

M G Abbas Malik



Address Résidence Houille Blanche–F763, 2 av. des jeux olympiques, 38029 Grenoble
 Nationality Pakistani
 Date of Birth 14 October 1979
 Voice: Mob: +33 (0)6 18 37 00 06 Off: +33 (0)4 76 51 43 79
 Email: abbas.malik@gmail.com abbas.malik@imag.fr
 Home Page: www.puran.info

Objective To develop Computational Linguistic resources for South Asian Languages and making them available for research and development

Education

2006 – Date: *Doctorate in Computer Sciences*
 Department of Computer Sciences and Applied Mathematics
 University of Joseph Fourier (Grenoble) www.ujf-grenoble.fr

2005 – 06: *Master Research in Computational Linguistics*
 Department of Computer Sciences & Linguistics
 University of Paris 7 <http://www.univ-paris-diderot.fr>

2001 – 03: *Master in Computer Sciences*
 Punjab University College of Information Technology,
 University of the Punjab, www.pucit.edu.pk

Experience

2009 – Date: Lecturer, University of Joseph Fourier (Grenoble), France
 Working as lecturer in Department of Computer Sciences and Applied Mathematics (IMA – Informatique et Mathématiques Appliqués) from September 2009.

2006 – Date: Research Scholar at University Joseph Fourier, Grenoble, France
 Working as researcher and developer. Doing research on Machine Translation (MT), Natural Language Processing (NLP) and mainly working on computation of languages of India and Pakistan and their translation. Demo software are available at www.puran.info

Nov 2002 – Jun 2005: European Union – Punjab Lok Sujag (NGO)
 Worked as System Analyst and principal researcher on a **European Union** project PLATT (Punjabi Language and Transliteration Tool). I am also responsible for the project management. This project was a research project for designing a software which will be able to transliterate the Shahmukhi (Right to left script used for Punjabi in Pakistan and derived from Persian/Arabic script) into Gurmukhi (Left to right script used for Punjabi in India and derived from old script of Indian Subcontinent named Landa) and vice versa. See Demo version at www.puran.info/pmt

Workshop Organized

South and Southeast Asian Natural Language Processing (SSANLP) Workshop
 I am organizing a workshop on South and Southeast Asian Natural Language Processing with special emphasis *Morphology and Segmentation* (the scope of the present workshop) at **International Conference on Computational Linguistics COLING – 2010**, August 2010, Beijing, China.

Publications

A Hybrid Model for Urdu Hindi Transliteration (2009)
 Named Entities Workshop: Joint Conference the 47th Annual Meeting of the Association of Computational Linguistics (ACL) and the 4th International Joint Conference on Natural Language Processing, August 2009, Singapore.
<http://www.aclweb.org/anthology/W/W09/W09-3536.pdf>

Hindi Urdu Machine Transliteration using Finite-state Transducers (2008)
 22nd International Conference on Computational Linguistics (COLING 08), August 2008, Manchester, UK.
<http://aclweb.org/anthology/C/C08/C08-1068.pdf>

- Punjabi Machine Transliteration (2006)**
 21st International Conference on Computational Linguistics (COLING 06) and
 44th Annual Meeting of ACL, pages 1137 – 1144, July 2006, Sydney,
 Australia.
<http://www.aclweb.org/anthology/P/P06/P06-1143.pdf>
- Towards a Unicode Compatible Punjabi Character Set (2005)**
 27th Internationalization and Unicode Conference, April 2005, Berlin,
 Germany.
- Talks Delivered**
- Combining Finite-state Machines and Statistical Language Models for Urdu to Hindi Transliteration (July 2009)** at XEROX Research Center Europe (XRCE), Grenoble, France.
- Computational Linguistics: a Case Study of Punjabi (January 2009)** at *Oriental College, University of the Punjab, Lahore, Pakistan*
- Statistical Machine Translation (December 2008)** at *Punjab University College of Information Technology, University of the Punjab, Lahore, Pakistan*
- Urdu/Punjabi Font Development (May 2005)** at *Punjab University College of Information Technology, University of the Punjab, Lahore, Pakistan*
- Software Development**
- Punjabi Machine Transliteration Tool (Puran 1.0)**
 This system transliterates Pakistani Punjabi (Shahmukhi – a right to left script, derived from Perso-Arabic script) into Indian Punjabi (Gurmukhi – a left to right script like Devanagari). www.puran.info/pmt
- Hindi Urdu Machine Translation System (HUMT 1.0)**
 HUMT system translates Hindi into Urdu and vice versa.
www.puran.info/humt
- Font Development**
- Fajer Pakistani Nasta'liq 1.0 beta** (font for Urdu, Punjabi, Sindhi, Pashto, Kashmiri and Balochi, 2008)
 This font is the continuation of my previous font Fajer Noori Nasta'liq. This may first ever try to give a taste of Nasta'liq to Sindhi and Pashto languages. According to my best knowledge, no one ever tried to give a Nasta'liq font for Sindhi and Pashto languages. I have also enhanced Fajer Noorin Nasta'liq font for other languages like Kashmir and Balochi. This font is also made according to current international standards of Unicode and Open Type Font (OTF) technology.
- Fajer Noori Nasta'liq 1.0 beta** (font for Urdu/Punjabi - 2003)
 This font is among the pioneer fonts that were developed for Urdu language in accordance with the international language encoding standards – Unicode and the current font development technology – Open Type Font (OTF) technology. This font is the first ever Nasta'liq font for Punjabi language. I have developed this font as *Final Year Project* for my Masters degree in **Computer Sciences at Punjab University College of Information Technology, University of the Punjab, Lahore, Pakistan**. It is also used in the European Union's research project.
- Punjabi Naskh 1.0 beta** (font for Urdu/Punjabi, 2003)
 I have developed this Naskh font for Urdu and Punjabi using the international standards for font development and the first ever standard code page for Punjabi language in accordance with Unicode, described in my paper "Towards a Unicode Compatible Punjabi Character Set". This is the first ever font for Punjabi language using Unicode and OTF.
- Professional Memberships**
- Member of ACL (Association of Computational Linguistics)
 - Member of ACM (Association of Computing Machines)
- Tools, Languages and Operating Systems**
- OS**
- Windows XP
 - Windows 2000

- Windows 98
- Linux

Languages

- C/C++
- Visual C/C++
- Visual C/C++ .Net
- Visual Basic .Net
- Java
- SQL
- HTML/DHTML

Tools

- XEROX Research Center Europe (XRCE) Integrated Environment for Linguistic Programming (XIP)
- XEROX Finite-state Tool (XFST)
- Moses (A decoder for statistical MT)
- SRILM (A statistical language modeling toolkit)
- GIZA++ (Statistical Alignment Tool)
- Microsoft Office (Word, Excel, Access, Power point, etc.)
- Microsoft SQL Server 2000
- Microsoft VOLT (Visual Open Type Font Layout Tool)
- Font Lab 4.6 (Font designing tool)
- Scan Font (Font designing tool)

Languages

- English (Bilingual)
- French (Good)
- Punjabi (Mother Tongue)
- Urdu (Mother Tongue)
- Hindi (Very Good)

References

- Dr. Christian Boitet, Prof. of Computer Sciences, University of Joseph Fourier (Grenoble), France. Christian.Boitet@imag.fr
- Dr. Pushpak Bhattacharyya, Prof. of Computer Sciences, Indian Institute of Technology, Bombay, India. pb@cse.iitb.ac.in