

Report of NEWS 2012 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 2012), an ACL 2012 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. 7 teams participated in the evaluations. Finally, 57 standard and 1 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 ad¹⁰

dress 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. In NEWS 2011 (Zhang et al., 2011a; Zhang et al., 2011b), we significantly increased the hand-crafted parallel named entities corpora to include 14 different language pairs from 11 language families, and made them available as the common dataset for the shared task. NEWS 2012 was a continued effort of NEWS 2011, 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 rationale 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.”*

Following NEWS 2011, in NEWS 2012, we still keep the three back-transliteration tasks. We define back-transliteration as a process of restoring transliterated words to their original languages. For example, NEWS 2012 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 of NEWS workshop series, the shared task at NEWS 2012 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 (more than 1 month after the release of the training data) and the final result submission (4 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 2012 leverages on the success of NEWS 2011 by utilizing the training set of NEWS 2011 as the training data of NEWS 2012 and the dev data of NEWS 2011 as the dev data of NEWS 2012. NEWS 2012 provides entirely new test data across all 14 tasks for evaluation.

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) → 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 2012 represent perhaps the most diverse and largest corpora to be used for any common transliteration tasks today.¹²

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 2011, 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.

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

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

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	1K	EnCh
Western	Chinese	English	Institute for Infocomm Research	28K	2.7K	2.2K	1K	ChEn
Western	English	Korean Hangul	CJK Institute	7K	1K	609	1K	EnKo
Western	English	Japanese Katakana	CJK Institute	26K	2K	1.8K	1K	EnJa
Japanese	English	Japanese Kanji	CJK Institute	10K	2K	571	1K	JnJk
Arabic	Arabic	English	CJK Institute	27K	2.5K	2.6K	1K	ArEn
Mixed	English	Hindi	Microsoft Research India	12K	1K	1K	1K	EnHi
Mixed	English	Tamil	Microsoft Research India	10K	1K	1K	1K	EnTa
Mixed	English	Kannada	Microsoft Research India	10K	1K	1K	1K	EnKa
Mixed	English	Bangla	Microsoft Research India	13K	1K	1K	1K	EnBa
Western	English	Thai	NECTEC	27K	2K	2K	1K	EnTh
Western	Thai	English	NECTEC	25K	2K	1.9K	1K	ThEn
Western	English	Persian	Sarvnaz Karimi / RMIT	10K	2K	2K	1K	EnPe
Western	English	Hebrew	Microsoft Research India	9.5K	1K	1K	1K	EnHe

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

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 (LCS) between a candidate and a reference:

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

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

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

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.)

13

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

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

4 Participation in Shared Task

7 teams submitted their transliteration results. Table 3 shows the details of registration tasks. Teams are required to submit at least one standard run for every task they participated in. In total, we receive 57 standard and 1 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.

	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	14	5	2	2	2	2	2
Non-standard runs	0	0	0	0	0	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	3	4	4	5	4	4	4
Non-standard runs	0	1	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	University of Alberta											x			
2	NICT	x	x	x	x	x	x	x	x	x	x	x	x	x	x
3	MIT@Lab of HIT		x												
4	IASL, Academia Sinica	x													
5	Yahoo Japan Corporation	x	x						x	x	x		x	x	x
6	Yuan Ze University									x					
7	CMU	x	x	x	x	x	x	x	x	x	x	x	x	x	x

Table 3: Participation of teams in different tasks.

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. Similar to their NEWS 2011 system, Finch et al. (2012) 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. Different from their NEWS 2011 system, in order to solve the data sparseness issue, they use two RNN-based LM to project the grapheme set onto a smaller hidden representation: one for the target grapheme sequence and the other for the sequence of grapheme sequence pair used to generate the target.

Zhang et al. (2012) also use the statistical phrase-based SMT framework. They propose the fine-grained English segmentation algorithm and other new features and achieve very good performance. Wu et al. (2012) uses m2m-aligner and DirecTL-p decoder and two re-ranking methods: co-occurrence at web corpus and JLIS-Reranking method based on the features from alignment results. They report very good performance at English-Korean tasks. Okuno (2012) studies the mpaligner (an improvement of m2m-aligner) and

shows that mpaligner is more effective than m2m-aligner. They also find that de-romanization is crucial to JnJk task and mora is the best alignment unit for EnJa task. Ammar et al. (2012) use CRF as the basic model but with two innovations: a training objective that optimizes toward any of a set of possible correct labels (i.e., multiple references) and a k-best reranking with non-local features. Their results on ArEn show that the two features are very effective in accuracy improvement. Kondrak et al. (2012) study the language-specific adaptations in the context of two language pairs: English to Chinese (Pinyin representation) and Arabic to English (letter mapping). They conclude that Pinyin representation is useful while letter mapping is less effective. Kuo et al. (2012) explore two-stage CRF for English-to-Chinese task and show that the two-stage CRF outperform traditional one-stage CRF.

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 2012, only one non-standard run (Wu et al., 2012) was submitted. Their reported web-based re-validation method is very effective.

6 Conclusions and Future Plans

The Machine Transliteration Shared Task in NEWS 2012 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), CRF, re-ranking, DirecTL-p decoder, Non-Parametric Bayesian Co-segmentation (Finch et al., 2011), and Multi-to-Multi Joint Source Channel Model (Chen et al., 2011) in the NEWS 2011, we are delighted to see that several new techniques have been proposed and explored with promising results reported, including RNN-based LM (Finch et al., 2012), English Segmentation algorithm (Zhang et al., 2012), JLI⁵

reranking method (Wu et al., 2012), improved m2m-aligner (Okuno, 2012), multiple reference-optimized CRF (Ammar et al., 2012), language dependent adaptation (Kondrak et al., 2012) and two-stage CRF (Kuo et al., 2012). 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 2012 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 Approaches to Semitic Languages*, Philadelphia, PA, USA.
- Waleed Ammar, Chris Dyer, and Noah Smith. 2012. Transliteration by sequence labeling with lattice encodings and reranking. In *Proc. Named Entities Workshop at ACL 2012*.
- 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/>.
- 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*.
- Andrew Finch, Paul Dixon, and Eiichiro Sumita. 2012. Rescoring a phrase-based machine transliteration system with recurrent neural network language models. In *Proc. Named Entities Workshop at ACL 2012*.
- 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.
- 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*.
- Grzegorz Kondrak, Xingkai Li, and Mohammad Salameh. 2012. Transliteration experiments on chinese and arabic. In *Proc. Named Entities Workshop at ACL 2012*.
- A Kumaran and T. Kellner. 2007. A generic framework for machine transliteration. In *Proc. SIGIR*, pages 721–722.
- Chan-Hung Kuo, Shih-Hung Liu, Mike Tian-Jian Jiang, Cheng-Wei Lee, and Wen-Lian Hsu. 2012. Cost-benefit analysis of two-stage conditional random fields based english-to-chinese machine transliteration. In *Proc. Named Entities Workshop at ACL 2012*.
- 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>.
- Jong-Hoon Oh and Key-Sun Choi. 2002. An English-Korean transliteration model using pronunciation and contextual rules. In *Proc. COLING 2002*, Taipei, Taiwan.
- Yoh Okuno. 2012. Applying mpaligner to machine transliteration with japanese-specific heuristics. In *Proc. Named Entities Workshop at ACL 2012*.
- 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.
- Chun-Kai Wu, Yu-Chun Wang, and Richard Tzong-Han Tsai. 2012. English-korean named entity transliteration using substring alignment and re-ranking methods. In *Proc. Named Entities Workshop at ACL 2012*.
- Dmitry Zelenko and Chinatsu Aone. 2006. Discriminative methods for transliteration. In *Proc. EMNLP*, pages 612–617, Sydney, Australia, July.
- Min Zhang, A Kumaran, and Haizhou Li. 2011a. Whitepaper of news 2011 shared task on machine transliteration. In *Proc. Named Entities Workshop at IJCNLP 2011*.
- Min Zhang, Haizhou Li, A Kumaran, and Ming Liu. 2011b. Report of news 2011 machine transliteration shared task. In *Proc. Named Entities Workshop at IJCNLP 2011*.
- Chunyue Zhang, Tingting Li, and Tiejun Zhao. 2012. Syllable-based machine transliteration with extra phrase features. In *Proc. Named Entities Workshop at ACL 2012*.

Team ID	ACC	F -score	MRR	MAP_{ref}	Organisation
Primary runs					
3	0.330357	0.66898	0.413062	0.320285	MIT@Lab of HIT
1	0.325397	0.67228	0.418079	0.316296	University of Alberta
2	0.310516	0.66585	0.44664	0.307788	NICT
4	0.310516	0.662467	0.37696	0.299266	IASL, Academia Sinica
5	0.300595	0.655091	0.376025	0.292252	Yahoo Japan Corporation
7	0.031746	0.430698	0.055574	0.030265	CMU
Non-primary standard runs					
3	0.330357	0.676232	0.407755	0.3191	MIT@Lab of HIT
1	0.325397	0.673053	0.409452	0.316055	University of Alberta
1	0.324405	0.668165	0.424517	0.316248	University of Alberta
3	0.31746	0.666551	0.399476	0.308187	MIT@Lab of HIT
4	0.298611	0.658836	0.362263	0.288725	IASL, Academia Sinica
5	0.298611	0.656974	0.357481	0.289373	Yahoo Japan Corporation
4	0.294643	0.651988	0.357495	0.284274	IASL, Academia Sinica
4	0.290675	0.653565	0.370733	0.282545	IASL, Academia Sinica

Table 4: Runs submitted for English to Chinese task.

Team ID	ACC	F -score	MRR	MAP_{ref}	Organisation
Primary runs					
2	0.20314	0.736058	0.308801	0.199569	NICT
3	0.176644	0.701791	0.257324	0.172991	MIT@Lab of HIT
7	0.030422	0.489705	0.048211	0.03004	CMU
5	0.012758	0.258962	0.017354	0.012758	Yahoo Japan Corporation
Non-primary standard runs					
5	0.007851	0.258013	0.012163	0.007851	Yahoo Japan Corporation

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

Team ID	ACC	F -score	MRR	MAP_{ref}	Organisation
Primary runs					
2	0.122168	0.746824	0.183318	0.122168	NICT
7	0.000809	0.288585	0.001883	0.000809	CMU

Table 6: Runs submitted for English to Thai task.

Team ID	ACC	F -score	MRR	MAP_{ref}	Organisation
Primary runs					
2	0.139968	0.765534	0.21551	0.139968	NICT
7	0	0.417451	0.000566	0	CMU

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

Team ID	ACC	F -score	MRR	MAP_{ref}	Organisation
Primary runs					
2	0.668	0.923347	0.73795	0.661278	NICT
7	0.048	0.645666	0.087842	0.048528	CMU

Table 8: Runs submitted for English to Hindi task.

Team ID	ACC	F -score	MRR	MAP_{ref}	Organisation
Primary runs					
2	0.592	0.908444	0.67881	0.5915	NICT
7	0.052	0.638029	0.083728	0.052	CMU

Table 9: Runs submitted for English to Tamil task.

Team ID	ACC	F -score	MRR	MAP_{ref}	Organisation
Primary runs					
2	0.546	0.900557	0.640534	0.545361	NICT
7	0.116	0.737857	0.180234	0.11625	CMU

Table 10: Runs submitted for English to Kannada task.

Team ID	ACC	F -score	MRR	MAP_{ref}	Organisation
Primary runs					
2	0.400774	0.810109	0.522758	0.397386	NICT
5	0.362052	0.802701	0.468973	0.35939	Yahoo Japan Corporation
7	0	0.147441	0.00038	0	CMU

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

Team ID	ACC	F -score	MRR	MAP_{ref}	Organisation
Primary runs					
6	0.398095	0.731212	0.398095	0.396905	Yuan Ze University
2	0.38381	0.721247	0.464553	0.383095	NICT
5	0.334286	0.687794	0.411264	0.334048	Yahoo Japan Corporation
7	0	0	0.00019	0	CMU
Non-standard runs					
6	0.458095	0.756755	0.484048	0.458095	Yuan Ze University

Table 12: Runs submitted for English to Korean task.

Team ID	ACC	F -score	MRR	MAP_{ref}	Organisation
Primary runs					
2	0.513242	0.693184	0.598304	0.418708	NICT
5	0.512329	0.693029	0.581803	0.400505	Yahoo Japan Corporation
7	0	0	0	0	CMU
Non-primary standard runs					
5	0.511416	0.691131	0.580485	0.402127	Yahoo Japan Corporation

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

Team ID	ACC	F -score	MRR	MAP_{ref}	Organisation
Primary runs					
2	0.588235	0.929787	0.709003	0.506991	NICT
7	0.58391	0.925292	0.694338	0.367162	CMU
1	0.583045	0.932959	0.670457	0.42041	University of Alberta
Non-primary standard runs					
7	0.57699	0.93025	0.678898	0.330353	CMU
7	0.573529	0.925306	0.675125	0.328782	CMU

Table 14: Runs submitted for Arabic to English task.

Team ID	ACC	F -score	MRR	MAP_{ref}	Organisation
Primary runs					
2	0.46	0.891476	0.582944	0.458417	NICT
5	0.404	0.882395	0.514541	0.402917	Yahoo Japan Corporation
7	0.178	0.783893	0.248674	0.177139	CMU
Non-primary standard runs					
5	0.398	0.880286	0.510148	0.396528	Yahoo Japan Corporation

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

Team ID	ACC	F -score	MRR	MAP_{ref}	Organisation
Primary runs					
5	0.658349	0.940642	0.761223	0.639873	Yahoo Japan Corporation
2	0.65547	0.941044	0.773843	0.642663	NICT
7	0.18618	0.803002	0.311881	0.184961	CMU
Non-primary standard runs					
5	0.054702	0.627335	0.082754	0.054367	Yahoo Japan Corporation

Table 16: Runs submitted for English to Persian task.

Team ID	ACC	F -score	MRR	MAP_{ref}	Organisation
Primary runs					
5	0.190909	0.808491	0.253575	0.19	Yahoo Japan Corporation
2	0.153636	0.787254	0.228649	0.152727	NICT
7	0.097273	0.759444	0.130955	0.096818	CMU
Non-primary standard runs					
5	0.165455	0.803019	0.241948	0.164545	Yahoo Japan Corporation

Table 17: Runs submitted for English to Hebrew task.