

# Report of NEWS 2010 Transliteration Generation Shared Task

Haizhou Li<sup>†</sup>, A Kumaran<sup>‡</sup>, Min Zhang<sup>†</sup> and Vladimir Pervouchine<sup>†</sup>

<sup>†</sup>Institute for Infocomm Research, A\*STAR, Singapore 138632  
{hli,mzhang,vpervouchine}@i2r.a-star.edu.sg

<sup>‡</sup>Multilingual Systems Research, Microsoft Research India  
A.Kumaran@microsoft.com

## Abstract

This report documents the Transliteration Generation Shared Task conducted as a part of the Named Entities Workshop (NEWS 2010), an ACL 2010 workshop. The shared task features machine transliteration of proper names from English to 9 languages and from 3 languages to English. In total, 12 tasks are provided. 7 teams from 5 different countries participated in the evaluations. Finally, 33 standard and 8 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 sys-

tems. 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. NEWS 2010 is a continued effort of NEWS 2009. It builds on the foundations established in the first transliteration shared task and extends the scope to include new language pairs.

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.”*

In NEWS 2010, we introduce three back-transliteration tasks. We define back-transliteration as a process of restoring transliterated words to their original languages. For example, NEWS 2010 offers the tasks to convert western names written in Chinese and Thai into their original English spellings, or romanized Japanese names into their original Kanji writings.

### 2.2 Shared Task Description

Following the tradition in NEWS 2009, the shared task at NEWS 2010 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 1,000 and 3,000 source names (approximately 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 every 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 12 tasks shown in Table 1 (Li et al., 2004; Kumaran and Kellner, 2007; MSRI, 2009; CJKI, 2010).

NEWS 2010 leverages on the success of NEWS 2009 by utilizing the training and dev data of NEWS 2009 as the training data of NEWS 2010 and the test data of NEWS 2009 as the dev data of NEWS 2010. NEWS 2010 provides totally new test data across all 12 tasks for evaluation. In addition to the 7 tasks inherited from NEWS 2009, NEWS 2010 is enhanced with 5 new tasks, three new languages (Arabic, Bangla and Thai) and two back-transliteration (Chinese to English and Thai to English).

The names given in the training sets for Chinese, Japanese, Korean and Thai 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 2010 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. 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 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:

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

Name origin	Source script	Target script	Data Owner	Data Size			Task ID
				Train	Dev	Test	
Western	English	Chinese	Institute for Infocomm Research	32K	6K	2K	EnCh
Western	Chinese	English	Institute for Infocomm Research	25K	5K	2K	ChEn
Western	English	Korean Hangul	CJK Institute	5K	2K	2K	EnKo
Western	English	Japanese Katakana	CJK Institute	23K	3K	3K	EnJa
Japanese	English	Japanese Kanji	CJK Institute	7K	3K	3K	JnJk
Arabic	Arabic	English	CJK Institute	25K	2.5K	2.5K	ArAe
Mixed	English	Hindi	Microsoft Research India	10K	2K	2K	EnHi
Mixed	English	Tamil	Microsoft Research India	8K	2K	2K	EnTa
Mixed	English	Kannada	Microsoft Research India	8K	2K	2K	EnKa
Mixed	English	Bangla	Microsoft Research India	10K	2K	2K	EnBa
Western	English	Thai	NECTEC	26K	2K	2K	EnTh
Western	Thai	English	NECTEC	24K	2K	2K	ThEn

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

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.)

### 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.

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

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

### 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 from 5 countries and regions (Canada, Hong Kong, India, 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 33 standard and 8. Table 2 shows the number of standard and non-standard runs submitted for each task. It is clear that the most “popular” task is transliteration from English to Hindi attempted by 5 participants. The next most popular are other Indic scripts (Tamil, Kannada, Bangla) and Thai, attempted by 3 participants. This is somewhat different from NEWS 2009, where the two most popular tasks were English to Hindi and English to Chinese transliteration.

	English to Chinese	Chinese to English	English to Thai	Thai to English	English to Hindi	English to Tamil
Language pair code	EnCh	ChEn	EnTh	ThEn	EnHi	EnTa
Standard runs	5	2	2	2	7	3
Non-standard runs	0	0	1	1	2	1

---

	English to Kannada	English to Japanese Katakana	English to Korean Hangul	English to Japanese Kanji	Arabic to English	English to Bengali (Bangla)
Language pair code	EnKa	EnJa	EnKo	JnJk	ArAe	EnBa
Standard runs	3	2	1	1	2	3
Non-standard runs	1	0	0	0	0	2

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	ArAe	EnBa
1*	IIT, Bombay					x							
2	University of Alberta	x	x	x	x	x	x	x	x	x	x	x	x
3						x							
4	City University of Hong Kong	x	x										
5	NICT			x	x	x	x	x	x			x	x
6				x	x								
7	Jadavpur University					x	x	x					x

Table 3: Participation of teams in different tasks. \*Participation without a system paper.

## 5 Task Results and Analysis

### 5.1 Standard runs

All the results are presented numerically in Tables 4–15, for all evaluation metrics. These are the official evaluation results published for this edition of the transliteration shared task.

Among the four submitted system papers<sup>1</sup>, Song et al. (2010) and Finch and Sumita (2010) adopt the approach of phrase-based statistical machine transliteration (Finch and Sumita, 2008), an approach initially developed for machine translation (Koehn et al., 2003) while Das et al. (2010) adopts the approach of Conditional Random Fields (CRF) (Lafferty et al., 2001). Jiampojamarn et al. (2010) further develop DirectTL approach presented at the previous NEWS workshop (Jiampojamarn et al., 2009), achieving very good performance in the NEWS 2010.

An example of a completely language-

<sup>1</sup>To maintain anonymity, papers of the teams that submitted anonymous results are not cited in this report.

independent approach is (Finch and Sumita, 2010). Other participants used language-independent approach but added language-specific pre- or post-processing (Jiampojamarn et al., 2010; Das et al., 2010; Song et al., 2010), including name origin recognition for English to Hindi task (Jiampojamarn et al., 2010).

Combination of different models via re-ranking of their outputs has been used in most of the systems (Das et al., 2010; Song et al., 2010; Finch and Sumita, 2010). In fact, one system (Song et al., 2010) is mostly devoted to re-ranking of the system output to achieve significant improvement of the *ACC* (accuracy in top-1) results compared to the same system in NEWS 2009 workshop (Song, 2009).

Compared the same seven tasks among the NEWS 2009 and the NEWS 2010 (almost same training sets, but different test sets), we can see that the performance in the NEWS 2010 drops except the English to Korean task. This could be due to the fact that NEWS 2010 introduces a entirely

new test set, which come from different sources than the train and dev sets, while NEWS 2009 have all train, dev and test sets from the same sources.

As far as back-transliteration is concerned, we can see that English-to-Thai and Thai-to-English have the similar performance. However, Chinese-to-English back transliteration performs much worse than English-to-Chinese forward transliteration. This could be due to the fact that Thai and English are alphabet languages in nature while Chinese is not. As a result, Chinese have much fewer transliteration units than English and Thai. In other words, Chinese to English transliteration is a one-to-many mapping while English-to-Chinese is a many-to-one mapping. The later one has fewer mapping ambiguities.

## 5.2 Non-standard runs

For the non-standard runs there exist 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 2010, the approaches used in non-standard runs are typical and may be summarised as follows:

- Pronunciation dictionaries to convert words to their phonetic transcription (Jiampojamarn et al., 2010).
- Web search. First, transliteration candidates are generated. A Web search is then performed to re-affirm or re-rank the candidacy (Das et al., 2010).

Unfortunately, these additional knowledge used in the non-standard runs is not helpful since all non-standard runs perform worse than their corresponding standard runs. This would be an interesting issue to look into.

## 6 Conclusions and Future Plans

The Transliteration Generation Shared Task in NEWS 2010 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. The most popular techniques such as Phrase-Based Machine Transliteration (Koehn

et al., 2003), system combination and re-ranking, are inspired by recent progress in statistical machine translation. 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 larger linguistic corpora. It is also noted that two systems have reported significant performance improvement over their NEWS 2009 systems.

NEWS 2010 Shared Task represents a successful debut of a community effort in driving machine transliteration techniques forward. 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 2010 Shared Task would like to thank the Institute for Infocomm Research (Singapore), Microsoft Research India, CJK Institute (Japan) and National Electronics and Computer Technology Center (Thailand) 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.
- 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. ACL Named Entities Workshop Shared Task*.
- 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 and Eiichiro Sumita. 2010. Transliteration using a phrase-based statistical machine translation system to re-score the output of a joint multigram model. In *Proc. ACL Named Entities Workshop Shared Task*.
- 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.
- Sittichai Jiampojarn, Aditya Bhargava, Qing Dou, Kenneth Dwyer, and Grzegorz Kondrak. 2009. DirecTL: a language-independent approach to transliteration. In *Proc. ACL/IJCNLP Named Entities Workshop Shared Task*.
- Sittichai Jiampojarn, Kenneth Dwyer, Shane Bergsma, Aditya Bhargava, Qing Dou, Mi-Young Kim, and Grzegorz Kondrak. 2010. Transliteration generation and mining with limited training resources. In *Proc. ACL Named Entities Workshop Shared Task*.
- 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*.
- A Kumaran and T. Kellner. 2007. A generic framework for machine transliteration. In *Proc. SIGIR*, pages 721–722.
- 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*.
- 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.
- Tarek Sherif and Grzegorz Kondrak. 2007. Substring-based transliteration. In *Proc. 45th Annual Meeting of the ACL*, pages 944–951, Prague, Czech Republic, June.

- Yan Song, Chunyu Kit, and Hai Zhao. 2010. Reranking with multiple features for better transliteration. In *Proc. ACL Named Entities Workshop Shared Task*.
- Yan Song. 2009. Name entities transliteration via improved statistical translation on character-level chunks. In *Proc. ACL/IJCNLP Named Entities Workshop Shared Task*.
- 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.
- 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					
4	0.477333	0.740494	0.506209	0.455491	City University of Hong Kong
2	0.363333	0.707435	0.430168	0.347701	University of Alberta
Non-primary standard runs					
2	0.362667	0.704284	0.428854	0.347500	University of Alberta
2	0.360333	0.706765	0.428990	0.345215	University of Alberta
2	0.357000	0.702902	0.419415	0.341567	University of Alberta

Table 4: Runs submitted for English to Chinese task.

Team ID	ACC	$F$ -score	MRR	$MAP_{ref}$	Organisation
Primary runs					
4	0.226766	0.749237	0.268557	0.226090	City University of Hong Kong
2	0.137209	0.740364	0.197665	0.136702	University of Alberta

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

Team ID	ACC	$F$ -score	MRR	$MAP_{ref}$	Organisation
Primary runs					
5	0.391000	0.872526	0.505264	0.391000	NICT
2	0.377500	0.866254	0.467328	0.377500	University of Alberta
Non-standard runs					
6	0.247000	0.842063	0.366959	0.247000	

Table 6: Runs submitted for English to Thai task.

Team ID	ACC	$F$ -score	MRR	$MAP_{ref}$	Organisation
Primary runs					
5	0.396690	0.872642	0.524511	0.396690	NICT
2	0.352056	0.861207	0.450472	0.352056	University of Alberta
Non-standard runs					
6	0.092778	0.706995	0.131779	0.092778	

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

Team ID	ACC	$F$ -score	MRR	$MAP_{ref}$	Organisation
Primary runs					
2	0.456456	0.884199	0.559212	0.456456	University of Alberta
5	0.445445	0.883841	0.574195	0.445445	NICT
3	0.381381	0.860320	0.403172	0.381381	
1	0.158158	0.810309	0.231594	0.158158	IIT, Bombay
7	0.150150	0.714490	0.307674	0.150150	Jadavpur University
Non-primary standard runs					
2	0.456456	0.885122	0.558203	0.456456	University of Alberta
1	0.142142	0.799092	0.205945	0.142142	IIT, Bombay
Non-standard runs					
7	0.254254	0.751766	0.369072	0.254254	Jadavpur University
7	0.170170	0.738777	0.314335	0.170170	Jadavpur University

Table 8: Runs submitted for English to Hindi task.

Team ID	ACC	$F$ -score	MRR	$MAP_{ref}$	Organisation
Primary runs					
2	0.390000	0.890692	0.515298	0.390000	University of Alberta
5	0.390000	0.886560	0.522088	0.390000	NICT
7	0.013000	0.562917	0.121233	0.013000	Jadavpur University
Non-standard runs					
7	0.082000	0.759856	0.142317	0.082000	Jadavpur University

Table 9: Runs submitted for English to Tamil task.

Team ID	ACC	$F$ -score	MRR	$MAP_{ref}$	Organisation
Primary runs					
5	0.371000	0.871131	0.506010	0.371000	NICT
2	0.341000	0.867133	0.460189	0.341000	University of Alberta
7	0.056000	0.663196	0.111500	0.056000	Jadavpur University
Non-standard runs					
7	0.055000	0.662106	0.168750	0.055000	Jadavpur University

Table 10: Runs submitted for English to Kannada task.

Team ID	ACC	$F$ -score	MRR	$MAP_{ref}$	Organisation
Primary runs					
2	0.397933	0.791233	0.507828	0.398062	University of Alberta
5	0.378295	0.782682	0.510096	0.377778	NICT

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

Team ID	ACC	$F$ -score	MRR	$MAP_{ref}$	Organisation
Primary runs					
2	0.553604	0.770168	0.672665	0.553835	University of Alberta

Table 12: Runs submitted for English to Korean task.

Team ID	ACC	$F$ -score	MRR	$MAP_{ref}$	Organisation
Primary runs					
2	0.125937	0.426349	0.201497	0.127339	University of Alberta

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

Team ID	ACC	$F$ -score	MRR	$MAP_{ref}$	Organisation
Primary runs					
2	0.463679	0.923826	0.535097	0.265379	University of Alberta
5	0.403014	0.891443	0.512337	0.327418	NICT

Table 14: Runs submitted for Arabic to English task.

Team ID	ACC	$F$ -score	MRR	$MAP_{ref}$	Organisation
Primary runs					
5	0.411705	0.882858	0.549913	0.411705	NICT
2	0.394551	0.876947	0.511876	0.394551	University of Alberta
7	0.232089	0.818470	0.325345	0.232089	Jadavpur University
Non-standard runs					
7	0.429869	0.875349	0.526152	0.429869	Jadavpur University
7	0.369324	0.845273	0.450589	0.369324	Jadavpur University

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