Named Entities Workshop: Shared Task on Transliteration 2009

(NEWS 2009)

Held at ACL-IJCNLP 2009

Singapore
7 August 2009

# Table of Contents

*Report of NEWS 2009 Machine Transliteration Shared Task*
Haizhou Li, A Kumaran, Vladimir Pervouchine and Min Zhang .......................... 1

*Whitepaper of NEWS 2009 Machine Transliteration Shared Task*
Haizhou Li, A Kumaran, Min Zhang and Vladimir Pervouchine .......................... 19

*Automata for Transliteration and Machine Translation*
Kevin Knight ......................................................... 27

*DirecTL: a Language Independent Approach to Transliteration*
Sittichai Jiampojamarn, Aditya Bhargava, Qing Dou, Kenneth Dwyer and Grzegorz Kondrak . . . 28

*Named Entity Transcription with Pair n-Gram Models*
Martin Jansche and Richard Sproat ........................................ 32

*Machine Transliteration using Target-Language Grapheme and Phoneme: Multi-engine Transliteration Approach*
Jong-Hoon Oh, Kiyotaka Uchimoto and Kentaro Torisawa ................................. 36

*A Language-Independent Transliteration Schema Using Character Aligned Models at NEWS 2009*
Praneeth Shishtla, Surya Ganesh V, Sethuramalingam Subramaniam and Vasudeva Varma ...... 40

*Experiences with English-Hindi, English-Tamil and English-Kannada Transliteration Tasks at NEWS 2009*
Manoj Kumar Chinnakotla and Om P. Damani ........................................ 44

*Testing and Performance Evaluation of Machine Transliteration System for Tamil Language*
Kommaluri Vijayanand .................................................. 48

*Transliteration by Bidirectional Statistical Machine Translation*
Andrew Finch and Eiichiro Sumita ........................................ 52

*Transliteration of Name Entity via Improved Statistical Translation on Character Sequences*
Yan Song, Chunyu Kit and Xiao Chen ........................................ 57

*Learning Multi Character Alignment Rules and Classification of Training Data for Transliteration*
Dipankar Bose and Sudeshna Sarkar ........................................ 61

*Fast Decoding and Easy Implementation: Transliteration as Sequential Labeling*
Eiji Aramaki and Takeshi Abekawa ........................................ 65

*NEWS 2009 Machine Transliteration Shared Task System Description: Transliteration with Letter-to-Phoneme Technology*
Colin Cherry and Hisami Suzuki ........................................ 69

*Combining a Two-step Conditional Random Field Model and a Joint Source Channel Model for Machine Transliteration*
Dong Yang, Paul Dixon, Yi-Cheng Pan, Tasuku Oonishi, Masanobu Nakamura and Sadaoki Furui 72
Phonological Context Approximation and Homophone Treatment for NEWS 2009 English-Chinese Transliteration Shared Task
Oi Yee Kwong ................................................................. 76

English to Hindi Machine Transliteration System at NEWS 2009
Amitava Das, Asif Ekbal, Tapabrata Mondal and Sivaji Bandyopadhyay ...................... 80

Improving Transliteration Accuracy Using Word-Origin Detection and Lexicon Lookup
Mitesh Khapra and Pushpak Bhattacharyya .................................................. 84

A Noisy Channel Model for Grapheme-based Machine Transliteration
Jia Yuxiang, Zhu Danqing and Yu Shiwen .................................................... 88

Substring-based Transliteration with Conditional Random Fields
Sravana Reddy and Sonjia Waxmonsksy ....................................................... 92

A Syllable-based Name Transliteration System
Xue Jiang, Le Sun and Dakun Zhang .......................................................... 96

Transliteration System Using Pair HMM with Weighted FSTs
Peter Nabende ......................................................................................... 100

Rejwanul Haque, Sandipan Dandapat, Ankit Kumar Srivastava, Sudip Kumar Naskar and Andy Way .......................................................... 104

A Hybrid Approach to English-Korean Name Transliteration
Gumwon Hong, Min-Jeong Kim, Do-Gil Lee and Hae-Chang Rim ......................... 108

Language Independent Transliteration System Using Phrase-based SMT Approach on Substrings
Sara Noeman ...................................................................................... 112

Combining MDL Transliteration Training with Discriminative Modeling
Dmitry Zelenko ..................................................................................... 116

ε-extension Hidden Markov Models and Weighted Transducers for Machine Transliteration
Balakrishnan Varadarajan and Delip Rao .................................................... 120

Modeling Machine Transliteration as a Phrase Based Statistical Machine Translation Problem
Taraka Rama and Karthik Gali .................................................................. 124

Maximum n-Gram HMM-based Name Transliteration: Experiment in NEWS 2009 on English-Chinese Corpus
Yilu Zhou ............................................................................................ 128

Name Transliteration with Bidirectional Perceptron Edit Models
Dayne Freitag and Zhiqiang Wang .......................................................... 132

Bridging Languages by SuperSense Entity Tagging
Dave Picca, Alfio Massimiliano Gliozzo and Simone Campora............................. 136

Chinese-English Organization Name Translation Based on Correlative Expansion
Feiliang Ren, Muhua Zhu, Huizhen Wang and Jingbo Zhu .................................. 143

Name Matching between Roman and Chinese Scripts: Machine Complements Human
Ken Samuel, Alan Rubenstein, Sherri Condon and Alex Yeh ............................. 152
Analysis and Robust Extraction of Changing Named Entities
Masatoshi Tsuchiya, Shoko Endo and Seiichi Nakagawa ................................. 161

Tag Confidence Measure for Semi-Automatically Updating Named Entity Recognition
Kuniko Saito and Kenji Imamura ................................................................. 168

A Hybrid Model for Urdu Hindi Transliteration
Abbas Malik, Laurent Besacier, Christian Boitet and Pushpak Bhattacharyya ........ 177

Graphemic Approximation of Phonological Context for English-Chinese Transliteration
Oi Yee Kwong .......................................................... 186

Czech Named Entity Corpus and SVM-based Recognizer
Jana Kravalova and Zdenek Zabokrtsky ......................................................... 194

Voted NER System using Appropriate Unlabeled Data
Asif Ekbal and Sivaji Bandyopadhyay ......................................................... 202