot all learn every language or script in use worldwide. Therefore, if we can get around the language barrier or at least the script barrier, we can access much more of the world's culture and can explore its abundant richness.

This research work is an effort to fulfill the above stated objective for Punjabi language. Incidentally, the existence of Shahmukhi (Urdu script used for writing Punjabi Language) and Gurmukhi scripts for Punjabi has created a script barrier between the Punjabi literature written in India and in Pakistan. Notably, more than 60 per cent of Punjabi literature of medieval period (500-1450 AD) is available in Shahmukhi script only, while most of the modern Punjabi writings are available in both scripts. Hence, a machine transliteration system that overcomes script barriers of this cultural extravaganza needs to handle these Punjabi scripts with different origins, different direction of writings, different set of alphabet, and different kind of writing system conventions.

Therefore, PhD research describes a complete Shahmukhi to Gurmukhi script transliteration system which has been implemented and tested on available online and offline datasets. An overall 97% word accuracy has been achieved. To the best of our knowledge, this is the first time that an attempt has been made to develop a Shahmukhi to Gurmukhi transliteration system that can run on Shahmukhi script with missing short vowels and other diacritics. However, in the existing system study, we found one rule based system for Shahmukhi script with manually pre-processing task of adding missing short vowels and other diacritical marks before processing the input text. Clearly, as compared to our methodology the approach of existing system is not a practical in nature, not very challenging and simple to implement. We have studied in depth the problems in the various stages of the development of a complete transliteration system. In the

beginning stage we found that there is an urgent need for InPage to Unicode converter for developing Unicode based resources and that have been achieved. This achievement has further contributed to gradually develop a large Shahmukhi corpus. The statistical analysis of Shahmukhi and Gurmukhi corpora give the insight of scripts, their similarities and differences, Pakistani dialect and its influence on written forms etc.

The transliteration engine works well with Shahmukhi script with missing short vowels, but a careful evaluation had shown that like Urdu and Persian Shahmukhi text has affected with two types of segmentation problems. Therefore, to enhance the output accuracy at the transliteration stage we have implemented two separate segmentation algorithms for space insertion and space omission problems in Shahmukhi script. Furthermore, the post-processing task of word disambiguation is addressed with two promising solutions within the limitation of availability of lexical resources.

Almost all the aspects related to corpus based transliteration of Shahmukhi text have been touched upon. Surprisingly, the implemented unique solutions for the challenging tasks like supplying missing vowels, word segmentation detection and word disambiguation for multiple interpretations are achieved using target script resources. **In addition, script writing rules, conversion rules, machine learning with heuristics are utilized in the various phases of transliteration.**

a) Research Projects Completed:

Year	Funding Agency	Project	My role	Budget
2016-2017	PSEB, Mohali	Online Punjabi Teaching	Co-Coordinator Ongoing	31 Lakh
2014-2015	The Information Society Innovation Fund (ISIF) Grants, Australia	Enhancing Communication and Co-operation across South Asia: An ICT Solution to Script Barriers	Co-Coordinator	AU\$ 29,819
2009-2010	The Information Society Innovation Fund (ISIF) Grants, Australia	Web based Transliteration and Translation System between Urdu and Hindi Languages	Co-Coordinator	US\$ 29,896
2006-2008	Pan Asia ICT Grants, Singapore	Shahmukhi to Gurmukhi Transliteration Solution for Networking	Project Developer	US\$ 30,000

b) Research Publications:

Book:

Title: **Shahmukhi To Gurmukhi Transliteration System**
(*Addressing Script Challenges With Statistical And Rule Based Approach*)

ISBN-13: **978-3-659-26200-5**

ISBN-10: **3659262005**

EAN: **9783659262005**

Published by: LAP Lambert Academic Publishing

Number of pages: 236

Published on: 2012-11-26

Journals:

1. T S Saini and G S Lehal "Shahmukhi to Gurmukhi Transliteration System: A Corpus based Approach", Research in Computing Science (Mexico), Volume 33, pp. 151-162 (2008). **Refereed and Indexed**
2. Gurpreet Singh Lehal, Tejinder Singh Saini, Pritpal Kaur Buttar, "Automatic Bilingual Legacy-Fonts Identification and Conversion System", Research in Computing Science, Vol. 86, pp. 9-23 (2014). **Refereed and Indexed**
3. Kawarbir Singh Dhanju, Gurpreet Singh Lehal, Tejinder Singh Saini and Arshdeep Kaur, "Design and Implementation of Shahmukhi Spell Checker", Indian Journal of Science and Technology, Vol 8(27), pp.1-12. 2015, DOI: 10.17485/ijst/2015/v8i27/83917
4. Gurpreet Singh Lehal, and Tejinder Singh Saini, "A Transliteration Based Word Segmentation System for Shahmukhi Script", Proceedings of International Conference on Information Systems for Indian Languages, ICISIL 2011, Patiala, Communications in Computer and Information Science, Vol. 139, **Springer-Verlag** Berlin Heidelberg, Germany, pp. 136-143. (2011). **Peer-Reviewed and Indexed**
5. Gurpreet Singh Lehal, Tejinder Singh Saini and V. S. Kalra, "Urdu to Hindi and Reverse Transliteration System", Proceedings of International Conference on Information Systems for Indian Languages, ICISIL 2011, Patiala, Communications in Computer and Information Science, Vol. 139, **Springer-Verlag** Berlin Heidelberg, Germany, pp. 305-306. (2011) **Peer-Reviewed and Indexed**

Conference Proceedings:

6. Saini, T. S., Lehal, G. S., and Kalra, V. S. 2008. Shahmukhi to Gurmukhi transliteration system. In 22nd international Conference on Computational Linguistics: Demonstration Papers (Manchester, United Kingdom, August 18 - 22, 2008). Coling: International Conference on Computational Linguistics. Association for Computational Linguistics, Morristown, NJ, 177-180.
7. G. S. Lehal and T. S. Saini, "A Hindi to Urdu Transliteration System", Proceedings of 8th International Conference on Natural Language Processing, pp. 235-240, Kharagpur, India. (2010)
8. Tejinder Singh Saini and Gurpreet Singh Lehal, "Word Disambiguation in Shahmukhi to Gurmukhi Transliteration", Proceedings of the 9th Workshop on Asian Language Resources, **IJCNLP 2011, Chiang Mai, Thailand**, pp. 79-87. (2011)
9. Gurpreet Singh Lehal and Tejinder Singh Saini, "Conversion between Scripts of Punjabi: Beyond Simple Transliteration", Proceedings of the COLING 2012: Posters, Mumbai, pp. 633-642. (2012)
10. Gurpreet Singh Lehal and Tejinder Singh Saini, "Development of a Complete Urdu-Hindi Transliteration System", Proceedings of the COLING 2012: Posters, Mumbai, pp. 643-652. (2012)
11. Gurpreet Singh Lehal, Tejinder Singh Saini and Savleen Kaur Chowdhary, "An Omni-Font Gurmukhi to Shahmukhi Transliteration System", Proceedings of the COLING 2012: Demonstration papers, Mumbai, pp. 313-320. (2012)
12. 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. 33-56. (2014)

C) Invited Talks/Resource Person:

1. Subject expert in two day workshop on "ਪੰਜਾਬੀ ਭਾਸ਼ਾ ਦੀ ਵਰਤੋਂ ਲਈ ਸਾਫਟਵੇਅਰ ਅਤੇ ਐਪਸ ਦੀ ਜਾਣਕਾਰੀ" Organised by Punjabi Academy Delhi, New Delhi, 22-23 March, 2017.

2. Expert Speaker on "**Transliteration Challenges between Perso Arabic and Indian Scripts**" in the TEQIP-II sponsored short term training programme on soft computing, Pattern recognition and Image processing, organised by DCS&E at Sant Longowal Institute of Engineering & Technology from September 27-October 01, 2016.
3. **International Mother Language Day 2016: Punjabi Typing Training.** Organized by Reserve Bank of India (RBI), Chandigarh
4. **Conversion between Scripts of Punjabi: Beyond Simple Transliteration,** TEQIP-II Sponsored FDP on Natural Language Processing, Organized by Department of Computer Science and Engineering, GNE, Ludhiana, 13-17, January 2014
5. **Shahmukhi to Gurmukhi Transliteration System,** UGC and CSIR sponsored one week National Workshop on Multilingual Technologies, Organized by Department of Computer Science, Punjabi University, Patiala, 11-17, November 2013
6. **Demonstrated e-Governance and Localization Prototype for Punjab Police,** Workshop on National e-Governance Plan (NeGP), Organized at Punjabi University, Patiala with Department of IT, Ministry of Communication & IT Govt. of India, December 22, 2011
7. **Internet and You,** Technology up-Gradation Workshop for Administrative Staff, Organized by Department of Distance Education, Punjabi University, Patiala, 4th-6th March 2011
8. **ਪੰਜਾਬੀ ਭਾਸ਼ਾ ਦੇ ਕੰਪਿਊਟਰੀਕਰਨ ਵਿੱਚ ਪੰਜਾਬੀ ਯੂਨੀਵਰਸਿਟੀ ਪਟਿਆਲਾ ਦਾ ਯੋਗਦਾਨ,** National Seminar on "ਕੰਪਿਊਟਰ ਉਤੇ ਪੰਜਾਬੀ ਭਾਸ਼ਾ ਦੀ ਵਰਤੋਂ ਤੇ ਉਪਯੋਗਤਾ", Organized at Government Post-Graduate College, Karnal with the support of Haryana Punjabi Sahitya Akademi , Govt of Haryana, 26th February 2010
9. **Technical Development of Punjabi Language,** Refresher course in Computer Science and Application organized by Academic Staff College, Punjabi University, Patiala, 20th May, 2010
10. **Shahmukhi to Gurmukhi Transliteration System and other NLP Tools,** Workshop on NLP Generation Next for World Communication, Organized by Chitkara Institute of Engineering & Technology, 21-23, December 2009

d) Software and Technologies developed/Contributed:

1. **Counseling Software** for B.Ed. 2008, Punjabi University, Patiala
2. **PunjabiKhoj,** Customized Search Engine for Punjabi (Project Developer)
"www.punjabikhoj.com" (*Released by Sardar Parkash Singh Badal, Hon'ble Chief Minister of Punjab in 2008*)
3. Online **Gurmukhi Unicode Typing Pad**
4. Offline **Unitype** for Windows (**Team Member**)

5. **InPage Urdu Text to Unicode** Converter (Used in Thesis Plagiarism)
6. Multimedia enabled **Punjabi (Gurmukhi & Shahmukhi) to English Dictionary**
7. **Sodhak: Online/Offline Punjabi Spellchecker (Team Member)**
8. **Web based Sangam: A Perso-Arabic to Indic Script Machine Transliteration Model**
9. **Shahmukhi Spell Checker (Team Member)**
10. **Akhar-IWP (Indic Word Processor) (Team Member)**

e) Working Knowledge:

- Web Technologies HTML/JavaScript/JQuery/CSS/PHP
- XML
- Operating System
- Object Oriented Programming with JAVA, JDBC
- Linux & Unix
- C# and C++
- Web Programming using Microsoft Asp.NET

Scholarship: National Level scholarship from 7th to 10th standard.

Conference/Event Organized as Team Member:

1. **iRafit-2012: Member Web Content Development Committee**, Organized by DCS, Punjabi University, Patiala
2. **icisil-2011: Internet Facility and Technical Support**, Organized by DCS, Punjabi University, Patiala
3. **Techdisha-2004: Technical Support**, Organized by Tech. Society of DCSE, Punjabi University, Patiala

Guest Faculty:

1. DCS, Punjabi University, Patiala
2. Department of Distance Education (DDE), Punjabi University, Patiala
3. UCoE, Punjabi University, Patiala
4. PGDCA (Evening) at University Computer Centre.

Personal Details:

<i>Father's Name</i>	Late sh. Gurdev Chand Saini
<i>Date of Birth</i>	October 24th 1973
<i>Status</i>	Married / Male / Indian
<i>Hobbies</i>	Painting
<i>Languages known</i>	Punjabi, English, Hindi and Urdu

Country Visited:

- ❖ United Kingdom (UK) 2008

DOM: Oct/2016