

# **Named Entities Workshop: Shared Task on Transliteration 2009**

**(NEWS 2009)**

**Held at ACL-IJCNLP 2009**

**Singapore  
7 August 2009**

**ISBN: 978-1-61738-457-8**

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2009) by the Association for Computational Linguistics  
All rights reserved.

Printed by Curran Associates, Inc. (2010)

For permission requests, please contact the Association for Computational Linguistics  
at the address below.

Association for Computational Linguistics  
209 N. Eighth Street  
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006  
Fax: 1-570-476-0860

[acl@aclweb.org](mailto:acl@aclweb.org)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2634  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

## Table of Contents

<i>Report of NEWS 2009 Machine Transliteration Shared Task</i> Haizhou Li, A Kumaran, Vladimir Pervouchine and Min Zhang .....	1
<i>Whitepaper of NEWS 2009 Machine Transliteration Shared Task</i> Haizhou Li, A Kumaran, Min Zhang and Vladimir Pervouchine .....	19
<i>Automata for Transliteration and Machine Translation</i> Kevin Knight .....	27
<i>DirectTL: a Language Independent Approach to Transliteration</i> Sittichai Jiampojamarn, Aditya Bhargava, Qing Dou, Kenneth Dwyer and Grzegorz Kondrak . . .	28
<i>Named Entity Transcription with Pair n-Gram Models</i> Martin Jansche and Richard Sproat .....	32
<i>Machine Transliteration using Target-Language Grapheme and Phoneme: Multi-engine Transliteration Approach</i> Jong-Hoon Oh, Kiyotaka Uchimoto and Kentaro Torisawa .....	36
<i>A Language-Independent Transliteration Schema Using Character Aligned Models at NEWS 2009</i> Praneeth Shishtla, Surya Ganesh V, Sethuramalingam Subramaniam and Vasudeva Varma . . . . .	40
<i>Experiences with English-Hindi, English-Tamil and English-Kannada Transliteration Tasks at NEWS 2009</i> Manoj Kumar Chinnakotla and Om P. Damani .....	44
<i>Testing and Performance Evaluation of Machine Transliteration System for Tamil Language</i> Kommaluri Vijayanand .....	48
<i>Transliteration by Bidirectional Statistical Machine Translation</i> Andrew Finch and Eiichiro Sumita .....	52
<i>Transliteration of Name Entity via Improved Statistical Translation on Character Sequences</i> Yan Song, Chunyu Kit and Xiao Chen .....	57
<i>Learning Multi Character Alignment Rules and Classification of Training Data for Transliteration</i> Dipankar Bose and Sudeshna Sarkar .....	61
<i>Fast Decoding and Easy Implementation: Transliteration as Sequential Labeling</i> Eiji Aramaki and Takeshi Abekawa .....	65
<i>NEWS 2009 Machine Transliteration Shared Task System Description: Transliteration with Letter-to-Phoneme Technology</i> Colin Cherry and Hisami Suzuki .....	69
<i>Combining a Two-step Conditional Random Field Model and a Joint Source Channel Model for Machine Transliteration</i> Dong Yang, Paul Dixon, Yi-Cheng Pan, Tasuku Oonishi, Masanobu Nakamura and Sadaoki Furui	72

<i>Phonological Context Approximation and Homophone Treatment for NEWS 2009 English-Chinese Transliteration Shared Task</i>	
Oi Yee Kwong .....	76
<i>English to Hindi Machine Transliteration System at NEWS 2009</i>	
Amitava Das, Asif Ekbal, Tapabrata Mondal and Sivaji Bandyopadhyay .....	80
<i>Improving Transliteration Accuracy Using Word-Origin Detection and Lexicon Lookup</i>	
Mitesh Khapra and Pushpak Bhattacharyya .....	84
<i>A Noisy Channel Model for Grapheme-based Machine Transliteration</i>	
Jia Yuxiang, Zhu Danqing and Yu Shiwen .....	88
<i>Substring-based Transliteration with Conditional Random Fields</i>	
Sravana Reddy and Sonjia Waxmonsky .....	92
<i>A Syllable-based Name Transliteration System</i>	
Xue Jiang, Le Sun and Dakun Zhang .....	96
<i>Transliteration System Using Pair HMM with Weighted FSTs</i>	
Peter Nabende .....	100
<i>English-Hindi Transliteration Using Context-Informed PB-SMT: the DCU System for NEWS 2009</i>	
Rejwanul Haque, Sandipan Dandapat, Ankit Kumar Srivastava, Sudip Kumar Naskar and Andy Way .....	104
<i>A Hybrid Approach to English-Korean Name Transliteration</i>	
Gumwon Hong, Min-Jeong Kim, Do-Gil Lee and Hae-Chang Rim .....	108
<i>Language Independent Transliteration System Using Phrase-based SMT Approach on Substrings</i>	
Sara Noeman .....	112
<i>Combining MDL Transliteration Training with Discriminative Modeling</i>	
Dmitry Zelenko .....	116
<i><math>\epsilon</math>-extension Hidden Markov Models and Weighted Transducers for Machine Transliteration</i>	
Balakrishnan Varadarajan and Delip Rao .....	120
<i>Modeling Machine Transliteration as a Phrase Based Statistical Machine Translation Problem</i>	
Taraka Rama and Karthik Gali .....	124
<i>Maximum n-Gram HMM-based Name Transliteration: Experiment in NEWS 2009 on English-Chinese Corpus</i>	
Yilu Zhou .....	128
<i>Name Transliteration with Bidirectional Perceptron Edit Models</i>	
Dayne Freitag and Zhiqiang Wang .....	132
<i>Bridging Languages by SuperSense Entity Tagging</i>	
Davide Picca, Alfio Massimiliano Gliozzo and Simone Campora .....	136
<i>Chinese-English Organization Name Translation Based on Correlative Expansion</i>	
Feiliang Ren, Muhua Zhu, Huizhen Wang and Jingbo Zhu .....	143
<i>Name Matching between Roman and Chinese Scripts: Machine Complements Human</i>	
Ken Samuel, Alan Rubenstein, Sherri Condon and Alex Yeh .....	152

<i>Analysis and Robust Extraction of Changing Named Entities</i> Masatoshi Tsuchiya, Shoko Endo and Seiichi Nakagawa .....	161
<i>Tag Confidence Measure for Semi-Automatically Updating Named Entity Recognition</i> Kuniko Saito and Kenji Imamura .....	168
<i>A Hybrid Model for Urdu Hindi Transliteration</i> Abbas Malik, Laurent Besacier, Christian Boitet and Pushpak Bhattacharyya .....	177
<i>Graphemic Approximation of Phonological Context for English-Chinese Transliteration</i> Oi Yee Kwong .....	186
<i>Czech Named Entity Corpus and SVM-based Recognizer</i> Jana Kravalova and Zdenek Zabokrtsky .....	194
<i>Voted NER System using Appropriate Unlabeled Data</i> Asif Ekbal and Sivaji Bandyopadhyay .....	202