Report of NEWS 2009 Machine Transliteration Shared Task

Haizhou Li[†], A Kumaran[‡], Vladimir Pervouchine[†] and Min Zhang[†]

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

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

Abstract

This report documents the details of the Machine Transliteration Shared Task conducted as a part of the Named Entities Workshop (NEWS), an ACL-IJCNLP 2009 workshop. The shared task features machine transliteration of proper names from English to a set of languages. This shared task has witnessed enthusiastic participation of 31 teams from all over the world, with diversity of participation for a given system and wide coverage for a given language pair (more than a dozen participants per language pair). Diverse transliteration methodologies are represented adequately in the shared task for a given language pair, thus underscoring the fact that the workshop may truly indicate the state of the art in machine transliteration in these language pairs. We measure and report 6 performance metrics on the submitted results. We believe that the shared task has successfully achieved the following objectives: (i) bringing together the community of researchers in the area of Machine Transliteration to focus on various research avenues, (ii) Calibrating systems on common corpora, using common metrics, thus creating a reasonable baseline for the state-of-the-art of transliteration systems, and (iii) providing a quantitative basis for meaningful comparison and analysis between various algorithmic approaches used in machine transliteration. We believe that the results of this shared task would uncover a host of interesting research problems, giving impetus to research in this significant research area.

1 Introduction

Names play a significant role in many Natural Language Processing (NLP) and Information Retrieval (IR) systems. They have a critical role in Cross Language Information Retrieval (CLIR) and Machine Translation (MT) systems as the systems' performances are 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 as they do not have sufficient coverage of names. New names are introduced to the vocabulary of a language every day.

All of the above point to the critical need for robust Machine Transliteration technology and systems. This has attracted attention from the research community. Over the last decade scores of papers on Machine Transliteration have appeared in the top Computational Linguistics, Information Retrieval and Data Management conferences, exploring diverse algorithmic approaches in a wide variety of different languages (Knight and Graehl, 1998; Li et al., 2004; Zelenko and Aone, 2006; Sproat et al., 2006; Sherif and Kondrak, 2007; Hermjakob et al., 2008; Goldwasser and Roth, 2008; Goldberg and Elhadad, 2008; Klementiev and Roth, 2006). However, there has not been any coordinated effort in calibrating the state-of-the-art technical capabilities of machine transliteration: the studies explore different algorithmic approaches in different language pairs and report their performance in different metrics and tested on different corpora.

The overarching objective of this shared task is to drive the machine transliteration technology forward, to measure and baseline the state-of-the-

Proceedings of the 2009 Named Entities Workshop, ACL-IJCNLP 2009, pages 1–18,
Suntec, Singapore, 7 August 2009.

art and to provide a meaningful comparison between the most promising algorithmic approaches in order to stimulate the discussions among the researchers. The NLP community in Asia is especially interested in transliteration as several major Asian languages do not use Latin script in their native writing systems. The Named Entity Workshop (NEWS 2009) in ACL-IJCNLP 2009 in Singapore provides an ideal platform for the shared task to take off. This is precisely what we address in this shared task on machine transliteration that is conducted as a part of the Named Entity Workshop (NEWS-2009), an ACL-IJCNLP 2009 workshop.

The shared task aims at achieving the following objectives:

- Providing a forum to bring together the community of researchers in the area of Machine Transliteration to focus on various research avenues in this important research area.

- Calibrating systems on common hand-crafted corpora, using common metrics, in many different languages, thus creating a reasonable baseline for the state-of-the-art of transliteration systems.

- Analysing the results so that a reasonable comparison of different algorithmic approaches and their trade-offs (such as, transliteration quality vs. generality of approach across languages vs. training data size, etc.) may be explored.

We believe that a substantial part of what we have set out to achieve has been accomplished, and we present this report as a record of the task process, system participation and results and our findings. It is our hope that this reporting will generate lively discussions during the NEWS workshop and subsequent research in this important area.

This introduction outlines the purpose of the transliteration shared task conducted as a part of the NEWS workshop. 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. Section 4 sketches the participation. Section 5 presents the results of the shared task and the analysis of the results. Section 6, summarises the queries and feedback we have received from the participants and Section 7 concludes, presenting some lessons learnt from the current edition of the shared task, and some ideas we want to pursue in the future plan for the Machine Transliteration tasks.

2 Transliteration Shared Task

In this section, we outline the definition of the task, the process followed and the rationale for the decisions.

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 ultimate utility for downstream applications, such as CLIR and MT. We have narrowed down 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."*

Given that the phoneme set of languages may not be exactly the same, the first requirement must be diluted to "close to", instead of "equivalent". The second requirement is needed to ensure that the target string is a valid string as per the target language phonology. The third requirement is introduced to produce what a normal user would expect (at least for the popular names), and in order to make it useful for downstream applications like MT or CLIR systems. Though the third requirement make systems produce target language strings that marginally violate the first or second requirements, it ensures that such transliteration system is of value to downstream systems. All the above requirements are implicitly enforced by the choice of name pairs used to define the training and test corpora in a given language pair. In cases where multiple equivalent target language names are possible for a source language name, we include all of them.

After much debate, we have also retained the task name as "transliteration", though our definition may be closest to the "popular transcription" (Halpern, 2007), due to the popularity of term "Machine Transliteration" among the language technology researchers.

2.2 Shared Task Description

The shared task 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 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 runs (designated as "non-standard") 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 2 months after the release of the training data) and the final result submission (5 days after the release of the test data).

2.3 Shared Task Corpora

We have had 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 7 languages shown in Table 1 for the task (Li et al., 2004; Kumaran and Kellner, 2007; MSRI, 2009; CJKI, 2009).

For all of the languages chosen, we have been able to procure paired names data between English and the respective languages and were able to make them available to the participants. In addition, we have been able to procure a specific corpus of about 40K Romanised Japanese names and their Kanji counterparts, and the corresponding language pair (Japanese names from their Romanised form to Kanji) has been included as one of the task language pair.

It should be noted here that each corpus has a definite skew in its characteristics: the names in the Chinese, Japanese and Korean (CJK) language corpora are Western names; the Indic languages (Hindi, Kannada and Tamil) corpora consists of a mix of Indian and Western names. The Romanised Kanji to Kanji corpus consists only of native Japanese names. While such characteristics may have provided us an opportunity to specifically measure the performance for forward transliterations (in CJK) and backward transliterations (in Romanised Kanji), we do not highlight such fine distinctions in this edition.

Finally, it should be noted here that the corpora procured and released for NEWS 2009 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 one standard and up to four non-standard

Source language	Target language	Data Source	Data Size (No. source names) Training	Development	Testing	Task ID
English	Hindi	Microsoft Research India	9,975	974	1,000	EnHi
English	Tamil	Microsoft Research India	7,974	987	1,000	EnTa
English	Kannada	Microsoft Research India	7,990	968	1,000	EnKa
English	Russian	Microsoft Research India	5,977	943	1,000	EnRu
English	Chinese	Institute for Infocomm Research	31,961	2,896	2,896	EnCh
English	Korean Hangul	CJK Institute	4,785	987	989	EnKo
English	Japanese Katakana	CJK Institute	23,225	1,492	1,489	EnJa
Japanese name (in English)	Japanese Kanji	CJK Institute	6,785	1,500	1,500	JnJk

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

runs. 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 6 evaluation metrics capturing different aspects of transliteration performance. 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 (LCS) between a candidate and a reference:

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

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

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

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

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

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

$$F_i = 2\frac{R_i \times P_i}{R_i + P_i} \tag{6}$$

- 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

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

3.5 MAP$_{10}$

MAP$_{10}$ measures the precision in the 10-best candidates for i-th source name provided by the candidate system. In general, the higher MAP$_{10}$ is, the better is the quality of the transliteration system in capturing the multiple references.

$$MAP_{10} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{10} \left(\sum_{k=1}^{10} num(i,k) \right) \quad (10)$$

3.6 MAP$_{sys}$

MAP$_{sys}$ measures the precision in the top K_i-best candidates produced by the system for i-th source name, for which n_i reference transliterations are available. This measure allows the systems to produce variable number of transliterations, based on their confidence in identifying and producing correct transliterations.

$$MAP_{sys} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{K_i} \left(\sum_{k=1}^{K_i} num(i,k) \right) \quad (11)$$

4 Participation in Shared Task

There have been 31 systems from around the world that participated in the shared task and submitted the transliteration results for a common test data, produced by their systems trained on the common training corpora.

A few teams have participated in all or almost all tasks (that is, language pairs); most others participated in 3 tasks on average. Each language pair has attracted on average around 13 teams. The participation details are shown in Table 3 and the demographics of the participating teams by country is shown in Figure 1.

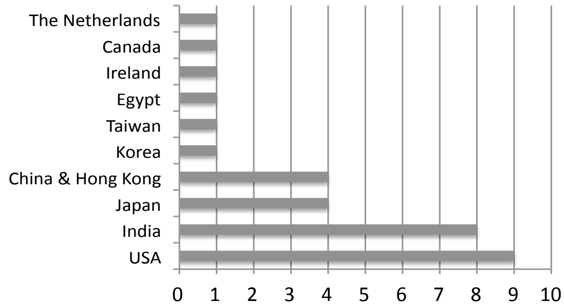


Figure 1: Participation by country.

Teams are required to submit at least one standard run for every task they participated in. In total 104 standard and 86 non-standard runs have been submitted. Table 2 shows the number of standard and non-standard runs submitted for each task. It is clear that the most "popular" tasks are transliteration from English to Hindi and from English to Chinese, attempted by 21 and 18 participants respectively. Overall, as can be noted from the results, each task has received significant participation.

5 Task Results and Analysis

5.1 Standard runs

The 8 individual plots in Figure 2 summarise (for each task) the results of standard runs via 3 measured metrics concerning output of at least one correct candidate per source word, namely, accuracy in top-1, F-score and Mean Reciprocal Rank (MRR). The plots in Figure 3 summarise (for each task) the results for 3 metrics on ranked ordered transliteration output of the systems, namely MAP$_{ref}$, MAP$_{10}$ and MAP$_{sys}$ metrics. All the results are presented numerically in Tables 8–11, for all evaluation metrics. These are the official

	English to Hindi	English to Tamil	English to Kannada	English to Russian	English to Chinese	English to Korean	English to Japanese Katakana	Japanese transliterated to Japanese Kanji
Language pair code	EnHi	EnTa	EnKa	EnRu	EnCh	EnKo	EnJa	JnJk
Standard runs	21	13	14	13	18	8	10	7
Non-standard runs	18	5	5	16	20	9	5	8

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

evaluation results published for this edition of the transliteration shared task. Note that two teams have updated their results (after fixing bugs in their systems) after the deadline; their results are identified specifically.

We find that two approaches to transliteration are most popular in the shared task submissions. One of these approaches is Phrase-based statistical machine transliteration (Finch and Sumita, 2008), an approach initially developed for machine translation (Koehn et al., 2003). Systems that adopted this approach are (Song, 2009; Haque et al., 2009; Noeman, 2009; Rama and Gali, 2009; Chinnakotla and Damani, 2009).[1] The other is Conditional Random Fields(Lafferty et al., 2001) (CRF), adopted by (Aramaki and Abekawa, 2009; Shishtla et al., 2009). With only a few exceptions, most implementations are based on approaches that are language-independent. Indeed, many of the participants fielded their systems on multiple languages, as can be seen from Table 3.

We also note that combination of several different models via re-ranking of their outputs (CRF, Maximum Entropy Model, Margin Infused Relaxed Algorithm) proves to be very successful (Oh et al., 2009); their system (reported as Team ID 6) produced the best or second-best transliteration performance consistently across all metrics, in all tasks, except Japanese back-transliteration. Examples of other model combinations are (Das et al., 2009).

At least two teams (reported as Team IDs 14 and 27) incorporate language origin detection in their system (Bose and Sarkar, 2009; Khapra and Bhattacharyya, 2009). The Indian language corpora contains names of both English and Indic origin. Khapra and Bhattacharyya (2009) demonstrate how much the transliteration performance can be improved when language of origin detection is employed, followed by a language-specific transliteration model for decoding.

Some systems merit specific mention as they adopt are rather unique approaches. Jiampojamarn et al. (2009) propose DirectTL discriminative sequence prediction model that is language-independent (reported as Team ID 7). Their transliteration accuracy is among the highest in several tasks (EnCh, EnHi and EnRu). Zelenko (2009) present an approach to the transliteration problem based on Minimum Description Length (MDL) principle. Freitag and Wang (2009) approach the problem of transliteration with bidirectional perceptron edit models.

Finally, in Figure 4 we present a plot where each point represents a standard run by a system, with different tasks marked with specific shape and colour. This plot gives a bird-eye-view of the system performances across two most uncorrelated evaluation metrics, namely accuracy in top-1 (ACC) and Mean F-score. Not surprisingly, we notice very high performance in terms of F-score for English to Russian transliteration task, likely because Russian orthography follows pronunciation very closely, except for characters like soft and hard signs that can hardly be recovered from English words.

We also observe that Japanese back-transliteration has proven to be much harder than other (forward-transliteration) tasks. In general, we note that a well-performing transliteration system performs well across all metrics. We are curious about the correlation between different metrics, and the results (specifically, the Spearman's rank correlation coefficient) are presented below:

- Accuracy in top-1 vs. F-score: 0.40
- Accuracy in top-1 vs. MRR: 0.97
- Accuracy in top-1 vs. MAP_{ref}: 0.997

[1]To maintain anonymity, papers of the teams that submitted anonymous results are not cited in this report.

- Accuracy in top-1 vs. MAP_{10}: 0.89
- Accuracy in top-1 vs. MAP_{sys}: 0.80

We find that F-score is the most uncorrelated metric: the Spearman's rank correlation coefficient between F-score and accuracy in top-1 is 0.40 and between F-score and MRR it is 0.44. This is likely because all metrics, except for F-score, are based on word accuracy, while F-score is based on word similarity allowing non-matching words to have scores well above 0.

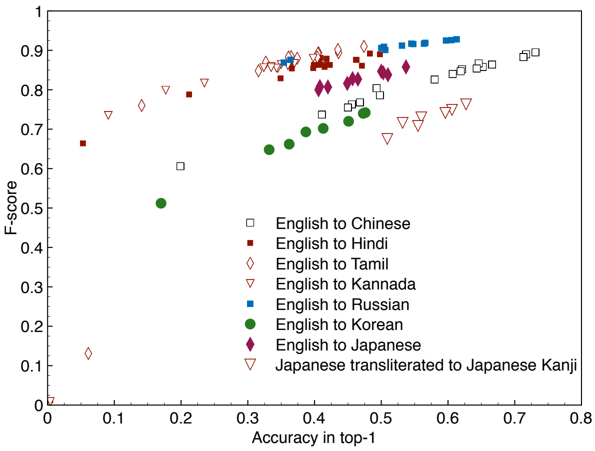


Figure 4: Accuracy in top-1 vs. F-score for different tasks.

5.2 Non-standard runs

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

- Dictionary lookup.
- Pronunciation dictionaries to convert words to their phonetic transcription.
- Additional corpora for training and dictionary lookup, such as LDC English-Chinese named entity list LDC2005T34 (Linguistic Data Consortium, 2005).
- Web search, and in particular, Wikipedia search. First, transliteration candidates are generated. Then a Web search is performed to see if any of the candidates appear in the search results. Based on the results, the candidates are re-ranked.

The results are shown in Tables 16–19. For English to Chinese and English to Russian transliteration tasks the accuracy in top-1 can go as high as 0.909 and 0.955 respectively when Web search is used to aid transliteration.

5.3 Post-evaluation

Two participants have found a bug in their system implementation and re-evaluated the results after the deadline. Their results are marked specifically in Tables 4–8 and 16.

6 Process Analysis and Fine-tuning

In this section we highlight some of the suggestions and feedback that we have received from the participants during the course of this shared task. While a few of them have been implemented in the current edition, many of these may be considered in the future editions of the shared task.

More or different languages There is quite a bit of interest in enhancing the list of language pairs short-listed. While we are constrained (in this edition) due to the availability of manually verified data, certainly more languages will be included in the future editions, as some specific data have already been promised for future editions.

Bidirectional transliteration Many participants express interest in transliterations into English; and this reflexive task will be added in the future editions. We believe it will encourage more participation as it will be easy to read and verify system output in English for those teams not familiar with the non-English side of the language.

Forward vs. backward transliteration There is quite a bit of interest expressed in specifically separating forward and backward transliteration tasks. However, such separation requires specific corpora with known origin for each name pair, and clearly we are constrained by the availability of corpora. When corpora is available, the task may be designated explicitly in future editions.

Number of standard runs The number of standard runs that may be submitted may be increased in the future editions, as many participants would like to submit many standard runs, trained with different parameters.

Errors in training and development corpora While we have taken all precautions in acquiring and creating the corpora, some errors still remain. We thank those who have sent us the errata. However, since the affected part is less than 0.5% of the data, we believe that the effect on final results is minimal. The errata will be made available to all participants.

7 Conclusions and Future Plans

We are pleased to report a comprehensive calibration and baselining of machine transliteration apporaches 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), and Conditional Random Fields (Lafferty et al., 2001) are inspired by recent progress in 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 several systems have reported improved performance over any previously reported results on similar corpora.

NEWS 2009 Shared Task represents a successful debut of a community effort in driving machine transliteration techniques forward. The overwhelming responses in the first shared task also warrant continuation of such an effort in future ACL or IJCNLP events.

Acknowledgements

The organisers of the NEWS 2009 Shared Task would like to thank the Institute for Infocomm Research (Singapore), Microsoft Research India and CJK Institute (Japan) 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 want to thank Monojit Choudhury for his contribution to metrics defined for the shared task. 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

Eiji Aramaki and Takeshi Abekawa. 2009. Fast decoding and easy implementation: Transliteration as a sequential labeling. In *Proc. ACL/IJCNLP Named Entities Workshop Shared Task*.

Dipankar Bose and Sudeshna Sarkar. 2009. Learning multi character alignment rules and classification of training data for transliteration. In *Proc. ACL/IJCNLP Named Entities Workshop Shared Task*.

Manoj Kumar Chinnakotla and Om P. Damani. 2009. Experiences with English-Hindi, English-Tamil and English-Kannada transliteration tasks at NEWS 2009. In *Proc. ACL/IJCNLP Named Entities Workshop Shared Task*.

CJKI. 2009. CJK Institute. http://www.cjk.org/.

Amitava Das, Asif Ekbal, Tapabrata Mondal, and Sivaji Bandyopadhyay. 2009. English to Hindi machine transliteration system at NEWS 2009. In *Proc. ACL/IJCNLP 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.

Dayne Freitag and Zhiqiang Wang. 2009. Name transliteration with bidirectional perceptron edit models. In *Proc. ACL/IJCNLP Named Entities Workshop Shared Task*.

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.*

Rejwanul Haque, Sandipan Dandapat, Ankit Kumar Srivastava, Sudip Kumar Naskar, and Andy Way. 2009. English-Hindi transliteration using context-informed PB-SMT. In *Proc. ACL/IJCNLP Named Entities Workshop Shared Task*.

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 Jiampojamarn, 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*.

Mitesh Khapra and Pushpak Bhattacharyya. 2009. Improving transliteration accuracy using word-origin detection and lexicon lookup. In *Proc. ACL/IJCNLP Named Entities Workshop Shared Task*.

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.

Linguistic Data Consortium. 2005. LDC Chinese-English name entity lists LDC2005T34.

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.

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

Sara Noeman. 2009. Language independent transliteration system using phrase based SMT approach on substring. In *Proc. ACL/IJCNLP Named Entities Workshop Shared Task*.

Jong-Hoon Oh, Kiyotaka Uchimoto, and Kentaro Torisawa. 2009. Machine transliteration with target-language grapheme and phoneme: Multi-engine transliteration approach. In *Proc. ACL/IJCNLP Named Entities Workshop Shared Task*.

Taraka Rama and Karthik Gali. 2009. Modeling machine transliteration as a phrase based statistical machine translation problem. In *Proc. ACL/IJCNLP Named Entities Workshop Shared Task*.

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

Praneeth Shishtla, V Surya Ganesh, S Sethuramalingam, and Vasudeva Varma. 2009. A language-independent transliteration schema using character aligned models. In *Proc. ACL/IJCNLP 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.

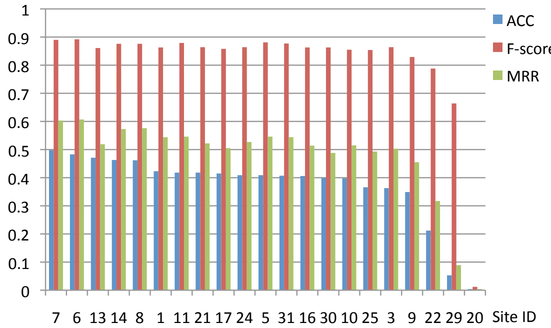
Raghavendra Udupa, K. Saravanan, Anton Bakalov, and Abhijit Bhole. 2009. "They are out there, if you know where to look": Mining transliterations of OOV query terms for cross-language information retrieval. In *LNCS: Advances in Information Retrieval*, volume 5478, pages 437–448. Springer Berlin / Heidelberg.

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

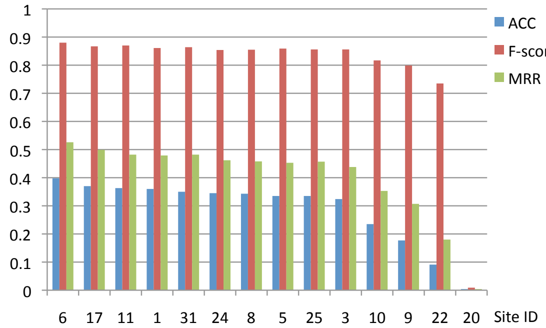
Dmitry Zelenko. 2009. Combining MDL transliteration training with discriminative modeling. In *Proc. ACL/IJCNLP Named Entities Workshop Shared Task*.

Team ID	Organisation	English to Hindi	English to Tamil	English to Kannada	English to Russian	English to Chinese	English to Korean	English to Japanese Katakana	Japanese transliterated to Japanese Kanji
		EnHi	EnTa	EnKa	EnRu	EnCh	EnKo	EnJa	JnJk
1	IIT Bombay	x	x	x					
2	Institution of Computational Linguistics Peking University					x			
3	University of Tokyo	x	x	x	x	x	x	x	
4*	University of Illinois, Urbana-Champaign				x	x			
5	IIT Bombay	x		x					
6	NICT	x	x	x	x	x	x	x	x
7	University of Alberta	x			x	x	x	x	x
8		x	x	x	x	x	x	x	x
9		x	x	x	x	x	x	x	x
10	Johns Hopkins University	x	x	x	x	x			
11		x	x	x					
12							x	x	
13	Jadavpur University	x							
14	IIIT Hyderabad	x							
15						x		x	x
16*	ARL-CACI	x							
17		x	x	x	x	x	x	x	x
18						x			
19*	Chaoyang University of Technology					x			
20	Pondicherry University	x	x	x					
21	Microsoft Research	x						x	
22	SRI International	x	x	x	x	x			
23	IBM Cairo TDC				x	x			
24	SRA	x	x	x	x	x	x	x	x
25	IIT Kharagpur	x	x	x					
26	Institute of Software Chinese Academy of Sciences					x			
27					x				
28	George Washington University					x			
29*		x							
30	Dublin City University	x							
31	IIIT	x	x	x	x	x			

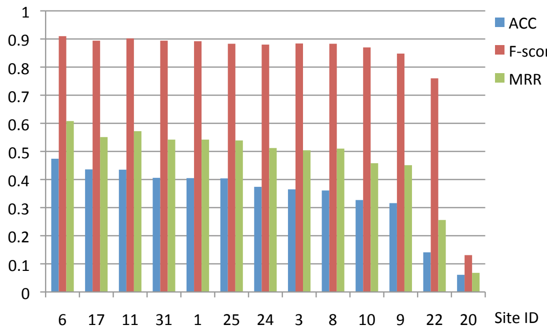
Table 3: Participation of teams in different tasks. *Participants without a system paper.



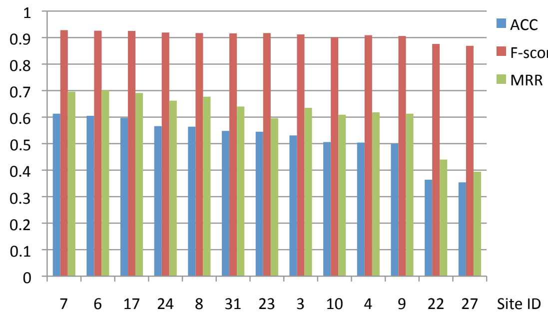
(a) English to Hindi



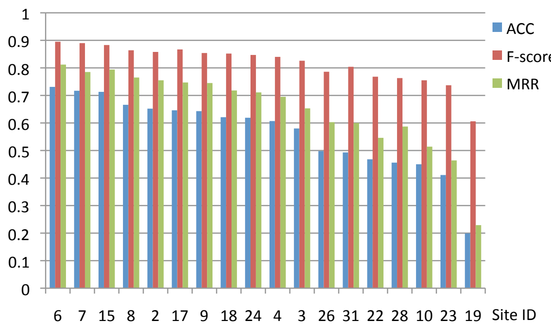
(b) English to Kannada



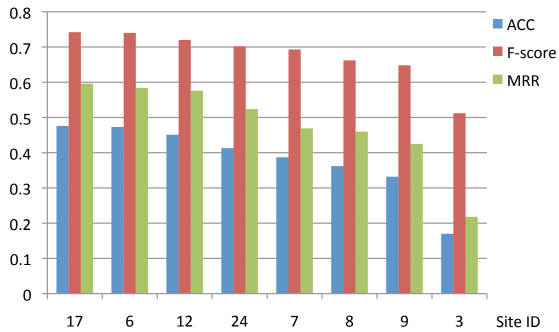
(c) English to Tamil



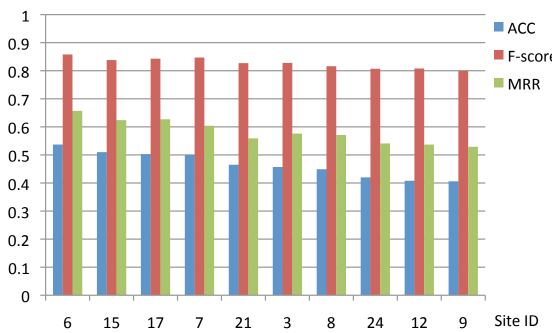
(d) English to Russian



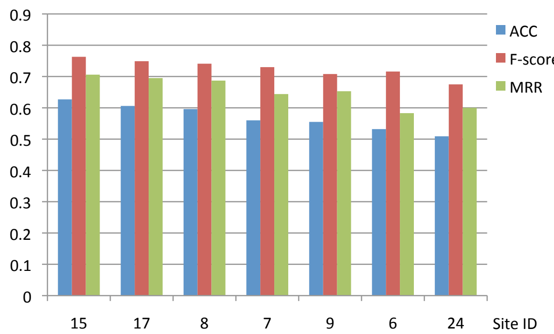
(e) English to Chinese



(f) English to Korean

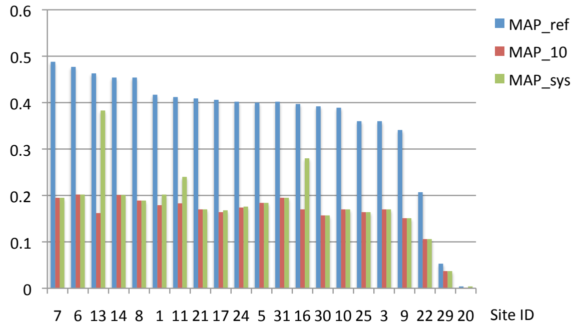


(g) English to Japanese Katakana

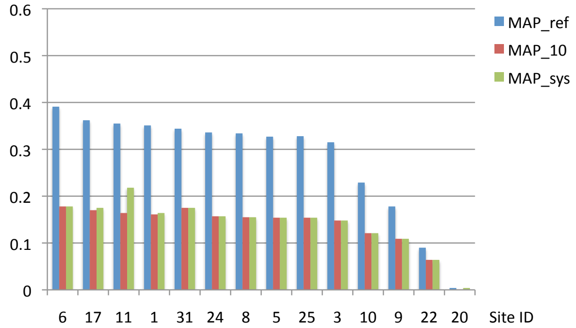


(h) Japanese transliterated to Japanese Kanji

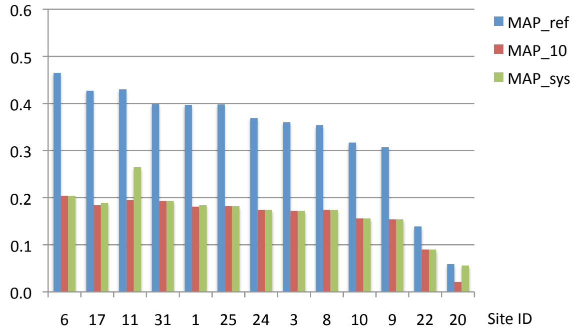
Figure 2: Accuracy in top-1, F-score and MRR for standard runs.



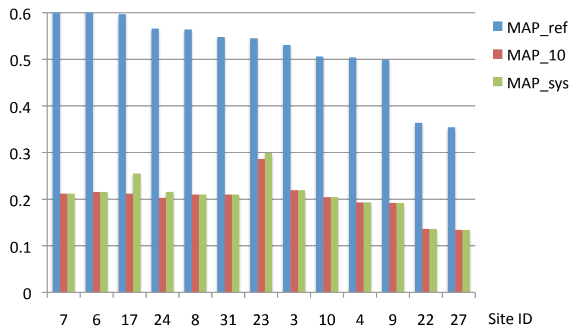
(a) English to Hindi



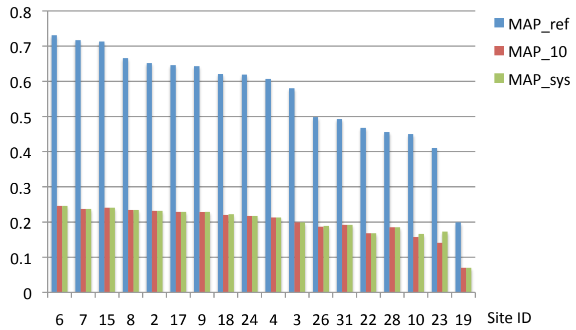
(b) English to Kannada



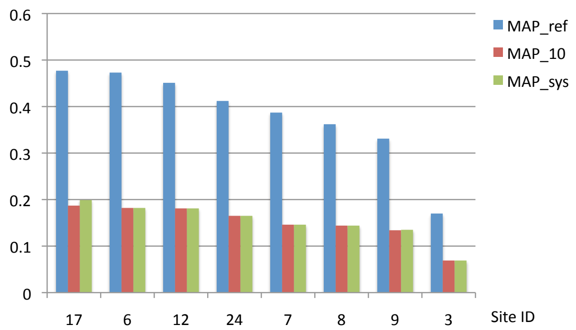
(c) English to Tamil



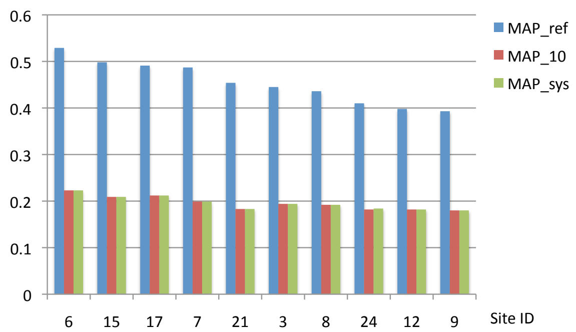
(d) English to Russian



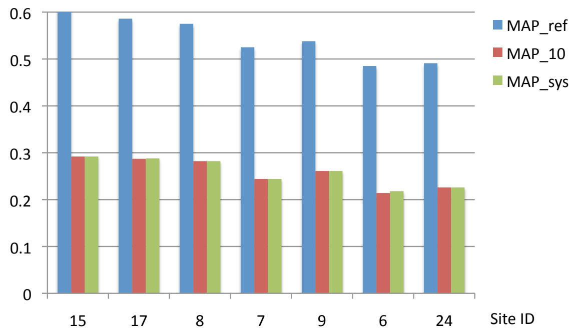
(e) English to Chinese



(f) English to Korean



(g) English to Japanese Katakana



(h) Japanese transliterated to Japanese Kanji

Figure 3: MAP_{ref}, MAP_{10} and MAP_{sys} scores for standard runs.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	MAP$_{10}$	MAP$_{sys}$	Organisation
7	0.498	0.890	0.603	0.488	0.195	0.195	University of Alberta
6	0.483	0.892	0.607	0.477	0.202	0.202	NICT
13	0.471	0.861	0.519	0.463	0.162	0.383	Jadavpur University
14	0.463	0.876	0.573	0.454	0.201	0.201	IIIT Hyderabad
8	0.462	0.876	0.576	0.454	0.189	0.189	
1	0.423	0.863	0.544	0.417	0.179	0.202	IIT Bombay
11	0.418	0.879	0.546	0.412	0.183	0.240	
21	0.418	0.864	0.522	0.409	0.170	0.170	Microsoft Research
17	0.415	0.858	0.505	0.406	0.164	0.168	
24	0.409	0.864	0.527	0.402	0.174	0.176	SRA
5	0.409	0.881	0.546	0.400	0.184	0.184	IIT Bombay
31	0.407	0.877	0.544	0.402	0.195	0.195	IIIT
16	0.406	0.863	0.514	0.397	0.170	0.280	ARL-CACI
30	0.399	0.863	0.488	0.392	0.157	0.157	Dublin City University
10	0.398	0.855	0.515	0.389	0.170	0.170	Johns Hopkins University
25	0.366	0.854	0.493	0.360	0.164	0.164	IIT Kharagpur
3	0.363	0.864	0.503	0.360	0.170	0.170	University of Tokyo
9	0.349	0.829	0.455	0.341	0.151	0.151	
22	0.212	0.788	0.317	0.207	0.106	0.106	SRI International
29	0.053	0.664	0.089	0.053	0.037	0.037	
20	0.004	0.012	0.004	0.004	0.001	0.004	Pondicherry University
21	0.466	0.881	0.567	0.457	0.183	0.183	Microsoft Research (post-evaluation)
22	0.465	0.886	0.567	0.458	0.185	0.185	SRI International (post-evaluation)

Table 4: Standard runs for English to Hindi task.

TeamID	ACC	F-score	MRR	MAP$_{ref}$	MAP$_{10}$	MAP$_{sys}$	Organisation
6	0.474	0.910	0.608	0.465	0.204	0.204	NICT
17	0.436	0.894	0.551	0.427	0.184	0.189	
11	0.435	0.902	0.572	0.430	0.195	0.265	
31	0.406	0.894	0.542	0.399	0.193	0.193	IIIT
1	0.405	0.892	0.542	0.397	0.181	0.184	IIT Bombay
25	0.404	0.883	0.539	0.398	0.182	0.182	IIT Kharagpur
24	0.374	0.880	0.512	0.369	0.174	0.174	SRA
3	0.365	0.884	0.504	0.360	0.172	0.172	University of Tokyo
8	0.361	0.883	0.510	0.354	0.174	0.174	
10	0.327	0.870	0.458	0.317	0.156	0.156	Johns Hopkins University
9	0.316	0.848	0.451	0.307	0.154	0.154	
22	0.141	0.760	0.256	0.139	0.090	0.090	SRI International
20	0.061	0.131	0.068	0.059	0.021	0.056	Pondicherry University
22	0.475	0.909	0.581	0.466	0.193	0.193	SRI International (post-evaluation)

Table 5: Standard runs for English to Tamil task.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	MAP$_{10}$	MAP$_{sys}$	Organisation
6	0.398	0.880	0.526	0.391	0.178	0.178	NICT
17	0.370	0.867	0.499	0.362	0.170	0.175	
11	0.363	0.870	0.482	0.355	0.164	0.218	
1	0.360	0.861	0.479	0.351	0.161	0.164	IIT Bombay
31	0.350	0.864	0.482	0.344	0.175	0.175	IIIT
24	0.345	0.854	0.462	0.336	0.157	0.157	SRA
8	0.343	0.855	0.458	0.334	0.155	0.155	
5	0.335	0.859	0.453	0.327	0.154	0.154	IIT Bombay
25	0.335	0.856	0.457	0.328	0.154	0.154	IIT Kharagpur
3	0.324	0.856	0.438	0.315	0.148	0.148	University of Tokyo
10	0.235	0.817	0.353	0.229	0.121	0.121	Johns Hopkins University
9	0.177	0.799	0.307	0.178	0.109	0.109	
22	0.091	0.735	0.180	0.090	0.064	0.064	SRI International
20	0.004	0.009	0.004	0.004	0.001	0.004	Pondicherry University
22	0.396	0.874	0.494	0.385	0.161	0.161	SRI International (post-evaluation)

Table 6: Standard runs for English to Kannada task.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	MAP$_{10}$	MAP$_{sys}$	Organisation
7	0.613	0.928	0.696	0.613	0.212	0.212	University of Alberta
6	0.605	0.926	0.701	0.605	0.215	0.215	NICT
17	0.597	0.925	0.691	0.597	0.212	0.255	
24	0.566	0.919	0.662	0.566	0.203	0.216	SRA
8	0.564	0.917	0.677	0.564	0.210	0.210	
31	0.548	0.916	0.640	0.548	0.210	0.210	IIIT
23	0.545	0.917	0.596	0.545	0.286	0.299	IBM Cairo TDC
3	0.531	0.912	0.635	0.531	0.219	0.219	University of Tokyo
10	0.506	0.901	0.609	0.506	0.204	0.204	Johns Hopkins University
4	0.504	0.909	0.618	0.504	0.193	0.193	University of Illinois, Urbana-Champaign
9	0.500	0.906	0.613	0.500	0.192	0.192	
22	0.364	0.876	0.440	0.364	0.136	0.136	SRI International
27	0.354	0.869	0.394	0.354	0.134	0.134	
22	0.609	0.928	0.686	0.609	0.209	0.209	SRI International (post-evaluation)

Table 7: Standard runs for English to Russian task.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	MAP$_{10}$	MAP$_{sys}$	Organisation
6	0.731	0.895	0.812	0.731	0.246	0.246	NICT
7	0.717	0.890	0.785	0.717	0.237	0.237	University of Alberta
15	0.713	0.883	0.794	0.713	0.241	0.241	
8	0.666	0.864	0.765	0.666	0.234	0.234	
2	0.652	0.858	0.755	0.652	0.232	0.232	Institution of Computational Linguistics Peking University China
17	0.646	0.867	0.747	0.646	0.229	0.229	
9	0.643	0.854	0.745	0.643	0.228	0.229	
18	0.621	0.852	0.718	0.621	0.220	0.222	
24	0.619	0.847	0.711	0.619	0.217	0.217	SRA
4	0.607	0.840	0.695	0.607	0.213	0.213	University of Illinois, Urbana-Champaign
3	0.580	0.826	0.653	0.580	0.199	0.199	University of Tokyo
26	0.498	0.786	0.603	0.498	0.187	0.189	Institute of Software Chinese Academy of Sciences
31	0.493	0.804	0.600	0.493	0.192	0.192	IIIT
22	0.468	0.768	0.546	0.468	0.168	0.168	SRI International
28	0.456	0.763	0.587	0.456	0.185	0.185	George Washington University
10	0.450	0.755	0.514	0.450	0.157	0.166	Johns Hopkins University
23	0.411	0.737	0.464	0.411	0.141	0.173	IBM Cairo TDC
19	0.199	0.606	0.229	0.199	0.070	0.070	Chaoyang University of Technology
22	0.671	0.872	0.725	0.672	0.218	0.218	SRI International (post-evaluation)

Table 8: Standard runs for English to Chinese task.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	MAP$_{10}$	MAP$_{sys}$	Organisation
17	0.476	0.742	0.596	0.477	0.187	0.199	
6	0.473	0.740	0.584	0.473	0.182	0.182	NICT
12	0.451	0.720	0.576	0.451	0.181	0.181	
24	0.413	0.702	0.524	0.412	0.165	0.165	SRA
7	0.387	0.693	0.469	0.387	0.146	0.146	University of Alberta
8	0.362	0.662	0.460	0.362	0.144	0.144	
9	0.332	0.648	0.425	0.331	0.134	0.135	
3	0.170	0.512	0.218	0.170	0.069	0.069	University of Tokyo

Table 9: Standard runs for English to Korean task.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	MAP$_{10}$	MAP$_{sys}$	Organisation
6	0.537	0.858	0.657	0.529	0.223	0.223	NICT
15	0.510	0.838	0.624	0.498	0.209	0.209	
17	0.503	0.843	0.627	0.491	0.212	0.212	
7	0.500	0.847	0.604	0.487	0.199	0.199	University of Alberta
21	0.465	0.827	0.559	0.454	0.183	0.183	Microsoft Research
3	0.457	0.828	0.576	0.445	0.194	0.194	University of Tokyo
8	0.449	0.816	0.571	0.436	0.192	0.192	
24	0.420	0.807	0.541	0.410	0.182	0.184	SRA
12	0.408	0.808	0.537	0.398	0.182	0.182	
9	0.406	0.800	0.529	0.393	0.180	0.180	
21	0.469	0.834	0.567	0.454	0.186	0.186	Microsoft Research (post-evaluation)

Table 10: Standard runs for English to Japanese Katakana task.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	MAP$_{10}$	MAP$_{sys}$	Organisation
15	0.627	0.763	0.706	0.605	0.292	0.292	
17	0.606	0.749	0.695	0.586	0.287	0.288	
8	0.596	0.741	0.687	0.575	0.282	0.282	
7	0.560	0.730	0.644	0.525	0.244	0.244	University of Alberta
9	0.555	0.708	0.653	0.538	0.261	0.261	
6	0.532	0.716	0.583	0.485	0.214	0.218	NICT
24	0.509	0.675	0.600	0.491	0.226	0.226	SRA

Table 11: Standard runs for Japanese Transliterated to Japanese Kanji task.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	MAP$_{10}$	MAP$_{sys}$	Organisation
7	0.509	0.893	0.610	0.498	0.198	0.198	University of Alberta
1	0.487	0.873	0.594	0.481	0.195	0.229	IIT Bombay
6	0.475	0.893	0.601	0.469	0.200	0.200	NICT
6	0.469	0.884	0.581	0.464	0.192	0.193	NICT
6	0.455	0.888	0.575	0.448	0.191	0.191	NICT
5	0.448	0.885	0.570	0.439	0.190	0.190	IIT Bombay
6	0.443	0.879	0.555	0.437	0.184	0.191	NICT
17	0.424	0.862	0.513	0.415	0.166	0.174	
30	0.421	0.864	0.519	0.415	0.171	0.171	Dublin City University
30	0.420	0.867	0.519	0.413	0.170	0.170	Dublin City University
30	0.419	0.868	0.464	0.419	0.338	0.338	Dublin City University
16	0.407	0.862	0.528	0.399	0.175	0.289	ARL-CACI
16	0.407	0.862	0.528	0.399	0.175	0.289	ARL-CACI
30	0.407	0.856	0.507	0.399	0.168	0.168	Dublin City University
16	0.400	0.864	0.516	0.391	0.171	0.212	ARL-CACI
13	0.389	0.831	0.487	0.385	0.160	0.328	Jadavpur University
13	0.384	0.828	0.485	0.380	0.160	0.325	Jadavpur University
16	0.273	0.796	0.358	0.266	0.119	0.193	ARL-CACI

Table 12: Non-standard runs for English to Hindi task.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	MAP$_{10}$	MAP$_{sys}$	Organisation
6	0.478	0.910	0.606	0.472	0.203	0.203	NICT
6	0.459	0.906	0.583	0.453	0.195	0.196	NICT
6	0.459	0.906	0.583	0.453	0.195	0.196	NICT
6	0.453	0.907	0.584	0.446	0.196	0.196	NICT
17	0.437	0.894	0.555	0.426	0.185	0.193	

Table 13: Non-standard runs for English to Tamil task.

Team ID	ACC	F-score	MRR	MAP_{ref}	MAP_{10}	MAP_{sys}	Organisation
6	0.399	0.881	0.522	0.391	0.176	0.176	NICT
6	0.386	0.877	0.503	0.379	0.169	0.169	NICT
6	0.380	0.869	0.488	0.370	0.163	0.163	NICT
17	0.374	0.868	0.502	0.366	0.170	0.176	
6	0.373	0.869	0.485	0.362	0.162	0.168	NICT

Table 14: Non-standard runs for English to Kannada task.

Team ID	ACC	F-score	MRR	MAP_{ref}	MAP_{10}	MAP_{sys}	Organisation
17	0.955	0.989	0.966	0.955	0.284	0.504	
17	0.609	0.928	0.701	0.609	0.214	0.263	
7	0.608	0.927	0.694	0.608	0.212	0.212	University of Alberta
7	0.607	0.927	0.690	0.607	0.211	0.211	University of Alberta
6	0.600	0.927	0.634	0.600	0.189	0.189	NICT
6	0.600	0.926	0.699	0.600	0.214	0.214	NICT
7	0.591	0.928	0.679	0.591	0.208	0.208	University of Alberta
6	0.561	0.918	0.595	0.561	0.178	0.182	NICT
6	0.557	0.920	0.596	0.557	0.179	0.233	NICT
23	0.545	0.917	0.618	0.545	0.188	0.206	IBM Cairo TDC
23	0.524	0.913	0.602	0.524	0.184	0.203	IBM Cairo TDC
23	0.524	0.913	0.579	0.524	0.277	0.291	IBM Cairo TDC
4	0.496	0.908	0.613	0.496	0.191	0.191	University of Illinois, Urbana-Champaign
27	0.338	0.872	0.408	0.338	0.128	0.128	
27	0.293	0.845	0.325	0.293	0.099	0.099	
27	0.162	0.849	0.298	0.162	0.188	0.188	

Table 15: Non-standard runs for English to Russian task.

Team ID	ACC	F-score	MRR	MAP_{ref}	MAP_{10}	MAP_{sys}	Organisation
17	0.909	0.960	0.933	0.909	0.276	0.276	
7	0.746	0.900	0.814	0.746	0.245	0.245	University of Alberta
7	0.734	0.895	0.807	0.734	0.244	0.244	University of Alberta
7	0.732	0.895	0.803	0.732	0.242	0.242	University of Alberta
6	0.731	0.894	0.812	0.731	0.246	0.246	NICT
6	0.715	0.890	0.741	0.715	0.220	0.231	NICT
6	0.699	0.884	0.729	0.699	0.216	0.232	NICT
6	0.684	0.873	0.711	0.684	0.211	0.211	NICT
22	0.663	0.867	0.754	0.663	0.230	0.230	SRI International
17	0.658	0.865	0.752	0.658	0.230	0.230	
18	0.587	0.834	0.665	0.587	0.203	0.330	
26	0.500	0.786	0.607	0.500	0.189	0.191	Institute of Software Chinese Academy of Sciences
22	0.487	0.787	0.622	0.487	0.196	0.196	SRI International
28	0.462	0.764	0.564	0.462	0.175	0.175	George Washington University
28	0.458	0.763	0.602	0.458	0.191	0.191	George Washington University
23	0.411	0.737	0.464	0.411	0.141	0.173	IBM Cairo TDC
19	0.279	0.668	0.351	0.279	0.110	0.110	Chaoyang University of Technology
28	0.058	0.353	0.269	0.058	0.101	0.101	George Washington University
28	0.050	0.359	0.260	0.050	0.098	0.098	George Washington University
4	0.001	0.249	0.001	0.001	0.000	0.000	University of Illinois, Urbana-Champaign
22	0.674	0.873	0.763	0.674	0.232	0.232	SRI International (post-evaluation)
22	0.500	0.793	0.636	0.500	0.200	0.200	SRI International (post-evaluation)

Table 16: Non-standard runs for English to Chinese task.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	MAP$_{10}$	MAP$_{sys}$	Organisation
17	0.794	0.894	0.836	0.793	0.249	0.323	
12	0.785	0.887	0.840	0.785	0.252	0.441	
12	0.784	0.889	0.840	0.784	0.252	0.484	
12	0.781	0.885	0.839	0.781	0.252	0.460	
12	0.740	0.868	0.806	0.740	0.243	0.243	
6	0.461	0.737	0.576	0.461	0.180	0.180	NICT
6	0.457	0.734	0.506	0.457	0.153	0.153	NICT
6	0.447	0.718	0.493	0.447	0.149	0.149	NICT
6	0.369	0.679	0.406	0.369	0.123	0.123	NICT

Table 17: Non-standard runs for English to Korean task.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	MAP$_{10}$	MAP$_{sys}$	Organisation
6	0.535	0.858	0.656	0.526	0.222	0.222	NICT
6	0.517	0.850	0.567	0.495	0.177	0.188	NICT
6	0.513	0.854	0.567	0.495	0.178	0.178	NICT
7	0.510	0.848	0.614	0.496	0.202	0.202	University of Alberta
6	0.500	0.842	0.547	0.480	0.170	0.196	NICT

Table 18: Non-standard runs for English to Japanese Katakana task.

Team ID	ACC	F-score	MRR	MAP$_{ref}$	MAP$_{10}$	MAP$_{sys}$	Organisation
17	0.717	0.818	0.784	0.691	0.319	0.319	
17	0.703	0.805	0.768	0.673	0.311	0.311	
17	0.698	0.805	0.774	0.676	0.317	0.317	
17	0.681	0.790	0.755	0.657	0.308	0.309	
6	0.525	0.713	0.607	0.503	0.248	0.249	NICT
6	0.525	0.712	0.606	0.502	0.248	0.248	NICT
6	0.523	0.712	0.572	0.479	0.211	0.213	NICT
6	0.517	0.705	0.603	0.496	0.248	0.249	NICT

Table 19: Non-standard runs for Japanese Transliterated to Japanese Kanji task.