Report of NEWS 2011 Machine Transliteration Shared Task

Min Zhang[†], **Haizhou Li**[†], **A Kumaran**[‡] **and Ming Liu** [†]

[†]Institute for Infocomm Research, A*STAR, Singapore 138632
{mzhang,hli,mliu}@i2r.a-star.edu.sg

[‡]Multilingual Systems Research, Microsoft Research India
A.Kumaran@microsoft.com

Abstract

This report documents the Machine Transliteration Shared Task conducted as a part of the Named Entities Workshop (NEWS 2011), an IJCNLP 2011 workshop. The shared task features machine transliteration of proper names from English to 11 languages and from 3 languages to English. In total, 14 tasks are provided. 10 teams from 7 different countries participated in the evaluations. Finally, 73 standard and 4 non-standard runs are submitted, where diverse transliteration methodologies are explored and reported on the evaluation data. We report the results with 4 performance metrics. We believe that the shared task has successfully achieved its objective by providing a common benchmarking platform for the research community to evaluate the state-of-the-art technologies that benefit the future research and development.

1 Introduction

Names play a significant role in many Natural Language Processing (NLP) and Information Retrieval (IR) systems. They are important in Cross Lingual Information Retrieval (CLIR) and Machine Translation (MT) as the system performance has been shown to positively correlate with the correct conversion of names between the languages in several studies (Demner-Fushman and Oard, 2002; Mandl and Womser-Hacker, 2005; Hermjakob et al., 2008; Udupa et al., 2009). The traditional source for name equivalence, the bilingual dictionaries — whether handcrafted or statistical — offer only limited support because new names always emerge.

All of the above point to the critical need for robust Machine Transliteration technology and systems. Much research effort has been made to address the transliteration issue in the research community (Knight and Graehl, 1998; Meng et al., 2001; Li et al., 2004; Zelenko and Aone, 2006; Sproat et al., 2006; Sherif and Kondrak, 2007; Hermjakob et al., 2008; Al-Onaizan and Knight, 2002; Goldwasser and Roth, 2008; Goldberg and Elhadad, 2008; Klementiev and Roth, 2006; Oh and Choi, 2002; Virga and Khudanpur, 2003; Wan and Verspoor, 1998; Kang and Choi, 2000; Gao et al., 2004; Zelenko and Aone, 2006; Li et al., 2009b; Li et al., 2009a). These previous work fall into three categories, i.e., grapheme-based, phoneme-based and hybrid methods. Grapheme-based method (Li et al., 2004) treats transliteration as a direct orthographic mapping and only uses orthography-related features while phoneme-based method (Knight and Graehl, 1998) makes use of phonetic correspondence to generate the transliteration. Hybrid method refers to the combination of several different models or knowledge sources to support the transliteration generation.

The first machine transliteration shared task (Li et al., 2009b; Li et al., 2009a) was held in NEWS 2009 at ACL-IJCNLP 2009. It was the first time to provide common benchmarking data in diverse language pairs for evaluation of state-of-the-art techniques. While the focus of the 2009 shared task was on establishing the quality metrics and on baselining the transliteration quality based on those metrics, the 2010 shared task (Li et al., 2010a; Li et al., 2010b) expanded the scope of the transliteration generation task to about a dozen languages, and explored the quality depending on the direction of transliteration, between the languages. NEWS 2011 was a continued effort of NEWS 2010 and NEWS 2009.

The rest of the report is organised as follows. Section 2 outlines the machine transliteration task and the corpora used and Section 3 discusses the metrics chosen for evaluation, along with the ratio-

Proceedings of the 2011 Named Entities Workshop, IJCNLP 2011, pages 1–13,
Chiang Mai, Thailand, November 12, 2011.

nale for choosing them. Sections 4 and 5 present the participation in the shared task and the results with their analysis, respectively. Section 6 concludes the report.

2 Transliteration Shared Task

In this section, we outline the definition and the description of the shared task.

2.1 "Transliteration": A definition

There exists several terms that are used interchangeably in the contemporary research literature for the conversion of names between two languages, such as, transliteration, transcription, and sometimes Romanisation, especially if Latin scripts are used for target strings (Halpern, 2007).

Our aim is not only at capturing the name conversion process from a source to a target language, but also at its practical utility for downstream applications, such as CLIR and MT. Therefore, we adopted the same definition of transliteration as during the NEWS 2009 workshop (Li et al., 2009a) to narrow down "transliteration" to three specific requirements for the task, as follows: *"Transliteration is the conversion of a given name in the source language (a text string in the source writing system or orthography) to a name in the target language (another text string in the target writing system or orthography), such that the target language name is: (i) phonemically equivalent to the source name (ii) conforms to the phonology of the target language and (iii) matches the user intuition of the equivalent of the source language name in the target language, considering the culture and orthographic character usage in the target language."*

In NEWS 2011, we introduce three back-transliteration tasks. We define back-transliteration as a process of restoring transliterated words to their original languages. For example, NEWS 2011 offers the tasks to convert western names written in Chinese and Thai into their original English spellings, and romanized Japanese names into their original Kanji writings.

2.2 Shared Task Description

Following the tradition in NEWS 2010, the shared task at NEWS 2011 is specified as development of machine transliteration systems in one or more of the specified language pairs. Each language pair of the shared task consists of a source and a target language, implicitly specifying the transliteration direction. Training and development data in each of the language pairs have been made available to all registered participants for developing a transliteration system for that specific language pair using any approach that they find appropriate.

At the evaluation time, a standard hand-crafted test set consisting of between 500 and 3,000 source names (approximately 5-10% of the training data size) have been released, on which the participants are required to produce a ranked list of transliteration candidates in the target language for each source name. The system output is tested against a reference set (which may include multiple correct transliterations for some source names), and the performance of a system is captured in multiple metrics (defined in Section 3), each designed to capture a specific performance dimension.

For every language pair each participant is required to submit at least one run (designated as a "standard" run) that uses only the data provided by the NEWS workshop organisers in that language pair, and no other data or linguistic resources. This standard run ensures parity between systems and enables meaningful comparison of performance of various algorithmic approaches in a given language pair. Participants are allowed to submit more "standard" runs, up to 4 in total. If more than one "standard" runs is submitted, it is required to name one of them as a "primary" run, which is used to compare results across different systems. In addition, up to 4 "non-standard" runs could be submitted for every language pair using either data beyond that provided by the shared task organisers or linguistic resources in a specific language, or both. This essentially may enable any participant to demonstrate the limits of performance of their system in a given language pair.

The shared task timelines provide adequate time for development, testing (approximately 1 month after the release of the training data) and the final result submission (7 days after the release of the test data).

2.3 Shared Task Corpora

We considered two specific constraints in selecting languages for the shared task: language diversity and data availability. To make the shared task interesting and to attract wider participation, it is important to ensure a reasonable variety among

the languages in terms of linguistic diversity, orthography and geography. Clearly, the ability of procuring and distributing a reasonably large (approximately 10K paired names for training and testing together) hand-crafted corpora consisting primarily of paired names is critical for this process. At the end of the planning stage and after discussion with the data providers, we have chosen the set of 14 tasks shown in Table 1 (Li et al., 2004; Kumaran and Kellner, 2007; MSRI, 2009; CJKI, 2010).

NEWS 2011 leverages on the success of NEWS 2010 by utilizing the training and dev data of NEWS 2010 as the training data of NEWS 2011 and the test data of NEWS 2010 as the dev data of NEWS 2011. NEWS 2011 provides entirely new test data across all 14 tasks for evaluation. In addition to the 12 tasks inherited from NEWS 2010, NEWS 2011 is augmented with 2 new tasks with two new languages (Persian, Hebrew).

The names given in the training sets for Chinese, Japanese, Korean, Thai, Persian and Hebrew languages are Western names and their respective transliterations; the Japanese Name (in English) $\rightarrow$ Japanese Kanji data set consists only of native Japanese names; the Arabic data set consists only of native Arabic names. The Indic data set (Hindi, Tamil, Kannada, Bangla) consists of a mix of Indian and Western names.

For all of the tasks chosen, we have been able to procure paired names data between the source and the target scripts and were able to make them available to the participants. For some language pairs, such as English-Chinese and English-Thai, there are both transliteration and back-transliteration tasks. Most of the task are just one-way transliteration, although Indian data sets contained mixture of names of both Indian and Western origins. The language of origin of the names for each task is indicated in the first column of Table 1.

Finally, it should be noted here that the corpora procured and released for NEWS 2011 represent perhaps the most diverse and largest corpora to be used for any common transliteration tasks today.

3 Evaluation Metrics and Rationale

The participants have been asked to submit results of up to four standard and four non-standard runs. One standard run must be named as the primary submission and is used for the performance summary. Each run contains a ranked list of up to 10 candidate transliterations for each source name. The submitted results are compared to the ground truth (reference transliterations) using 4 evaluation metrics capturing different aspects of transliteration performance. The same as the NEWS 2010, we have dropped two MAP metrics used in NEWS 2009 because they don't offer additional information to MAP_{ref}. Since a name may have multiple correct transliterations, all these alternatives are treated equally in the evaluation, that is, any of these alternatives is considered as a correct transliteration, and all candidates matching any of the reference transliterations are accepted as correct ones.

The following notation is further assumed:

N	:	Total number of names (source words) in the test set
n_i	:	Number of reference transliterations for i-th name in the test set ($n_i \geq 1$)
$r_{i,j}$	:	j-th reference transliteration for i-th name in the test set
$c_{i,k}$	:	k-th candidate transliteration (system output) for i-th name in the test set ($1 \leq k \leq 10$)
K_i	:	Number of candidate transliterations produced by a transliteration system

3.1 Word Accuracy in Top-1 (ACC)

Also known as Word Error Rate, it measures correctness of the first transliteration candidate in the candidate list produced by a transliteration system. $ACC = 1$ means that all top candidates are correct transliterations i.e. they match one of the references, and $ACC = 0$ means that none of the top candidates are correct.

$$ACC = \frac{1}{N} \sum_{i=1}^{N} \left\{ \begin{array}{l} 1 \text{ if } \exists r_{i,j} \, : \, r_{i,j} = c_{i,1}; \\ 0 \text{ otherwise} \end{array} \right\} \tag{1}$$

3.2 Fuzziness in Top-1 (Mean F-score)

The mean F-score measures how different, on average, the top transliteration candidate is from its closest reference. F-score for each source word is a function of Precision and Recall and equals 1 when the top candidate matches one of the references, and 0 when there are no common characters between the candidate and any of the references.

Precision and Recall are calculated based on the length of the Longest Common Subsequence

Name origin	Source script	Target script	Data Owner	Data Size			Task ID
				Train	Dev	Test	
Western	English	Chinese	Institute for Infocomm Research	37K	2.8K	2K	EnCh
Western	Chinese	English	Institute for Infocomm Research	28K	2.7K	2K	ChEn
Western	English	Korean Hangul	CJK Institute	7K	1K	1K	EnKo
Western	English	Japanese Katakana	CJK Institute	26K	2K	3K	EnJa
Japanese	English	Japanese Kanji	CJK Institute	10K	2K	3K	JnJk
Arabic	Arabic	English	CJK Institute	27K	2.5K	2.5K	ArEn
Mixed	English	Hindi	Microsoft Research India	12K	1K	2K	EnHi
Mixed	English	Tamil	Microsoft Research India	10K	1K	2K	EnTa
Mixed	English	Kannada	Microsoft Research India	10K	1K	2K	EnKa
Mixed	English	Bangla	Microsoft Research India	13K	1K	2K	EnBa
Western	English	Thai	NECTEC	27K	2K	2K	EnTh
Western	Thai	English	NECTEC	25K	2K	2K	ThEn
Western	English	Persian	Sarvnaz Karimi/RMIT	10K	2K	1K	EnPe
Western	English	Hebrew	Microsoft Research India	9.5K	1K	2K	EnHe

Table 1: Source and target languages for the shared task on transliteration.

(LCS) between a candidate and a reference:

$$LCS(c,r) = \frac{1}{2}\left(|c| + |r| - ED(c,r)\right) \quad (2)$$

where ED is the edit distance and $|x|$ is the length of x. For example, the longest common subsequence between "abcd" and "afcde" is "acd" and its length is 3. The best matching reference, that is, the reference for which the edit distance has the minimum, is taken for calculation. If the best matching reference is given by

$$r_{i,m} = \arg\min_j \left(ED(c_{i,1}, r_{i,j})\right) \quad (3)$$

then Recall, Precision and F-score for i-th word are calculated as

$$R_i = \frac{LCS(c_{i,1}, r_{i,m})}{|r_{i,m}|} \quad (4)$$

$$P_i = \frac{LCS(c_{i,1}, r_{i,m})}{|c_{i,1}|} \quad (5)$$

$$F_i = 2\frac{R_i \times P_i}{R_i + P_i} \quad (6)$$

- The length is computed in distinct Unicode characters.
- No distinction is made on different character types of a language (e.g., vowel vs. consonants vs. combining diereses etc.)

3.3 Mean Reciprocal Rank (MRR)

Measures traditional MRR *for any right answer* produced by the system, from among the candidates. $1/MRR$ tells approximately the average rank of the correct transliteration. MRR closer to 1 implies that the correct answer is mostly produced close to the top of the n-best lists.

$$RR_i = \left\{ \begin{array}{l} \min_j \frac{1}{j} \text{ if } \exists r_{i,j}, c_{i,k} : r_{i,j} = c_{i,k}; \\ 0 \text{ otherwise} \end{array} \right\} \quad (7)$$

$$MRR = \frac{1}{N}\sum_{i=1}^{N} RR_i \quad (8)$$

3.4 MAP$_{ref}$

Measures tightly the precision in the n-best candidates for i-th source name, for which reference transliterations are available. If all of the references are produced, then the MAP is 1. Let's denote the number of correct candidates for the i-th source word in k-best list as $num(i,k)$. MAP$_{ref}$ is then given by

$$MAP_{ref} = \frac{1}{N}\sum_{i}^{N}\frac{1}{n_i}\left(\sum_{k=1}^{n_i} num(i,k)\right) \quad (9)$$

4 Participation in Shared Task

10 teams from 7 countries and regions (Canada, Hong Kong/Mainland China, Iran, Germany, USA, Japan, Thailand) submitted their transliteration results.

Two teams have participated in all or almost all tasks while others participated in 1 to 4 tasks. Each language pair has attracted on average around 4 teams. The details are shown in Table 3.

Teams are required to submit at least one standard run for every task they participated in. In total, we receive 73 standard and 4 non-standard runs. Table 2 shows the number of standard and non-standard runs submitted for each task. It is

clear that the most "popular" task is the transliteration from English to Chinese being attempted by 7 participants. The next most popular is back-transliteration from Chinese to English being attempted by 6 participants. This is somewhat different from NEWS 2010, where the two most popular tasks were English to Hindi and English to other Indic scripts (Tamil,Kannada,Bangla) and Thai transliteration.

5 Task Results and Analysis

5.1 Standard runs

All the results are presented numerically in Tables 4–17, for all evaluation metrics. These are the official evaluation results published for this edition of the transliteration shared task.

The methodologies used in the ten submitted system papers are summarized as follows. Finch et al. (2011) employ non-Parametric Bayesian method to co-segment bilingual named entities for model training and report very good performance. This system is based on phrase-based statistical machine transliteration (SMT) (Finch and Sumita, 2008), an approach initially developed for machine translation (Koehn et al., 2003), where the SMT system's log-linear model is augmented with a set of features specifically suited to the task of transliteration. In particular, the model utilizes a feature based on a joint source-channel model, and a feature based on a maximum entropy model that predicts target grapheme sequences using the local context of graphemes and grapheme sequences in both source and target languages.

Jiang et al. (2011) extensively explore the use of accessor variety (a similarity measure) of the source graphemes as a feature under CRF framework for machine transliteration and report promising results. Kruengkrai et al. (2011) study discriminative training based on the Margin Infused Relaxed Algorithm with simple character alignments under SMT framework for machine transliteration. They report very impressive results. Bhargava et al. (2011) attemp to improve transliteration performance by leveraging transliterations from multiple languages. Dasigi and Diab (2011) adopt the approach of phrase-based statistical machine transliteration (Finch and Sumita, 2008). Chen et al. (2011) extend the joint source-channel model (Li et al., 2004) on the transliteration task into a multi-to-multi joint source-channel model, which allows alignments between substrings of arbitrary lengths in both source and target strings. Qin and Chen (2011) adopt the approach of Conditional Random Fields (CRF) (Lafferty et al., 2001).

Kwong (2011) present their transliteration system with a syllable-based Backward Maximum Matching method. The system uses the Onset First Principle to syllabify English names and align them with Chinese names. The bilingual lexicon containing aligned segments of various syllable lengths subsequently allows direct transliteration by chunks. Wang and Tsai (2011) adopt the substring-based transliteration approach which groups the characters of named entity in both source and target languages into substrings and then formulate the transliteration as a sequential tagging problem to tag the substrings in the source language with the substrings in the target language. The CRF algorithm is then used to deal with this tagging problem. They also construct a rule-based transliteration method for comparison. Nejad et al. (2011) report three systems for transliteration: the first system is a maximum entropy model with a newly proposed alignment algorithm. The second system is Sequitur g2p tool, an open source grapheme to phoneme convertor. The third system is Moses, a phrased based statistical machine translation system. In addition, several new features are introduced to enhance the overall accuracy in the maximum entropy model. Their results show that the combination of maximum entropy system with Sequitur g2p tool and Moses lead to a considerable improvement over individual systems.

5.2 Non-standard runs

For the non-standard runs, we pose no restrictions on the use of data or other linguistic resources. The purpose of non-standard runs is to see how best personal name transliteration can be, for a given language pair. In NEWS 2011, the approaches used in non-standard runs are typical and may be summarised as follows:

- with supplemental transliteration data from other languages of NEWS 2011 data. (Bhargava et al., 2011). Significant performance improvement is reported with this additional knowledge.

- with English phonemic information from CMU Pronouncing Dictionary v0.7a1

	English to Chinese	Chinese to English	English to Thai	Thai to English	English to Hindi	English to Tamil	English to Kannada
Language pair code	EnCh	ChEn	EnTh	ThEn	EnHi	EnTa	EnKa
Standard runs	15	13	4	4	9	4	4
Non-standard runs	0	0	0	0	1	0	0
	English to Japanese Katakana	English to Korean Hangul	English to Japanese Kanji	Arabic to English	English to Bengali (Bangla)	English to Persian	English to Hebrew
Language pair code	EnJa	EnKo	JnJk	ArEn	EnBa	EnPe	EnHe
Standard runs	2	2	1	3	3	6	3
Non-standard runs	0	3	0	0	0	0	0

Table 2: Number of runs submitted for each task. Number of participants coincides with the number of standard runs submitted.

Team ID	Organisation	EnCh	ChEn	EnTh	ThEn	EnHi	EnTa	EnKa	EnJa	EnKo	JnJk	ArEn	EnBa	EnPe	EnHe
1	Amirkabir University of Technology													x	
2	NICT	x	x	x	x	x	x	x	x	x	x	x	x	x	x
3	Beijing Foreign Studies University	x	x												
4	DFKI GmbH	x	x												
5	City University of Hong Kong	x	x												
6	NECTEC	x	x	x	x	x	x	x					x	x	x
7	University of Alberta	x				x			x						
8	Yuan Ze University and National Taiwan University									x					
9	National Tsing Hua University	x	x												
10	Columbia University					x						x		x	

Table 3: Participation of teams in different tasks.

(http://www.speech.cs.cmu.edu/cgi-bin/cmudict) (Das et al., 2010). However, performance drops very much when using the English phonemic information.

6 Conclusions and Future Plans

The Machine Transliteration Shared Task in NEWS 2011 shows that the community has a continued interest in this area. This report summarizes the results of the shared task. Again, we are pleased to report a comprehensive calibration and baselining of machine transliteration approaches as most state-of-the-art machine transliteration techniques are represented in the shared task. In addition to the most popular techniques such as Phrase-Based Machine Transliteration (Koehn et al., 2003), system combination and re-ranking in the NEWS 2010, we are delighted to see that several new techniques have been proposed and explored with promising results reported, including Non-Parametric Bayesian Co-segmentation (Finch et al., 2011), Multi-to-Multi Joint Source Channel Model (Chen et al., 2011), Leveraging Transliterations from Multiple Languages (Bhargava et al., 2011) and discriminative training based on the Margin Infused Relaxed Algorithm (Kruengkrai et al., 2011) . As the standard runs are limited by the use of corpus, most of the systems are implemented under the direct orthographic mapping (DOM) framework (Li et al., 2004). While the standard runs allow us

to conduct meaningful comparison across different algorithms, we recognise that the non-standard runs open up more opportunities for exploiting a variety of additional linguistic corpora.

Encouraged by the success of the NEWS workshop series, we would like to continue this event in the future conference to promote the machine transliteration research and development.

Acknowledgements

The organisers of the NEWS 2011 Shared Task would like to thank the Institute for Infocomm Research (Singapore), Microsoft Research India, CJK Institute (Japan), National Electronics and Computer Technology Center (Thailand) and Sarvnaz Karim / RMIT for providing the corpora and technical support. Without those, the Shared Task would not be possible. We thank those participants who identified errors in the data and sent us the errata. We also want to thank the members of programme committee for their invaluable comments that improve the quality of the shared task papers. Finally, we wish to thank all the participants for their active participation that have made this first machine transliteration shared task a comprehensive one.

References

Yaser Al-Onaizan and Kevin Knight. 2002. Machine transliteration of names in arabic text. In *Proc. ACL-2002 Workshop: Computational Apporaches to Semitic Languages*, Philadelphia, PA, USA.

Aditya Bhargava, Bradley Hauer, and Grzegorz Kondrak. 2011. Leveraging transliterations from multiple languages. In *Proc. Named Entities Workshop at IJCNLP 2011*.

Yu Chen, Rui Wang, and Yi Zhang. 2011. Statistical machine transliteration with multi-to-multi joint source channel model. In *Proc. Named Entities Workshop at IJCNLP 2011*.

CJKI. 2010. CJK Institute. http://www.cjk.org/.

Amitava Das, Tanik Saikh, Tapabrata Mondal, Asif Ekbal, and Sivaji Bandyopadhyay. 2010. English to Indian languages machine transliteration system at NEWS 2010. In *Proc. Named Entities Workshop at ACL 2010*.

Pradeep Dasigi and Mona Diab. 2011. Named entity transliteration using a statistical machine translation framework. In *Proc. Named Entities Workshop at IJCNLP 2011*.

D. Demner-Fushman and D. W. Oard. 2002. The effect of bilingual term list size on dictionary-based cross-language information retrieval. In *Proc. 36-th Hawaii Int'l. Conf. System Sciences*, volume 4, page 108.2.

Andrew Finch and Eiichiro Sumita. 2008. Phrase-based machine transliteration. In *Proc. 3rd Int'l. Joint Conf NLP*, volume 1, Hyderabad, India, January.

Andrew Finch, Paul Dixon, and Eiichiro Sumita. 2011. Integrating models derived from non-parametric bayesian co-segmentation into a statistical machine transliteration system. In *Proc. Named Entities Workshop at IJCNLP 2011*.

Wei Gao, Kam-Fai Wong, and Wai Lam. 2004. Phoneme-based transliteration of foreign names for OOV problem. In *Proc. IJCNLP*, pages 374–381, Sanya, Hainan, China.

Yoav Goldberg and Michael Elhadad. 2008. Identification of transliterated foreign words in Hebrew script. In *Proc. CICLing*, volume LNCS 4919, pages 466–477.

Dan Goldwasser and Dan Roth. 2008. Transliteration as constrained optimization. In *Proc. EMNLP*, pages 353–362.

Jack Halpern. 2007. The challenges and pitfalls of Arabic romanization and arabization. In *Proc. Workshop on Comp. Approaches to Arabic Script-based Lang.*

Ulf Hermjakob, Kevin Knight, and Hal Daumé. 2008. Name translation in statistical machine translation: Learning when to transliterate. In *Proc. ACL*, Columbus, OH, USA, June.

Mike Tian-Jian Jiang, Chan-Hung Kuo, and Wen-Lian Hsu. 2011. English-to-chinese machine transliteration using accessor variety features of source graphemes. In *Proc. Named Entities Workshop at IJCNLP 2011*.

Byung-Ju Kang and Key-Sun Choi. 2000. English-Korean automatic transliteration/back-transliteration system and character alignment. In *Proc. ACL*, pages 17–18, Hong Kong.

Alexandre Klementiev and Dan Roth. 2006. Weakly supervised named entity transliteration and discovery from multilingual comparable corpora. In *Proc. 21st Int'l Conf Computational Linguistics and 44th Annual Meeting of ACL*, pages 817–824, Sydney, Australia, July.

Kevin Knight and Jonathan Graehl. 1998. Machine transliteration. *Computational Linguistics*, 24(4).

P. Koehn, F. J. Och, and D. Marcu. 2003. Statistical phrase-based translation. In *Proc. HLT-NAACL*.

Canasai Kruengkrai, Thatsanee Charoenporn, and Virach Sornlertlamvanich. 2011. Simple discriminative training for machine transliteration. In *Proc. Named Entities Workshop at IJCNLP 2011*.

A Kumaran and T. Kellner. 2007. A generic framework for machine transliteration. In *Proc. SIGIR*, pages 721–722.

Oi Yee Kwong. 2011. English-chinese personal name transliteration by syllable-based maximum matching. In *Proc. Named Entities Workshop at IJCNLP 2011*.

J. Lafferty, A. McCallum, and F. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proc. Int'l. Conf. Machine Learning*, pages 282–289.

Haizhou Li, Min Zhang, and Jian Su. 2004. A joint source-channel model for machine transliteration. In *Proc. 42nd ACL Annual Meeting*, pages 159–166, Barcelona, Spain.

Haizhou Li, A Kumaran, Vladimir Pervouchine, and Min Zhang. 2009a. Report of NEWS 2009 machine transliteration shared task. In *Proc. Named Entities Workshop at ACL 2009*.

Haizhou Li, A Kumaran, Min Zhang, and Vladimir Pervouchine. 2009b. ACL-IJCNLP 2009 Named Entities Workshop — Shared Task on Transliteration. In *Proc. Named Entities Workshop at ACL 2009*.

Haizhou Li, A Kumaran, Min Zhang, and Vladimir Pervouchine. 2010a. Report of news 2010 transliteration generation shared task. In *Proc. Named Entities Workshop at ACL 2010*.

Haizhou Li, A Kumaran, Min Zhang, and Vladimir Pervouchine. 2010b. Whitepaper of news 2010 shared task on transliteration generation. In *Proc. Named Entities Workshop at ACL 2010*.

T. Mandl and C. Womser-Hacker. 2005. The effect of named entities on effectiveness in cross-language information retrieval evaluation. In *Proc. ACM Symp. Applied Comp.*, pages 1059–1064.

Helen M. Meng, Wai-Kit Lo, Berlin Chen, and Karen Tang. 2001. Generate phonetic cognates to handle name entities in English-Chinese cross-language spoken document retrieval. In *Proc. ASRU*.

MSRI. 2009. Microsoft Research India. http://research.microsoft.com/india.

Najmeh Mousavi Nejad, Shahram Khadivi, and Kaveh Taghipour. 2011. The machine transliteration system description for news 2011. In *Proc. Named Entities Workshop at IJCNLP 2011*.

Jong-Hoon Oh and Key-Sun Choi. 2002. An English-Korean transliteration model using pronunciation and contextual rules. In *Proc. COLING 2002*, Taipei, Taiwan.

Ying Qin and GuoHua Chen. 2011. Forward-backward machine transliteration between english and chinese based on combined crfs. In *Proc. Named Entities Workshop at IJCNLP 2011*.

Tarek Sherif and Grzegorz Kondrak. 2007. Substring-based transliteration. In *Proc. 45th Annual Meeting of the ACL*, pages 944–951, Prague, Czech Republic, June.

Richard Sproat, Tao Tao, and ChengXiang Zhai. 2006. Named entity transliteration with comparable corpora. In *Proc. 21st Int'l Conf Computational Linguistics and 44th Annual Meeting of ACL*, pages 73–80, Sydney, Australia.

Raghavendra Udupa, K. Saravanan, Anton Bakalov, and Abhijit Bhole. 2009. "They are out there, if you know where to look": Mining transliterations of OOV query terms for cross-language information retrieval. In *LNCS: Advances in Information Retrieval*, volume 5478, pages 437–448. Springer Berlin / Heidelberg.

Paola Virga and Sanjeev Khudanpur. 2003. Transliteration of proper names in cross-lingual information retrieval. In *Proc. ACL MLNER*, Sapporo, Japan.

Stephen Wan and Cornelia Maria Verspoor. 1998. Automatic English-Chinese name transliteration for development of multilingual resources. In *Proc. COLING*, pages 1352–1356.

Yu-Chun Wang and Richard Tzong-Han Tsai. 2011. English-korean named entity transliteration using statistical substring-based and rule-based approaches. In *Proc. Named Entities Workshop at IJCNLP 2011*.

Dmitry Zelenko and Chinatsu Aone. 2006. Discriminative methods for transliteration. In *Proc. EMNLP*, pages 612–617, Sydney, Australia, July.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	Organisation
			Primary runs		
2	0.3485	0.700095	0.462495	0.341924	NICT
6	0.342	0.701729	0.40574	0.331184	NECTEC
7	0.3405	0.691719	0.4203	0.331469	University of Alberta
9	0.3265	0.688231	0.423711	0.318296	National Tsing Hua University
4	0.3195	0.673834	0.396812	0.308382	DFKI GmbH
3	0.308	0.666474	0.337148	0.305857	Beijing Foreign Studies University
5	0.3055	0.672302	0.377732	0.296502	City University of Hong Kong
			Non-primary standard runs		
6	0.328	0.695756	0.392008	0.318354	NECTEC
3	0.308	0.666474	0.337148	0.305857	Beijing Foreign Studies University
9	0.3035	0.675249	0.383354	0.293095	National Tsing Hua University
7	0.2875	0.661642	0.2875	0.27303	University of Alberta
5	0.2855	0.659605	0.349497	0.276169	City University of Hong Kong
4	0.26	0.638255	0.340081	0.250505	DFKI GmbH
9	0.2025	0.610451	0.282637	0.195431	National Tsing Hua University
9	0	0.124144	0.000063	0	National Tsing Hua University

Table 4: Runs submitted for English to Chinese task.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	Organisation
			Primary runs		
3	0.166814	0.764739	0.201932	0.166703	Beijing Foreign Studies University
5	0.154898	0.765737	0.215209	0.155119	City University of Hong Kong
2	0.144748	0.764534	0.242493	0.144417	NICT
4	0.132833	0.745695	0.210143	0.132723	DFKI GmbH
6	0.131068	0.729656	0.19266	0.131178	NECTEC
9	0.000883	0.014535	0.00248	0.000883	National Tsing Hua University
			Non-primary standard runs		
5	0.153575	0.756761	0.205823	0.153685	City University of Hong Kong
6	0.121359	0.726054	0.176186	0.121139	NECTEC
6	0.120035	0.713803	0.184312	0.119925	NECTEC
4	0.117387	0.730918	0.176915	0.117277	DFKI GmbH
6	0.113416	0.713676	0.169103	0.113305	NECTEC
3	0.097087	0.692511	0.127462	0.096867	Beijing Foreign Studies University
9	0	0.010269	0.000412	0	National Tsing Hua University

Table 5: Runs submitted for Chinese to English back-transliteration task.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	Organisation
			Primary runs		
6	0.3545	0.85371	0.450846	0.350021	NECTEC
2	0.338	0.85323	0.443537	0.335972	NICT
			Non-primary standard runs		
6	0.3545	0.857262	0.457232	0.350625	NECTEC
6	0.354	0.855659	0.456143	0.349931	NECTEC

Table 6: Runs submitted for English to Thai task.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	Organisation
			Primary runs		
2	0.29641	0.845061	0.427258	0.296617	NICT
6	0.28359	0.840587	0.401574	0.282973	NECTEC
			Non-primary standard runs		
6	0.282564	0.841174	0.400137	0.280754	NECTEC
6	0.280513	0.839531	0.397005	0.278251	NECTEC

Table 7: Runs submitted for Thai to English back-transliteration task.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	Organisation
			Primary runs		
2	0.478	0.879438	0.591206	0.4765	NICT
7	0.471	0.878619	0.571162	0.46975	University of Alberta
6	0.436	0.870378	0.53784	0.435	NECTEC
10	0.387	0.859914	0.51587	0.38675	Columbia University
			Non-primary standard runs		
7	0.493	0.883611	0.581677	0.492	University of Alberta
7	0.457	0.877803	0.551577	0.45475	University of Alberta
6	0.42	0.866161	0.518392	0.41875	NECTEC
6	0.417	0.867697	0.522927	0.41575	NECTEC
10	0.386	0.859778	0.515204	0.38575	Columbia University
			Non-standard runs		
7	0.521	0.896287	0.606057	0.5205	University of Alberta

Table 8: Runs submitted for English to Hindi task.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	Organisation
			Primary runs		
2	0.441	0.900489	0.577195	0.44	NICT
6	0.432	0.895693	0.55284	0.4305	NECTEC
			Non-primary standard runs		
6	0.42	0.890297	0.521162	0.4185	NECTEC
6	0.409	0.890383	0.511919	0.4075	NECTEC

Table 9: Runs submitted for English to Tamil task.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	Organisation
			Primary runs		
2	0.419	0.885498	0.539931	0.41725	NICT
6	0.398	0.877997	0.501557	0.396722	NECTEC
			Non-primary standard runs		
6	0.378	0.871573	0.469133	0.375861	NECTEC
6	0.371	0.869731	0.46439	0.368333	NECTEC

Table 10: Runs submitted for English to Kannada task.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	Organisation
			Primary runs		
7	0.434711	0.815425	0.434711	0.434435	University of Alberta
2	0.393939	0.802719	0.535614	0.393939	NICT

Table 11: Runs submitted for English to Japanese Katakana task.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	Organisation
			Primary runs		
8	0.430213	0.711027	0.430213	0.422824	Yuan Ze University and National Taiwan University
2	0.356322	0.68032	0.461892	0.352627	NICT
			Non-standard runs		
8	0.331691	0.653147	0.331691	0.325123	Yuan Ze University and National Taiwan University
8	0.331691	0.653147	0.466886	0.331691	Yuan Ze University and National Taiwan University
8	0.215107	0.474405	0.215107	0.208949	Yuan Ze University and National Taiwan University

Table 12: Runs submitted for English to Korean task.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	Organisation
			Primary runs		
2	0.45359	0.640551	0.568179	0.45359	NICT

Table 13: Runs submitted for English to Japanese Kanji back-transliteration task.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	Organisation
			Primary runs		
10	0.525502	0.928104	0.628327	0.386179	Columbia University
2	0.447063	0.910865	0.550146	0.351398	NICT
			Non-primary standard runs		
10	0.518547	0.926968	0.61153	0.382576	Columbia University

Table 14: Runs submitted for Arabic to English task.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	Organisation
			Primary runs		
2	0.478	0.89183	0.596738	0.4765	NICT
6	0.455	0.886901	0.556766	0.453	NECTEC
			Non-primary standard runs		
6	0.456	0.884593	0.554751	0.4545	NECTEC

Table 15: Runs submitted for English to Bengali (Bangla) task.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	Organisation
			Primary runs		
1	0.872	0.979153	0.912697	0.869435	Amirkabir University of Technology
6	0.6435	0.942838	0.744343	0.629047	NECTEC
2	0.6145	0.93794	0.741716	0.603994	NICT
10	0.6055	0.933434	0.696681	0.589026	Columbia University
			Non-primary standard runs		
6	0.642	0.943011	0.747032	0.626604	NECTEC
10	0.6045	0.933263	0.696521	0.588117	Columbia University

Table 16: Runs submitted for English to Persian task.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	Organisation
			Primary runs		
6	0.602	0.931385	0.701797	0.602	NECTEC
2	0.6	0.928666	0.715443	0.6	NICT
			Non-primary standard runs		
6	0.601	0.929689	0.697298	0.601	NECTEC

Table 17: Runs submitted for English to Hebrew task.