Report of NEWS 2018 Named Entity Transliteration Shared Task

Nancy Chen[1], Rafael E. Banchs[2], Min Zhang[3], Xiangyu Duan[3], Haizhou Li[4]

[1] Singapore University of Technology and Design, Singapore

nancychen@alum.mit.edu

[2] Nanyang Technological University, Singapore

rbanchs@ntu.edu.sg

[3] Soochow University, China

{minzhang,xiangyuduan}@suda.edu.cn

[4] National University of Singapore, Singapore

haizhou.li@nus.edu.sg

Abstract

This report presents the results from the Named Entity Transliteration Shared Task conducted as part of The Seventh Named Entities Workshop (NEWS 2018) held at ACL 2018 in Melbourne, Australia. Similar to previous editions of NEWS, the Shared Task featured 19 tasks on proper name transliteration, including 13 different languages and two different Japanese scripts. A total of 6 teams from 8 different institutions participated in the evaluation, submitting 424 runs, involving different transliteration methodologies. Four performance metrics were used to report the evaluation results. The NEWS shared task on machine transliteration has successfully achieved its objectives by providing a common ground for the research community to conduct comparative evaluations of state-of-the-art technologies that will benefit the future research and development in this area.

1 Introduction

Names play an important role in the performance of most natural language processing and information retrieval applications. They are also critical in cross-lingual applications such as machine translation and cross-language information retrieval, as it has been shown that system performance correlates positively with the quality of name conversion across languages (Demner-Fushman and Oard 2002, Mandl and Womser-Hacker 2005, Hermjakob et al. 2008, Udupa et al. 2009). Bilingual dictionaries constitute the traditional source of information for name conversion across languages, however they offer very limited support as in most languages names are continuously emerging and evolving.

All of the above points to the critical need for robust machine transliteration methods and systems. Significant efforts has been conducted by the research community to address the problem of machine transliteration (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, Li et al. 2009a, Li et al. 2009b). These efforts fall into three main categories: grapheme-based, phoneme-based and hybrid methods. Grapheme based methods (Li et al. 2004) treat transliteration as a direct orthographic mapping and only uses orthography-related features while phoneme-based methods (Knight and Graehl 1998) make use of phonetic correspondences to generate the transliteration. The hybrid approach refers to the combination of several different models or knowledge sources to support the transliteration generation process. Recently, neural network approaches have been explored with varying successes, depending on the size of the training data.

The first machine transliteration shared task (Li et al. 2009a, Li et al. 2009b) was organized and conducted as part of NEWS 2009 at ACL-IJCNLP 2009. It was the first time that common benchmarking data in diverse language pairs was provided for evaluating state-of-the-art machine transliteration. While the focus of the 2009 shared task was on establishing the quality metrics and on setting up a baseline for transliteration quality based on those metrics, the 2010 shared task (Li et al. 2010a, Li et al. 2010b) fo-

cused on expanding the scope of the transliteration generation task to about a dozen languages and on exploring the quality of the task depending on the direction of transliteration.

In NEWS 2011 (Zhang et al. 2011a, Zhang et al. 2011b), the focus was on significantly increasing the hand-crafted parallel corpora of named entities to include 14 different language pairs from 11 language families, and on making them available as the common dataset for the shared task.

The NEWS 2018 Shared Task on Named Entity Transliteration has been a continued effort for evaluating machine transliteration performance following the NEWS edition of 2012 (Zhang et al. 2012), 2015 (Zhang et al. 2015) and 2016 (Duan et al. 2016).

In this paper, we present in full detail the results of NEWS 2018 Named Entity Transliteration Shared Task. The rest of the paper is structured as follows. Section 2 provides as short review of the main characteristics of the machine transliteration task and the corpora used for it. Section 3 reviews the four metrics used for the evaluations. Section 4 reports specific details about participation in the shared task, and section 5 presents and discusses the evaluation results. Finally, section 6 presents our main conclusions and future plans.

2 Shared Task on Transliteration

Transliteration, sometimes also called Romanization, especially if Latin Scripts are used for target strings (Halpern 2007), deals with the conversion of names between two languages and/or script systems. Within the context of this transliteration shared task, we are aiming not only at addressing the name conversion process but also its practical utility for downstream applications, such as machine translation and cross-language information retrieval.

In this context, we adopt the same definition of transliteration as proposed during NEWS 2009 (Li et al. 2009a): transliteration is understood as 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) conditioned to the following specific requirements regarding the name representation in the target language:

- it is phonetically equivalent to the source name,

- it conforms to the phonology of the target language, and
- it matches the user intuition on its equivalence with respect to the source language name.

Following previous editions of NEWS some back-transliteration tasks are considered. Back-transliteration attempts to restore transliterated names back into their original source language. NEWS 2018 included a total of six back-transliteration tasks.

2.1 Shared Task Description

As in previous editions of the workshop series, the shared task in NEWS 2018 consists of developing 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 was made available to all registered participants for developing their transliteration systems.

At the evaluation time, hand-crafted test sets of source names were released to the participants, who were required to produce a ranked list of transliteration candidates in the target language for each source name. The system outputs were tested against their corresponding reference sets (which may include multiple correct transliterations for some source names). The performance of a system is quantified using multiple metrics (defined in Section 3).

In this edition of the workshop, only *standard* runs (restricted to the train and development data provided) were considered. No other data or linguistic resources were allowed for standard runs. This ensures parity between systems and enables meaningful comparison of performance of various algorithmic approaches in a given language pair. Participants were allowed to submit one or more standard runs for each task they participated in. If more than one standard runs were submitted, it was required to select one as the "primary" run by publishing it into the leaderboard. The primary runs are the ones used to compare results across different systems.

The NEWS 2018 Shared Task was run on CodaLab (http://codalab.org/).

2.2 Shared Task Corpora

Two specific constraints were considered when 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 of linguistic diversity, orthography and geography. Following NEWS 2016, the tasks were grouped into five categories based on the specific organizations providing the datasets. The 19 tasks for NEWS 2018 are shown in Tables 1.a-e. In addition to the 14 tasks from NEWS 2016, five new tasks (highlighted in *italics*) have been included this year. This year, new evaluation data was generated and used.

Task ID	Type	Origin	Source	Target
T-EnTh	Trans.	Western	English	Thai
B-ThEn	Back.	Western	Thai	English

Table 1.a: NEWS 2018 Dataset_01

Task ID	Type	Origin	Source	Target
T-EnPe	trans.	western	English	Persian
B-PeEn	*back.*	*western*	*Persian*	*English*

Table 1.b: NEWS 2018 Dataset_02

Task ID	Type	Origin	Source	Target
T-EnCh	trans.	western	English	Chinese
B-ChEn	back.	western	Chinese	English
T-EnVi	*trans.*	*western*	*English*	*Vietnamese*

Table 1.c: NEWS 2018 Dataset_03

Task ID	Type	Origin	Source	Target
M-EnHi	mixed	mixed	English	Hindi
M-EnTa	mixed	mixed	English	Tamil
M-EnKa	mixed	mixed	English	Kannada
M-EnBa	mixed	mixed	English	Bangla
T-EnHe	trans.	western	English	Hebrew
B-HeEn	*back.*	*western*	*Hebrew*	*English*

Table 1.d: NEWS 2018 Dataset_04

Task ID	Type	Origin	Source	Target
T-EnJa	trans.	western	English	Katakana
B-JnJk	back.	japanese	English	Kanji
T-EnKo	trans.	western	English	Hangul
T-ArEn	trans.	arabic	Arabic	English
T-PeEn	*trans.*	*persian*	*Persian*	*English*
T-EnPe	*back.*	*persian*	*English*	*Persian*

Table 1.e: NEWS 2018 Dataset_05

In Tables 1.a-e, *Type* refers to the type of task (transliteration, back-transliteration or mixed); *Origin* refers to the origin of the names; and *Source*/*Target* refer to the source/target scripts.

3   Evaluation Metrics and Rationale

The participants have been asked to submit standard and, optionally, non-standard runs. One of the standard runs must be named as the primary submission, which was the one used for the performance summary. Each run must contain a ranked list of up to ten candidate transliterations for each source name. The submitted results are compared to the ground truth (reference transliterations) using four evaluation metrics capturing different aspects of transliteration performance. The four considered evaluation metrics are

- Word Accuracy in Top-1 (ACC),
- Fuzziness in Top-1 (Mean F-score) (Powers 2011),
- Mean Reciprocal Rank (MRR) (Voorhees 1999), and
- Mean Average Precision (MAP_{ref}) (Powers 2011).

In the next subsections, we present a brief description of the four considered evaluation metrics. 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.

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

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:

$$LCS(c,r) = \frac{1}{2}\big(|c| + |r| - ED(c,r)\big) \qquad \text{(Eq.2)}$$

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, i.e. the reference for which the edit distance has the minimum, is taken for calculation. If the best matching reference is given by

$$r_{i,m} = \arg\min_j\big(ED(c_{i,1}, r_{i,j})\big) \qquad \text{(Eq.3)}$$

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

$$R_i = \frac{LCS(c_{i,1}, r_{i,m})}{|r_{i,m}|} \qquad \text{(Eq.4)}$$

$$P_i = \frac{LCS(c_{i,1}, r_{i,m})}{|c_{i,1}|} \qquad \text{(Eq.5)}$$

$$F_i = 2\frac{R_i \times P_i}{R_i + P_i} \qquad \text{(Eq.6)}$$

The lengths are computed with respect to distinct Unicode characters, and no distinctions are made for different character types of a language (e.g. vowel vs. consonant vs. combining diereses).

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.

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

$$MRR = \frac{1}{N}\sum_{i=1}^{N} RR_i \qquad \text{(Eq.8)}$$

3.4 Mean Average Precision (MAP$_{ref}$)

This metric 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. If we denote the number of correct candidates for the i-th source word in k-best list as $num(i,k)$, then MAP$_{ref}$ is given by:

$$MAP_{ref} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{n_i}\Big(\sum_{k=1}^{n_i} num(i,k)\Big) \qquad \text{(Eq.9)}$$

4 Participation in the Shared Task

A total of six teams from eight different institutions participated in the NEWS 2018 Shared Task. More specifically, the participating teams were from University of Alberta (UALB), University of Edinburgh (EDI), University of Jadavpur and Universitat des Saarlandes (UJUS), Universite du Quebec a Montreal (UQAM), and team SINGA (from National University of Singapore and Singapore University of Technology and Design) and WIPO (World Intellectual Property Organization)[1].

In total, we received 424 standard runs. Table 2 summarizes the number of standard runs and the teams participated in each task.

Task	Std	Teams Participating
T-EnPe	13	UALB, EDI, UJUS, SINGA
B-ThEn	30	UALB, EDI, SINGA
T-EnTh	31	UALB, EDI, UJUS, SINGA
T-EnHe	27	UALB, EDI, UJUS, SINGA
M-EnBa	27	UALB, EDI, UJUS, SINGA
M-EnKa	29	UALB, EDI, UJUS, SINGA
M-EnTa	28	UALB, EDI, UJUS, SINGA
M-EnHi	30	UALB, EDI, UJUS, SINGA
T-ArEn	14	UALB, SINGA
B-JnJk	6	UALB
T-EnJa	17	UALB, SINGA
T-EnKo	15	UALB, SINGA
B-ChEn	29	UALB, EDI, UJUS, SINGA
T-EnCh	27	UALB, EDI, UJUS, WIPO, SINGA
T-PeEn	16	UALB, EDI, UJUS, SINGA
T-EnVi	29	UQAM, UALB, EDI, UJUS, SINGA
B-HeEn	29	UALB, EDI, UJUS, SINGA
B-EnPe	17	UALB, EDI, UJUS, SINGA
B-PeEn	10	EDI, SINGA
Overall	**424**	-

Table 2: Number of standard (Std) runs submitted, and teams participating in each task.

Table 2 shows that the most popular task continues to be the transliteration from English to Chinese (Zhang et al. 2012), followed by Chinese to English, English to Hindi, and English to Tamil.

5 Task Results and Analysis

In this section, we present the official results of the shared task along with brief descriptions of

---

[1] This last team did not submit a system paper, but we are including their submission result for the sake of completeness.

the different participant systems and some recommendations for future improvements.

5.1 Shared Task Results

Figure 1 summarizes the results of the NEWS 2018 Shared Task. In the figure, only F-scores over the NEWS 2018 evaluation test set for all primary standard submissions are depicted. A total of 66 primary standard submissions were received.

Most language pairs are able to achieve close to 80% or more in terms of F-score for at least some systems. An intriguing observation from Figure 1 is that for the language pair English-Chinese, the back-transliteration task from Chinese to English performs at least 15% better than the transliteration task from English to Chinese.

It also can be observed from the table that results for the T-EnPe and the B-PeEn tasks (western names) are significantly low. This resulted from a mismatch on scripting conventions used for the Persian language between the original train and development sets and the newly developed test set.

A much more comprehensive presentation of results for the NEWS 2018 Shared Task is provided in the Appendix at the end of this paper, where the resulting scores are reported for all received submissions for all four metrics, including non-primary submissions. All results are presented in 19 tables, each of which reports the scores for one transliteration task. In the tables, all primary standard runs are highlighted in bold-italic fonts.

5.2 Participant Systems

This year, the SINGA team (Snigdha et al. 2018) provided two baseline systems using Sequitur and Moses (phrase-based machine translation). All other systems used some version of neural modeling. It is interesting to note that non-neural systems by SINGA, while not the highest in performance, are generally comparable to neural systems or system combinations which include neural models.

Regarding the systems participating in this year evaluation, the UALB's system (Najafi et al. 2018) was based on multiple system combinations. They presented experimental results involving five different well-known transliteration approaches: DirecTL+ (Jiampojamarn et al. 2009), Sequitur (Bisani and Ney 2008), OpenNMT (Klein et al. 2017), BaseNMT (Sutskever et al. 2014), and RL-NMT (Najafi et al., 2018). They

showed improvements of up to 8% absolute over a baseline system by using system combination.
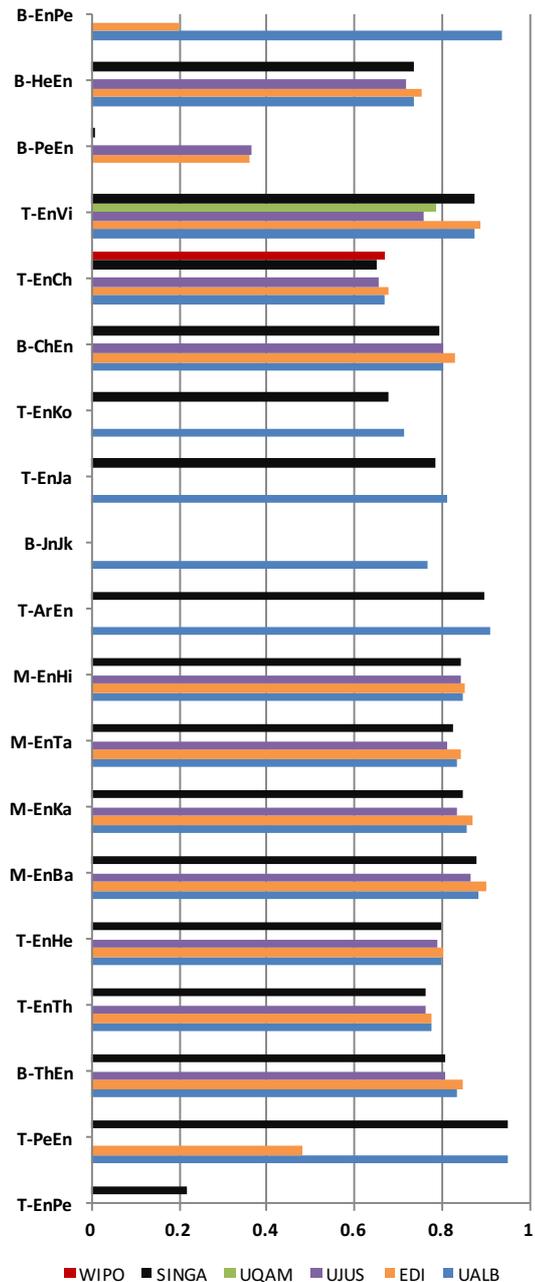
Figure 1: Mean F-scores (Top-1) on the evaluation set for all primary submissions and tasks.

The UJUS system (Kundu et al. 2018) used an RNN-based NMT framework and a CNN-based NMT framework, where both byte-pair encoding and character-based segmentation were employed for both cases. They also adopted an ensemble method to choose the hypothesis that has the highest frequency of occurrence to further improve accuracy.

The EDI system (Grundkiewicz et al. 2018) system uses a deep attention RNN encoder decoder model, which employed neural machine

translation techniques such as dropout regularization, model ensembling, and re-scoring with right-left models. The EDI system is competitive, outperforming other teams in most of the tasks it participated in.

The UQAM system (Le et al. 2018) aligned the sequences in the English Vietnamese language pair before an RNN based machine transliteration system was trained.

5.3 Issues and Recommendations

In this section, we report some issues encountered during the shared task execution along with recommendations for future improvement of the Shared Task on Named Entity Transliteration.[2]

- As mentioned in section 5.1, scripting discrepancies between the train/dev data and the test data occurred for Persian characters in the T-EnPe and B-PeEn tasks. Specifically, the newly developed test set happens to contain a mixture of the Persian and Arabic scripts, which includes visually similar characters that have distinct encodings. This dataset will be revised to resolve this problem for the next evaluation campaign.

- Some of the datasets for the shared task are available under specific licensing agreements that have to be undertaken directly by the participants from the data providers. The organizing team will explore alternative means to offer all the datasets in the shared task under a unique centralized licensing agreement, which should be ideally free of cost for the participants.

- Some of the participants experienced failures and delays during submissions to the CodaLab system. Most of these problems are due to server overloads. The organizing team will contact CodaLab support to see how these problems can be fully resolved, or at least minimized, in the future editions of the shared task.

- Participants also believe that better publicity for the shared task would result in increased participation in the task. NEWS workshop organizers receive a significant number of request for dataset and information about the shared task throughout the year. However, the total number of participants in the shared task does not reflect such actual interests from the research community on the data and

the tasks. Publicity strategies and shared task timelines will be revised accordingly.

6 Conclusions

The Shared Task on Named Entity Transliteration in NEWS 2018 has shown that the research community has a continued interest in this area. This report summarizes the results of the NEWS 2018 Shared Task.

We are pleased to report a comprehensive set of machine transliteration approaches and their evaluation results from 6 teams from 8 different institutions that participated in the shared task. This year, we received 424 runs in total. Most of the current state-of-the-art in machine transliteration is represented in the systems that have participated in the shared task.

Encouraged by the continued success of the NEWS workshop series, we plan to continue this event in the future to further promoting machine transliteration research and development.

References

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

---

[2] The organizers would like to thank all the participants, especially the University of Alberta team, for their valuable feedback and suggestions.

M. Bisani, H. Ney. 2008. Joint sequence models for grapheme-to-phoneme conversion. Speech Communication, 50(5):434–451.

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

D. Demner-Fushman, 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.

W. Gao, K.F. Wong, W. Lam. 2004. Phoneme-based transliteration of foreign names for OOV problem. In Proc. IJCNLP, pages 374–381, Sanya, Hainan, China.

Y. Goldberg, M. Elhadad. 2008. Identification of transliterated foreign words in Hebrew script. In Proc. CICLing, volume LNCS 4919, pages 466–477.

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

R. Grundkiewicz and K. Heafield. 2018. Neural Machine Translation Techniques for Named Entity Transliteration. In Proc. Named Entities Workshop at ACL 2018.

J. Halpern. 2007. The challenges and pitfalls of Arabic romanization and arabization. In Proc. Workshop on Comp. Approaches to Arabic Scriptbased Lang.

U. Hermjakob, K. Knight, H. Daum. 2008. Name translation in statistical machine translation: Learning when to transliterate. In Proc. ACL, Columbus, OH, USA, June.

S. Jiampojamarn, C. Cherry, G. Kondrak. 2010. Integrating joint n-gram features into a discriminative training framework. In Proceedings of NAACL-2010, Los Angeles, CA, June. Association for Computational Linguistics.

B.J. Kang, K.S. Choi. 2000. English-Korean automatic transliteration/ backtransliteration system and character alignment. In Proc. ACL, pages 17–18, Hong Kong.

G. Klein, Y. Kim, Y. Deng, J. Senellart, and A. M. Rush. 2017. Opennmt: Open-source toolkit for neural machine translation. In Proc. ACL.

A. Klementiev, D. 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.

K. Knight, J. Graehl. 1998. Machine transliteration. Computational Linguistics, 24(4).

P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, E. Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions, pages 177–180, Prague, Czech Republic.

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

S. Kundu, S. Paul, S. Pal. 2018. A Deep Learning Based Approach to Transliteration. In Proc. Named Entities Workshop at ACL 2018.

N. T. Le and F. Sadat. 2018. Low-Resource Machine Transliteration Using Recurrent Neural Networks of Asian Languages. In Proc. Named Entities Workshop at ACL 2018.

H. Li, M. Zhang, J. Su. 2004. A joint source-channel model for machine transliteration. In Proc. 42nd ACL Annual Meeting, pages 159–166, Barcelona, Spain.

H. Li, A. Kumaran, V. Pervouchine, M. Zhang. 2009a. Report of NEWS 2009 machine transliteration shared task. In Proc. Named Entities Workshop at ACL 2009.

H. Li, A. Kumaran, M. Zhang, V. Pervouchine. 2009b. ACL-IJCNLP 2009 Named Entities Workshop - Shared Task on Transliteration. In Proc. Named Entities Workshop at ACL 2009.

H. Li, A. Kumaran, M. Zhang, V. Pervouchine. 2010a. Report of news 2010 transliteration generation shared task. In Proc. Named Entities Workshop at ACL 2010.

H. Li, A. Kumaran, M. Zhang, V. Pervouchine. 2010b. Whitepaper of news 2010 shared task on transliteration generation. In Proc. Named Entities Workshop at ACL 2010.

T. Mandl, 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.

H.M. Meng, W.K. Lo, B. Chen, K. 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.

S. Najafi, Bradley Hauer, Rashed Rubby Riyadh, Lyeuan Yu, Gregorz Kondrak. 2018. Comparison of Assorted Models of Transliteration. In Proc. Named Entities Workshop at ACL 2018.

J.H. Oh, K.S. Choi. 2002. An English-Korean transliteration model using pronunciation and contextual rules. In Proc. COLING 2002, Taipei, Taiwan.

D. M. W. Powers. (2011). "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation" (PDF). Journal of Machine Learning Technologies. 2 (1): 37–63.

T. Sherif, G. Kondrak. 2007. Substringbased transliteration. In Proc. 45th Annual Meeting of the ACL, pages 944–951, Prague, Czech Republic, June.

S. Singhania, M. Nguyen, H. G. Ngo, N. F. Chen. 2018. Statistical Machine Transliteration Baselines for NEWS 2018. In Proc. Named Entities Workshop at ACL 2018.

R. Sproat, T. Tao, C.X. 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.

I. Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In Advances in Neural Information Processing Systems 27, pages 3104–3112.

R. Udupa, K. Saravanan, A. Bakalov, A. 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.

P. Virga, S. Khudanpur. 2003. Transliteration of proper names in cross-lingual information retrieval. In Proc. ACL MLNER, Sapporo, Japan.

E. M. Voorhees. 1999. Proceedings of the 8th Text Retrieval Conference. TREC-8 Question Answering Track Report. pp. 77–82.

S. Wan, C.M. Verspoor. 1998. Automatic English-Chinese name transliteration for development of multilingual resources. In Proc. COLING, pages 1352–1356.

D. Zelenko, C. Aone. 2006. Discriminative methods for transliteration. In Proc. EMNLP, pages 612–617, Sydney, Australia, July.

M. Zhang, A. Kumaran, H. Li. 2011a. Whitepaper of news 2011 shared task on machine transliteration. In Proc. Named Entities Workshop at IJCNLP 2011.

M. Zhang, H. Li, A. Kumaran, M. Liu. 2011b. Report of news 2011 machine transliteration shared task. In Proc. Named Entities Workshop at IJCNLP 2011.

M. Zhang, H. Li, A. Kumaran, M. Liu. 2012. Report of NEWS 2012 Machine Transliteration Shared Task. Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, pages 10–20, Jeju, Republic of Korea.

Appendix: Evaluation Results

Team	Test Set	Accuracy	F-score	MRR	MAP
EDI	**NEWS18**	**0.0010 (1)**	**0.2111 (4)**	**0.0044 (3)**	**0.0010 (1)**
EDI	NEWS18	0.0010 (1)	0.2063 (5)	0.0041 (4)	0.0010 (1)
UALB	**NEWS18**	**0.0010 (1)**	**0.2056 (6)**	**0.0010 (5)**	**0.0010 (1)**
UALB	NEWS18	0.0010 (1)	0.2042 (7)	0.0051 (1)	0.0010 (1)
UALB	NEWS18	0.0010 (1)	0.2034 (8)	0.0051 (1)	0.0010 (1)
UALB	NEWS18	0.0010 (1)	0.2012 (9)	0.0044 (2)	0.0010 (1)
SINGA	**NEWS18**	**0.0010 (1)**	**0.2167 (1)**	**0.0010 (5)**	**0.0010 (1)**
SINGA	NEWS18	0.0010 (1)	0.2167 (1)	0.0010 (5)	0.0010 (1)
SINGA	NEWS18	0.0010 (1)	0.2167 (1)	0.0010 (5)	0.0010 (1)
UJUS	**NEWS18**	**0.0010 (1)**	**0.2145 (2)**	**0.0010 (5)**	**0.0010 (1)**
UJUS	NEWS18	0.0010 (1)	0.2137 (3)	0.0010 (5)	0.0010 (1)
UJUS	NEWS18	0.0010 (1)	0.1928 (11)	0.0010 (5)	0.0010 (1)
UJUS	NEWS18	0.0000 (2)	0.1987 (10)	0.0000 (6)	0.0000 (2)

Table A1: Results for the English to Persian transliteration task (T-EnPe) on Evaluation Test. Numbers in parentheses refer to the ranking of the submitted system.

Team	Test Set	Accuracy	F-score	MRR	MAP
EDI	**NEWS18**	**0.0033 (2)**	**0.3590 (2)**	**0.0086 (2)**	**0.0030 (2)**
EDI	NEWS18	0.0022 (3)	0.3235 (4)	0.0053 (3)	0.0019 (3)
EDI	NEWS18	0.0000 (4)	0.0014 (8)	0.0000 (4)	0.0000 (4)
SINGA	**NEWS18**	**0.0000 (4)**	**0.0098 (5)**	**0.0000 (4)**	**0.0000 (4)**
SINGA	NEWS18	0.0000 (4)	0.0077 (6)	0.0000 (4)	0.0000 (4)
SINGA	NEWS18	0.0000 (4)	0.0074 (7)	0.0000 (4)	0.0000 (4)
SINGA	NEWS18	0.0000 (4)	0.0074 (7)	0.0000 (4)	0.0000 (4)
UJUS	**NEWS18**	**0.0088 (1)**	**0.3662 (1)**	**0.0088 (1)**	**0.0078 (1)**
UJUS	NEWS18	0.0000 (4)	0.3573 (3)	0.0000 (4)	0.0000 (4)
UJUS	NEWS18	0.0000 (4)	0.3573 (3)	0.0000 (4)	0.0000 (4)

Table A2: Results for the Persian to English transliteration task (B-PeEn) on Evaluation Test. Numbers in parentheses refer to the ranking of the submitted system.

Team	Test Set	ACC	F-score	MRR	MAP$_{ref}$
EDI	**NEWS18**	**0.0010 (12)**	**0.4817 (12)**	**0.0022 (13)**	**0.0012 (13)**
EDI	NEWS18	0.0000 (13)	0.0014 (13)	0.0000 (14)	0.0000 (14)
EDI	NEWS18	0.0000 (13)	0.0000 (14)	0.0000 (14)	0.0000 (14)
UALB	**NEWS18**	**0.6880 (1)**	**0.9515 (1)**	**0.7755 (3)**	**0.6081 (1)**
UALB	NEWS18	0.6820 (2)	0.9498 (3)	0.7777 (2)	0.6050 (2)
UALB	NEWS18	0.6800 (3)	0.9508 (2)	0.7780 (1)	0.6049 (3)
UALB	NEWS18	0.6450 (6)	0.9462 (5)	0.7476 (6)	0.5786 (4)
UALB	NEWS18	0.6440 (7)	0.9429 (7)	0.7546 (4)	0.5748 (5)
UALB	NEWS18	0.6380 (8)	0.9420 (8)	0.7516 (5)	0.5721 (6)
UALB	NEWS18	0.5070 (9)	0.9174 (9)	0.5070 (10)	0.4368 (9)
UALB	NEWS18	0.3930 (10)	0.9094 (10)	0.5075 (9)	0.3486 (10)
SINGA	**NEWS18**	**0.6580 (4)**	**0.9476 (4)**	**0.6580 (7)**	**0.5701 (7)**
SINGA	NEWS18	0.6560 (5)	0.9437 (6)	0.6560 (8)	0.5663 (8)

Team	Test Set	Accuracy	F-score	MRR	MAP
SINGA	NEWS18	0.6560 (5)	0.9437 (6)	0.6560 (8)	0.5663 (8)
SINGA	NEWS18	0.2460 (11)	0.9019 (11)	0.4812 (11)	0.2363 (11)
SINGA	NEWS18	0.2460 (11)	0.9019 (11)	0.2460 (12)	0.2060 (12)

Table A3: Results for the Persian to English transliteration task (T-PeEn) on Evaluation Test. Numbers in parentheses refer to the ranking of the submitted system.

Team	Test Set	Accuracy	F-score	MRR	MAP
EDI	**NEWS18**	**0.2367 (1)**	**0.8405 (1)**	**0.3291 (1)**	**0.2367 (1)**
EDI	NEWS18	0.2155 (2)	0.8361 (2)	0.3148 (2)	0.2155 (2)
EDI	NEWS18	0.0544 (21)	0.4591 (26)	0.0687 (23)	0.0544 (21)
EDI	NEWS18	0.0504 (22)	0.4577 (27)	0.0658 (24)	0.0504 (22)
UALB	**NEWS18**	**0.2135 (3)**	**0.8348 (3)**	**0.3078 (3)**	**0.2135 (3)**
UALB	NEWS18	0.2105 (4)	0.8314 (5)	0.3016 (5)	0.2105 (4)
UALB	NEWS18	0.2064 (5)	0.8332 (4)	0.3019 (4)	0.2064 (5)
UALB	NEWS18	0.1974 (6)	0.8271 (7)	0.2873 (6)	0.1974 (6)
UALB	NEWS18	0.1934 (7)	0.8304 (6)	0.2638 (9)	0.1934 (7)
UALB	NEWS18	0.1853 (8)	0.8175 (11)	0.2700 (8)	0.1853 (8)
UALB	NEWS18	0.1813 (11)	0.8217 (9)	0.1813 (13)	0.1813 (11)
UALB	NEWS18	0.1793 (12)	0.8159 (13)	0.2586 (10)	0.1793 (12)
SINGA	**NEWS18**	**0.1833 (9)**	**0.8260 (8)**	**0.1833 (11)**	**0.1833 (9)**
SINGA	NEWS18	0.1833 (9)	0.8260 (8)	0.1833 (11)	0.1833 (9)
SINGA	NEWS18	0.1833 (9)	0.8173 (12)	0.1833 (11)	0.1833 (9)
SINGA	NEWS18	0.1823 (10)	0.8126 (14)	0.2735 (7)	0.1823 (10)
SINGA	NEWS18	0.1813 (11)	0.7996 (20)	0.1813 (13)	0.1813 (11)
SINGA	NEWS18	0.1601 (16)	0.8176 (10)	0.1601 (18)	0.1601 (16)
SINGA	NEWS18	0.1581 (17)	0.7930 (23)	0.1581 (19)	0.1581 (17)
SINGA	NEWS18	0.0000 (23)	0.7668 (24)	0.0000 (25)	0.0000 (23)
UJUS	**NEWS18**	**0.1823 (10)**	**0.8076 (17)**	**0.1823 (12)**	**0.1823 (10)**
UJUS	NEWS18	0.1793 (12)	0.8100 (16)	0.1793 (14)	0.1793 (12)
UJUS	NEWS18	0.1702 (13)	0.8039 (18)	0.1702 (15)	0.1702 (13)
UJUS	NEWS18	0.1641 (14)	0.8109 (15)	0.1641 (16)	0.1641 (14)
UJUS	NEWS18	0.1631 (15)	0.7954 (22)	0.1631 (17)	0.1631 (15)
UJUS	NEWS18	0.1541 (18)	0.8006 (19)	0.1541 (20)	0.1541 (18)
UJUS	NEWS18	0.1460 (19)	0.7995 (21)	0.1460 (21)	0.1460 (19)
UJUS	NEWS18	0.1339 (20)	0.7591 (25)	0.1339 (22)	0.1339 (20)

Table A4: Results for the English to Tamil transliteration task (M-EnTa) on Evaluation Test. Numbers in parentheses refer to the ranking of the submitted system.

Team	Test Set	Accuracy	F-score	MRR	MAP
EDI	**NEWS18**	**0.3333 (2)**	**0.8515 (1)**	**0.4455 (1)**	**0.3333 (2)**
EDI	NEWS18	0.3283 (3)	0.8501 (2)	0.4426 (2)	0.3283 (3)
EDI	NEWS18	0.0400 (25)	0.4488 (27)	0.0568 (26)	0.0400 (25)
UALB	**NEWS18**	**0.3243 (4)**	**0.8472 (4)**	**0.4287 (4)**	**0.3243 (4)**
UALB	NEWS18	0.3233 (5)	0.8472 (5)	0.3935 (8)	0.3233 (5)
UALB	NEWS18	0.3223 (6)	0.8474 (3)	0.4291 (3)	0.3223 (6)
UALB	NEWS18	0.3193 (7)	0.8438 (6)	0.4235 (5)	0.3193 (7)
UALB	NEWS18	0.3033 (11)	0.8374 (16)	0.4083 (7)	0.3033 (11)
UALB	NEWS18	0.2943 (15)	0.8407 (12)	0.2943 (18)	0.2943 (15)
UALB	NEWS18	0.2683 (19)	0.8347 (20)	0.3873 (9)	0.2683 (19)
UALB	NEWS18	0.2543 (21)	0.8290 (21)	0.3741 (10)	0.2543 (21)
UALB	NEWS18	0.0000 (27)	0.0509 (29)	0.0000 (28)	0.0000 (27)
SINGA	**NEWS18**	**0.3343 (1)**	**0.8383 (14)**	**0.3343 (11)**	**0.3343 (1)**

SINGA	NEWS18	0.3333 (2)	0.8426 (8)	0.3333 (12)	0.3333 (2)
SINGA	NEWS18	0.3153 (8)	0.8417 (9)	0.3153 (13)	0.3153 (8)
SINGA	NEWS18	0.3143 (9)	0.8407 (11)	0.4167 (6)	0.3143 (9)
SINGA	NEWS18	0.3113 (10)	0.8369 (17)	0.3113 (14)	0.3113 (10)
SINGA	NEWS18	0.3013 (12)	0.8377 (15)	0.3013 (15)	0.3013 (12)
SINGA	NEWS18	0.3013 (12)	0.8377 (15)	0.3013 (15)	0.3013 (12)
SINGA	NEWS18	0.0010 (26)	0.3856 (28)	0.0010 (27)	0.0010 (26)
SINGA	NEWS18	0.0000 (27)	0.7784 (26)	0.0000 (28)	0.0000 (27)
UJUS	**NEWS18**	**0.2993 (13)**	**0.8401 (13)**	**0.2993 (16)**	**0.2993 (13)**
UJUS	NEWS18	0.2963 (14)	0.8429 (7)	0.2963 (17)	0.2963 (14)
UJUS	NEWS18	0.2923 (16)	0.8408 (10)	0.2923 (19)	0.2923 (16)
UJUS	NEWS18	0.2833 (17)	0.8359 (18)	0.2833 (20)	0.2833 (17)
UJUS	NEWS18	0.2773 (18)	0.8347 (19)	0.2773 (21)	0.2773 (18)
UJUS	NEWS18	0.2553 (20)	0.8195 (24)	0.2553 (22)	0.2553 (20)
UJUS	NEWS18	0.2502 (22)	0.8275 (22)	0.2502 (23)	0.2502 (22)
UJUS	NEWS18	0.2472 (23)	0.8223 (23)	0.2472 (24)	0.2472 (23)
UJUS	NEWS18	0.2312 (24)	0.7982 (25)	0.2312 (25)	0.2312 (24)

Table A5: Results for the English to Hindi transliteration task (M-EnHi) on Evaluation Test. Numbers in parentheses refer to the ranking of the submitted system.

Team	Test Set	Accuracy	F-score	MRR	MAP
EDI	**NEWS18**	**0.3404 (1)**	**0.8673 (1)**	**0.4588 (1)**	**0.3404 (1)**
EDI	NEWS18	0.3343 (2)	0.8638 (2)	0.4504 (2)	0.3343 (2)
EDI	NEWS18	0.0251 (23)	0.4087 (28)	0.0361 (25)	0.0251 (23)
EDI	NEWS18	0.0221 (24)	0.4091 (27)	0.0342 (26)	0.0221 (24)
UALB	**NEWS18**	**0.3042 (3)**	**0.8569 (3)**	**0.4198 (3)**	**0.3042 (3)**
UALB	NEWS18	0.3022 (4)	0.8563 (4)	0.4152 (4)	0.3022 (4)
UALB	NEWS18	0.2912 (5)	0.8528 (5)	0.4077 (5)	0.2912 (5)
UALB	NEWS18	0.2831 (8)	0.8486 (6)	0.4043 (6)	0.2831 (8)
UALB	NEWS18	0.2510 (15)	0.8391 (11)	0.3433 (10)	0.2510 (15)
UALB	NEWS18	0.2369 (16)	0.8405 (10)	0.3691 (8)	0.2369 (16)
UALB	NEWS18	0.2339 (17)	0.8385 (12)	0.2339 (20)	0.2339 (17)
UALB	NEWS18	0.2199 (19)	0.8362 (14)	0.3502 (9)	0.2199 (19)
SINGA	**NEWS18**	**0.2851 (6)**	**0.8453 (7)**	**0.2851 (11)**	**0.2851 (6)**
SINGA	NEWS18	0.2851 (6)	0.8422 (9)	0.3899 (7)	0.2851 (6)
SINGA	NEWS18	0.2841 (7)	0.8439 (8)	0.2841 (12)	0.2841 (7)
SINGA	NEWS18	0.2841 (7)	0.8439 (8)	0.2841 (12)	0.2841 (7)
SINGA	NEWS18	0.2781 (9)	0.8279 (20)	0.2781 (13)	0.2781 (9)
SINGA	NEWS18	0.2711 (10)	0.8313 (18)	0.2711 (14)	0.2711 (10)
SINGA	NEWS18	0.2691 (11)	0.8362 (13)	0.2691 (15)	0.2691 (11)
SINGA	NEWS18	0.0000 (25)	0.7809 (26)	0.0000 (27)	0.0000 (25)
UJUS	**NEWS18**	**0.2671 (12)**	**0.8298 (19)**	**0.2671 (16)**	**0.2671 (12)**
UJUS	NEWS18	0.2651 (13)	0.8338 (17)	0.2651 (17)	0.2651 (13)
UJUS	NEWS18	0.2641 (14)	0.8345 (16)	0.2641 (18)	0.2641 (14)
UJUS	NEWS18	0.2369 (16)	0.8192 (21)	0.2369 (19)	0.2369 (16)
UJUS	NEWS18	0.2239 (18)	0.8087 (24)	0.2239 (21)	0.2239 (18)
UJUS	NEWS18	0.2179 (20)	0.8346 (15)	0.2179 (22)	0.2179 (20)
UJUS	NEWS18	0.2108 (21)	0.8169 (23)	0.2108 (23)	0.2108 (21)
UJUS	NEWS18	0.1867 (22)	0.8169 (22)	0.1867 (24)	0.1867 (22)
UJUS	NEWS18	0.1867 (22)	0.7911 (25)	0.1867 (24)	0.1867 (22)

Table A6: Results for the English to Kannada transliteration task (M-EnKa) on Evaluation Test. Numbers in parentheses refer to the ranking of the submitted system.

Team	Test Set	Accuracy	F-score	MRR	MAP
EDI	**NEWS18**	**0.4610 (1)**	**0.9006 (1)**	**0.5927 (1)**	**0.4610 (1)**
EDI	NEWS18	0.4560 (2)	0.8994 (2)	0.5907 (2)	0.4560 (2)
EDI	NEWS18	0.4560 (2)	0.8994 (2)	0.5907 (2)	0.4560 (2)
UALB	**NEWS18**	**0.4120 (3)**	**0.8812 (5)**	**0.5312 (3)**	**0.4120 (3)**
UALB	NEWS18	0.4080 (4)	0.8840 (3)	0.5295 (4)	0.4080 (4)
UALB	NEWS18	0.4070 (5)	0.8827 (4)	0.5284 (5)	0.4070 (5)
UALB	NEWS18	0.3780 (10)	0.8701 (9)	0.5093 (7)	0.3780 (10)
UALB	NEWS18	0.3580 (12)	0.8680 (13)	0.4511 (10)	0.3580 (12)
UALB	NEWS18	0.3400 (14)	0.8714 (7)	0.4746 (8)	0.3400 (14)
UALB	NEWS18	0.3350 (15)	0.8701 (10)	0.4698 (9)	0.3350 (15)
UALB	NEWS18	0.3270 (17)	0.8635 (14)	0.3270 (20)	0.3270 (17)
UALB	NEWS18	0.3270 (17)	0.8635 (14)	0.3270 (20)	0.3270 (17)
SINGA	**NEWS18**	**0.4070 (5)**	**0.8793 (6)**	**0.4070 (11)**	**0.4070 (5)**
SINGA	NEWS18	0.4060 (6)	0.8682 (12)	0.4060 (12)	0.4060 (6)
SINGA	NEWS18	0.3950 (7)	0.8684 (11)	0.5126 (6)	0.3950 (7)
SINGA	NEWS18	0.3950 (7)	0.8684 (11)	0.3950 (13)	0.3950 (7)
SINGA	NEWS18	0.3930 (8)	0.8626 (16)	0.3930 (14)	0.3930 (8)
SINGA	NEWS18	0.3820 (9)	0.8713 (8)	0.3820 (15)	0.3820 (9)
SINGA	NEWS18	0.0010 (20)	0.3629 (24)	0.0010 (23)	0.0010 (20)
SINGA	NEWS18	0.0000 (21)	0.8215 (22)	0.0000 (24)	0.0000 (21)
UJUS	**NEWS18**	**0.3820 (9)**	**0.8618 (18)**	**0.3820 (15)**	**0.3820 (9)**
UJUS	NEWS18	0.3780 (10)	0.8621 (17)	0.3780 (16)	0.3780 (10)
UJUS	NEWS18	0.3760 (11)	0.8606 (19)	0.3760 (17)	0.3760 (11)
UJUS	NEWS18	0.3430 (13)	0.8631 (15)	0.3430 (18)	0.3430 (13)
UJUS	NEWS18	0.3340 (16)	0.8540 (20)	0.3340 (19)	0.3340 (16)
UJUS	NEWS18	0.2550 (18)	0.8291 (21)	0.2550 (21)	0.2550 (18)
UJUS	NEWS18	0.1180 (19)	0.7507 (23)	0.1180 (22)	0.1180 (19)

Table A7: Results for the English to Bangla (Bengali) transliteration task (M-EnBa) on Evaluation Test. Numbers in parentheses refer to the ranking of the submitted system.

Team	Test Set	Accuracy	F-score	MRR	MAP
UQAM	**NEWS18**	**0.0260 (16)**	**0.7831 (15)**	**0.0311 (18)**	**0.0260 (16)**
UQAM	NEWS18	0.0240 (17)	0.7502 (17)	0.0292 (20)	0.0240 (17)
UQAM	NEWS18	0.0240 (17)	0.7480 (18)	0.0309 (19)	0.0240 (17)
UQAM	NEWS18	0.0120 (19)	0.7423 (19)	0.0195 (23)	0.0120 (19)
EDI	**NEWS18**	**0.5020 (1)**	**0.8893 (1)**	**0.6046 (1)**	**0.5020 (1)**
EDI	NEWS18	0.5020 (1)	0.8893 (1)	0.6046 (1)	0.5020 (1)
EDI	NEWS18	0.4940 (2)	0.8858 (3)	0.5935 (3)	0.4940 (2)
EDI	NEWS18	0.4900 (3)	0.8885 (2)	0.5967 (2)	0.4900 (3)
UALB	**NEWS18**	**0.4540 (5)**	**0.8719 (5)**	**0.5447 (4)**	**0.4540 (5)**
UALB	NEWS18	0.4360 (7)	0.8641 (7)	0.5345 (6)	0.4360 (7)
UALB	NEWS18	0.4280 (9)	0.8605 (8)	0.5266 (7)	0.4280 (9)
UALB	NEWS18	0.4200 (10)	0.8592 (9)	0.5228 (8)	0.4200 (10)
UALB	NEWS18	0.3960 (11)	0.8533 (12)	0.4952 (9)	0.3960 (11)
UALB	NEWS18	0.3960 (11)	0.8525 (13)	0.4897 (10)	0.3960 (11)
UALB	NEWS18	0.3400 (12)	0.8448 (14)	0.4047 (14)	0.3400 (12)
UALB	NEWS18	0.0080 (20)	0.6021 (24)	0.0080 (24)	0.0080 (20)
SINGA	**NEWS18**	**0.4580 (4)**	**0.8583 (11)**	**0.4580 (11)**	**0.4580 (4)**
SINGA	NEWS18	0.4500 (6)	0.8730 (4)	0.4500 (12)	0.4500 (6)
SINGA	NEWS18	0.4500 (6)	0.8730 (4)	0.4500 (12)	0.4500 (6)
SINGA	NEWS18	0.4500 (6)	0.8658 (6)	0.5377 (5)	0.4500 (6)
SINGA	NEWS18	0.4500 (6)	0.8658 (6)	0.4500 (12)	0.4500 (6)

SINGA	NEWS18	0.4340 (8)	0.8587 (10)	0.4340 (13)	0.4340 (8)
SINGA	NEWS18	0.0260 (16)	0.6764 (22)	0.0260 (21)	0.0260 (16)
SINGA	NEWS18	0.0220 (18)	0.6690 (23)	0.0220 (22)	0.0220 (18)
UJUS	**NEWS18**	**0.2000 (13)**	**0.7560 (16)**	**0.2000 (15)**	**0.2000 (13)**
UJUS	NEWS18	0.1780 (14)	0.7399 (20)	0.1780 (16)	0.1780 (14)
UJUS	NEWS18	0.0940 (15)	0.6774 (21)	0.0940 (17)	0.0940 (15)
UJUS	NEWS18	0.0080 (20)	0.5863 (25)	0.0080 (24)	0.0080 (20)
UJUS	NEWS18	0.0080 (20)	0.5088 (26)	0.0080 (24)	0.0080 (20)

Table A8: Results for the English to Vietnamese transliteration task (T-EnVi) on Evaluation Test. Numbers in parentheses refer to the ranking of the submitted system.

Team	Test Set	Accuracy	F-score	MRR	MAP
EDI	**NEWS18**	**0.1670 (1)**	**0.7740 (4)**	**0.2547 (2)**	**0.1670 (2)**
EDI	NEWS18	0.1650 (3)	0.7728 (6)	0.2533 (3)	0.1650 (4)
EDI	NEWS18	0.1640 (5)	0.7760 (1)	0.2487 (4)	0.1640 (5)
EDI	NEWS18	0.1610 (6)	0.7712 (8)	0.2479 (5)	0.1610 (6)
UALB	**NEWS18**	**0.1660 (2)**	**0.7740 (5)**	**0.2352 (6)**	**0.1660 (3)**
UALB	NEWS18	0.1660 (2)	0.7654 (9)	0.2310 (9)	0.1660 (3)
UALB	NEWS18	0.1640 (5)	0.7712 (7)	0.2340 (7)	0.1640 (5)
UALB	NEWS18	0.1610 (6)	0.7745 (2)	0.2335 (8)	0.1610 (6)
UALB	NEWS18	0.1600 (7)	0.7606 (13)	0.2306 (10)	0.1600 (7)
UALB	NEWS18	0.1550 (8)	0.7596 (15)	0.1550 (16)	0.1550 (8)
UALB	NEWS18	0.1530 (9)	0.7627 (10)	0.2242 (11)	0.1530 (9)
UALB	NEWS18	0.1480 (10)	0.7615 (11)	0.2000 (15)	0.1480 (10)
UALB	NEWS18	0.1450 (11)	0.7586 (17)	0.2177 (12)	0.1450 (11)
UALB	NEWS18	0.1450 (11)	0.7578 (19)	0.1450 (17)	0.1450 (11)
UALB	NEWS18	0.1400 (16)	0.7590 (16)	0.2076 (14)	0.1400 (16)
SINGA	**NEWS18**	**0.1430 (13)**	**0.7578 (18)**	**0.2115 (13)**	**0.1430 (13)**
SINGA	NEWS18	0.1430 (13)	0.7578 (18)	0.1430 (19)	0.1430 (13)
SINGA	NEWS18	0.1420 (14)	0.7542 (21)	0.1420 (20)	0.1420 (14)
SINGA	NEWS18	0.1410 (15)	0.7604 (14)	0.1410 (21)	0.1410 (15)
SINGA	NEWS18	0.1390 (17)	0.7511 (22)	0.1390 (22)	0.1390 (17)
SINGA	NEWS18	0.1380 (18)	0.7481 (24)	0.1380 (23)	0.1380 (18)
SINGA	NEWS18	0.0380 (26)	0.4580 (29)	0.0380 (31)	0.0380 (26)
UJUS	**NEWS18**	**0.1450 (11)**	**0.7610 (12)**	**0.1450 (17)**	**0.1450 (11)**
UJUS	NEWS18	0.1440 (12)	0.7551 (20)	0.1440 (18)	0.1440 (12)
UJUS	NEWS18	0.1350 (19)	0.7484 (23)	0.1350 (24)	0.1350 (19)
UJUS	NEWS18	0.1300 (20)	0.7449 (25)	0.1300 (25)	0.1300 (20)
UJUS	NEWS18	0.1270 (21)	0.7383 (26)	0.1270 (26)	0.1270 (21)
UJUS	NEWS18	0.1080 (22)	0.7164 (27)	0.1080 (27)	0.1080 (22)
UJUS	NEWS18	0.0760 (23)	0.6632 (28)	0.0760 (28)	0.0760 (23)
UJUS	NEWS18	0.0750 (24)	0.3984 (30)	0.0750 (29)	0.0750 (24)
UJUS	NEWS18	0.0700 (25)	0.3947 (31)	0.0700 (30)	0.0700 (25)

Table A9: Results for the English to Thai transliteration task (T-EnTh) on Evaluation Test. Numbers in parentheses refer to the ranking of the submitted system.

Team	Test Set	Accuracy	F-score	MRR	MAP
UALB	**NEWS18**	**0.3400 (1)**	**0.7113 (1)**	**0.4301 (1)**	**0.3400 (1)**
UALB	NEWS18	0.3400 (1)	0.7110 (2)	0.4273 (2)	0.3400 (1)
UALB	NEWS18	0.3190 (2)	0.6954 (3)	0.4106 (3)	0.3190 (2)
UALB	NEWS18	0.2790 (6)	0.6775 (5)	0.3688 (5)	0.2790 (6)
UALB	NEWS18	0.2780 (7)	0.6822 (4)	0.3669 (6)	0.2780 (7)
UALB	NEWS18	0.2680 (9)	0.6672 (8)	0.3368 (7)	0.2680 (9)

UALB	NEWS18	0.2450 (11)	0.6286 (9)	0.3329 (8)	0.2450 (11)
UALB	NEWS18	0.2450 (11)	0.6286 (9)	0.3329 (8)	0.2450 (11)
UALB	NEWS18	0.0070 (12)	0.2646 (13)	0.0070 (14)	0.0070 (12)
SINGA	**NEWS18**	**0.3110 (3)**	**0.6093 (10)**	**0.3788 (4)**	**0.3110 (3)**
SINGA	NEWS18	0.3110 (3)	0.6093 (10)	0.3110 (9)	0.3110 (3)
SINGA	NEWS18	0.2910 (4)	0.5877 (11)	0.2910 (10)	0.2910 (4)
SINGA	NEWS18	0.2860 (5)	0.5836 (12)	0.2860 (11)	0.2860 (5)
SINGA	NEWS18	0.2730 (8)	0.6770 (6)	0.2730 (12)	0.2730 (8)
SINGA	NEWS18	0.2590 (10)	0.6747 (7)	0.2590 (13)	0.2590 (10)

Table A10: Results for the English to Korean Hangul transliteration task (T-EnKo) on Evaluation Test. Numbers in parentheses refer to the ranking of the submitted system.

Team	Test Set	ACC	F-score	MRR	MAP$_{ref}$
UALB	**NEWS18**	**0.3904 (1)**	**0.8098 (1)**	**0.5157 (1)**	**0.3893 (1)**
UALB	NEWS18	0.3844 (2)	0.8078 (3)	0.5116 (2)	0.3825 (2)
UALB	NEWS18	0.3814 (3)	0.8093 (2)	0.5110 (3)	0.3815 (3)
UALB	NEWS18	0.3684 (4)	0.8029 (5)	0.4979 (4)	0.3688 (4)
UALB	NEWS18	0.3644 (5)	0.8030 (4)	0.4977 (5)	0.3625 (5)
UALB	NEWS18	0.3594 (6)	0.8009 (7)	0.4924 (6)	0.3583 (6)
UALB	NEWS18	0.3504 (7)	0.8024 (6)	0.4897 (7)	0.3490 (7)
UALB	NEWS18	0.3463 (8)	0.7936 (8)	0.3463 (11)	0.3428 (8)
UALB	NEWS18	0.3293 (11)	0.7803 (15)	0.4258 (10)	0.3296 (11)
UALB	NEWS18	0.3203 (12)	0.7828 (12)	0.4602 (8)	0.3209 (13)
UALB	NEWS18	0.3033 (15)	0.7806 (14)	0.3033 (16)	0.3003 (16)
SINGA	**NEWS18**	**0.3393 (9)**	**0.7829 (11)**	**0.3393 (12)**	**0.3363 (9)**
SINGA	NEWS18	0.3313 (10)	0.7851 (9)	0.3313 (13)	0.3286 (12)
SINGA	NEWS18	0.3313 (10)	0.7848 (10)	0.4536 (9)	0.3322 (10)
SINGA	NEWS18	0.3183 (13)	0.7807 (13)	0.3183 (14)	0.3153 (14)
SINGA	NEWS18	0.3043 (14)	0.7745 (16)	0.3043 (15)	0.3008 (15)
SINGA	NEWS18	0.2913 (16)	0.7737 (17)	0.2913 (17)	0.2880 (17)

Table A11: Results for the English to Japanese Katakana transliteration task (T-EnJa) on Evaluation Test. Numbers in parentheses refer to the ranking of the submitted system.

Team	Test Set	ACC	F-score	MRR	MAP$_{ref}$
EDI	**NEWS18**	**0.1836 (1)**	**0.8042 (1)**	**0.2855 (1)**	**0.1807 (1)**
EDI	NEWS18	0.1778 (2)	0.8033 (2)	0.2776 (2)	0.1750 (2)
UALB	**NEWS18**	**0.1702 (4)**	**0.7983 (6)**	**0.1702 (12)**	**0.1663 (6)**
UALB	NEWS18	0.1702 (4)	0.7983 (6)	0.1702 (12)	0.1663 (6)
UALB	NEWS18	0.1683 (5)	0.7952 (12)	0.2741 (3)	0.1673 (5)
UALB	NEWS18	0.1683 (5)	0.7946 (13)	0.2600 (6)	0.1659 (7)
UALB	NEWS18	0.1683 (5)	0.7940 (14)	0.2555 (7)	0.1659 (7)
UALB	NEWS18	0.1625 (8)	0.7965 (9)	0.2627 (5)	0.1611 (13)
UALB	NEWS18	0.1606 (9)	0.7969 (8)	0.2636 (4)	0.1587 (15)
UALB	NEWS18	0.1530 (10)	0.7962 (11)	0.2211 (9)	0.1501 (17)
UALB	NEWS18	0.1530 (10)	0.7962 (11)	0.2211 (9)	0.1501 (17)
SINGA	**NEWS18**	**0.1778 (2)**	**0.7982 (7)**	**0.1778 (10)**	**0.1740 (3)**

Team	Test Set	ACC	F-score	MRR	MAP_ref
SINGA	NEWS18	0.1759 (3)	0.8000 (4)	0.1759 (11)	0.1721 (4)
SINGA	NEWS18	0.1683 (5)	0.7986 (5)	0.1683 (13)	0.1649 (9)
SINGA	NEWS18	0.1683 (5)	0.7964 (10)	0.1683 (13)	0.1654 (8)
SINGA	NEWS18	0.1644 (7)	0.8002 (3)	0.2484 (8)	0.1620 (11)
SINGA	NEWS18	0.1644 (7)	0.8002 (3)	0.1644 (15)	0.1611 (13)
SINGA	NEWS18	0.0000 (15)	0.7373 (24)	0.0000 (22)	0.0000 (22)
UJUS	**NEWS18**	**0.1663 (6)**	**0.7884 (15)**	**0.1663 (14)**	**0.1630 (10)**
UJUS	NEWS18	0.1644 (7)	0.7825 (17)	0.1644 (15)	0.1611 (13)
UJUS	NEWS18	0.1644 (7)	0.7812 (18)	0.1644 (15)	0.1616 (12)
UJUS	NEWS18	0.1625 (8)	0.7843 (16)	0.1625 (16)	0.1592 (14)
UJUS	NEWS18	0.1606 (9)	0.7789 (19)	0.1606 (17)	0.1573 (16)
UJUS	NEWS18	0.1453 (11)	0.7519 (23)	0.1453 (18)	0.1424 (18)
UJUS	NEWS18	0.1377 (12)	0.7560 (22)	0.1377 (19)	0.1348 (19)
UJUS	NEWS18	0.1300 (13)	0.7746 (20)	0.1300 (20)	0.1286 (20)
UJUS	NEWS18	0.1205 (14)	0.7691 (21)	0.1205 (21)	0.1185 (21)

Table A12: Results for the English to Hebrew transliteration task (T-EnHe) on Evaluation Test. Numbers in parentheses refer to the ranking of the submitted system

Team	Test Set	ACC	F-score	MRR	MAP_ref
WIPO	**NEWS18**	**0.2820 (4)**	**0.6686 (4)**	**0.4040 (4)**	**0.2820 (4)**
EDI	**NEWS18**	**0.3040 (1)**	**0.6791 (1)**	**0.4364 (2)**	**0.3040 (1)**
EDI	NEWS18	0.3030 (2)	0.6776 (3)	0.4267 (3)	0.3030 (2)
EDI	NEWS18	0.3010 (3)	0.6785 (2)	0.4383 (1)	0.3010 (3)
UALB	**NEWS18**	**0.2820 (4)**	**0.6680 (5)**	**0.3854 (5)**	**0.2820 (4)**
UALB	NEWS18	0.2750 (5)	0.6634 (6)	0.3771 (6)	0.2750 (5)
UALB	NEWS18	0.2750 (5)	0.6634 (6)	0.3771 (6)	0.2750 (5)
UALB	NEWS18	0.2710 (6)	0.6627 (7)	0.2710 (12)	0.2710 (6)
UALB	NEWS18	0.2600 (12)	0.6516 (10)	0.3664 (8)	0.2600 (12)
UALB	NEWS18	0.2560 (13)	0.6513 (12)	0.3646 (9)	0.2560 (13)
UALB	NEWS18	0.2460 (14)	0.6435 (19)	0.3108 (10)	0.2460 (14)
UALB	NEWS18	0.2280 (18)	0.6288 (21)	0.2280 (21)	0.2280 (18)
SINGA	**NEWS18**	**0.2750 (5)**	**0.6512 (13)**	**0.2750 (11)**	**0.2750 (5)**
SINGA	NEWS18	0.2700 (7)	0.6515 (11)	0.3736 (7)	0.2700 (7)
SINGA	NEWS18	0.2700 (7)	0.6515 (11)	0.2700 (13)	0.2700 (7)
SINGA	NEWS18	0.2670 (8)	0.6461 (17)	0.2670 (14)	0.2670 (8)
SINGA	NEWS18	0.2630 (9)	0.6489 (15)	0.2630 (15)	0.2630 (9)
SINGA	NEWS18	0.2620 (10)	0.6509 (14)	0.2620 (16)	0.2620 (10)
UJUS	**NEWS18**	**0.2610 (11)**	**0.6603 (8)**	**0.2610 (17)**	**0.2610 (11)**
UJUS	NEWS18	0.2610 (11)	0.6566 (9)	0.2610 (17)	0.2610 (11)
UJUS	NEWS18	0.2440 (15)	0.6443 (18)	0.2440 (18)	0.2440 (15)
UJUS	NEWS18	0.2400 (16)	0.6475 (16)	0.2400 (19)	0.2400 (16)
UJUS	NEWS18	0.2370 (17)	0.6358 (20)	0.2370 (20)	0.2370 (17)
UJUS	NEWS18	0.1870 (19)	0.6086 (22)	0.1870 (22)	0.1870 (19)
UJUS	NEWS18	0.1590 (20)	0.3497 (24)	0.1590 (23)	0.1590 (20)
UJUS	NEWS18	0.1540 (21)	0.3495 (25)	0.1540 (24)	0.1540 (21)

Team	Test Set	ACC	F-score	MRR	MAP_ref
UJUS	NEWS18	0.0410 (22)	0.4569 (23)	0.0410 (25)	0.0410 (22)

Table A13: Results for the English to Chinese transliteration task (T-EnCh) on Evaluation Test. Numbers in parentheses refer to the ranking of the submitted system.

Team	Test Set	ACC	F-score	MRR	MAP_ref
UALB	**NEWS18**	**0.3940 (1)**	**0.9087 (1)**	**0.3940 (9)**	**0.0586 (9)**
UALB	NEWS18	0.3910 (2)	0.9029 (2)	0.4949 (2)	0.1880 (1)
UALB	NEWS18	0.3900 (3)	0.9029 (3)	0.5012 (1)	0.1822 (2)
UALB	NEWS18	0.3730 (5)	0.9007 (4)	0.4688 (3)	0.1816 (3)
UALB	NEWS18	0.3730 (5)	0.8995 (5)	0.4632 (4)	0.1787 (5)
UALB	NEWS18	0.3630 (6)	0.8972 (7)	0.4559 (5)	0.1797 (4)
UALB	NEWS18	0.3520 (7)	0.8936 (9)	0.4413 (7)	0.1707 (6)
UALB	NEWS18	0.3300 (9)	0.8817 (10)	0.4167 (8)	0.1366 (8)
SINGA	**NEWS18**	**0.3750 (4)**	**0.8976 (6)**	**0.4552 (6)**	**0.1671 (7)**
SINGA	NEWS18	0.3750 (4)	0.8976 (6)	0.3750 (10)	0.0561 (10)
SINGA	NEWS18	0.3380 (8)	0.8964 (8)	0.3380 (11)	0.0507 (11)
SINGA	NEWS18	0.2910 (10)	0.8656 (12)	0.2910 (12)	0.0442 (12)
SINGA	NEWS18	0.2740 (11)	0.8645 (13)	0.2740 (13)	0.0413 (13)
SINGA	NEWS18	0.2620 (12)	0.8762 (11)	0.2620 (14)	0.0390 (14)

Table A14: Results for the Arabic to English transliteration task (T-ArEn) on Evaluation Test. Numbers in parentheses refer to the ranking of the submitted system.

Team	Test Set	ACC	F-score	MRR	MAP_ref
EDI	**NEWS18**	**0.3280 (2)**	**0.8454 (2)**	**0.4286 (2)**	**0.3278 (2)**
EDI	NEWS18	0.3050 (4)	0.8444 (3)	0.4101 (3)	0.3051 (4)
UALB	**NEWS18**	**0.3119 (3)**	**0.8089 (10)**	**0.3645 (8)**	**0.3118 (3)**
UALB	NEWS18	0.2840 (5)	0.8321 (4)	0.3741 (4)	0.2840 (5)
UALB	NEWS18	0.2819 (6)	0.8318 (5)	0.3732 (5)	0.2819 (6)
UALB	NEWS18	0.2729 (7)	0.8295 (6)	0.3657 (6)	0.2729 (7)
UALB	NEWS18	0.2715 (8)	0.8215 (9)	0.2715 (12)	0.2713 (8)
UALB	NEWS18	0.2687 (9)	0.8251 (8)	0.3645 (7)	0.2687 (9)
UALB	NEWS18	0.2617 (12)	0.8251 (7)	0.3529 (9)	0.2617 (12)
UALB	NEWS18	0.2212 (19)	0.8032 (16)	0.3081 (11)	0.2214 (19)
UALB	NEWS18	0.2031 (21)	0.8033 (15)	0.2712 (13)	0.2031 (22)
UALB	NEWS18	0.1298 (24)	0.7555 (24)	0.1786 (25)	0.1298 (25)
UALB	NEWS18	0.0223 (27)	0.0752 (29)	0.0292 (29)	0.0223 (28)
UALB	NEWS18	0.0223 (27)	0.0752 (29)	0.0292 (29)	0.0223 (28)
SINGA	**NEWS18**	**0.2554 (13)**	**0.7731 (22)**	**0.3308 (10)**	**0.2554 (13)**
SINGA	NEWS18	0.2554 (13)	0.7731 (22)	0.2554 (16)	0.2554 (13)
SINGA	NEWS18	0.2505 (14)	0.7376 (25)	0.2505 (17)	0.2505 (14)
SINGA	NEWS18	0.2338 (16)	0.7315 (26)	0.2338 (19)	0.2338 (16)
SINGA	NEWS18	0.2289 (17)	0.8067 (12)	0.2289 (20)	0.2289 (17)
SINGA	NEWS18	0.2233 (18)	0.8041 (14)	0.2233 (21)	0.2233 (18)
SINGA	NEWS18	0.0621 (26)	0.5045 (27)	0.0621 (28)	0.0621 (27)
UJUS	**NEWS18**	**0.2680 (10)**	**0.8079 (11)**	**0.2680 (14)**	**0.2678 (10)**
UJUS	NEWS18	0.2673 (11)	0.8046 (13)	0.2673 (15)	0.2671 (11)

Team	Test Set	ACC	F-score	MRR	MAP_ref
UJUS	NEWS18	0.2352 (15)	0.8005 (17)	0.2352 (18)	0.2350 (15)
UJUS	NEWS18	0.2289 (17)	0.7987 (18)	0.2289 (20)	0.2289 (17)
UJUS	NEWS18	0.2212 (19)	0.7906 (19)	0.2212 (22)	0.2210 (20)
UJUS	NEWS18	0.2114 (20)	0.7881 (20)	0.2114 (23)	0.2113 (21)
UJUS	NEWS18	0.1989 (22)	0.7776 (21)	0.1989 (24)	0.1987 (23)
UJUS	NEWS18	0.1689 (23)	0.7588 (23)	0.1689 (26)	0.1689 (24)
UJUS	NEWS18	0.0984 (25)	0.2960 (28)	0.0984 (27)	0.0982 (26)

Table A15: Results for the Thai to English back-transliteration task (B-ThEn) on Evaluation Test. Numbers in parentheses refer to the ranking of the submitted system.

Team	Test Set	ACC	F-score	MRR	MAP_ref
UALB	**NEWS18**	**0.5930 (1)**	**0.7678 (1)**	**0.6669 (1)**	**0.3740 (3)**
UALB	NEWS18	0.5690 (2)	0.7543 (2)	0.6492 (2)	0.3840 (1)
UALB	NEWS18	0.5650 (3)	0.7529 (3)	0.6459 (3)	0.3813 (2)
UALB	NEWS18	0.5530 (4)	0.7388 (4)	0.6399 (4)	0.3546 (4)
UALB	NEWS18	0.4660 (5)	0.6919 (5)	0.4660 (6)	0.1941 (6)
UALB	NEWS18	0.3850 (6)	0.6622 (6)	0.4837 (5)	0.2442 (5)

Table A16: Results for the English to Japanese Kanji back-transliteration task (B-JnJk) on Evaluation Test. Numbers in parentheses refer to the ranking of the submitted system.

Team	Test Set	ACC	F-score	MRR	MAP_ref
EDI	**NEWS18**	**0.1525 (2)**	**0.7532 (1)**	**0.2306 (1)**	**0.1521 (2)**
EDI	NEWS18	0.1068 (3)	0.7454 (2)	0.1803 (3)	0.1068 (3)
UALB	**NEWS18**	**0.1729 (1)**	**0.7240 (10)**	**0.2181 (2)**	**0.1725 (1)**
UALB	NEWS18	0.0915 (5)	0.7316 (8)	0.0915 (14)	0.0915 (5)
UALB	NEWS18	0.0881 (6)	0.7337 (4)	0.1505 (5)	0.0881 (6)
UALB	NEWS18	0.0864 (7)	0.7331 (5)	0.1498 (6)	0.0864 (7)
UALB	NEWS18	0.0864 (7)	0.7319 (6)	0.1494 (7)	0.0864 (7)
UALB	NEWS18	0.0780 (9)	0.7316 (7)	0.1477 (8)	0.0780 (9)
UALB	NEWS18	0.0780 (9)	0.7300 (9)	0.1436 (9)	0.0780 (9)
UALB	NEWS18	0.0678 (12)	0.7234 (11)	0.1148 (11)	0.0678 (12)
UALB	NEWS18	0.0644 (13)	0.7194 (13)	0.1261 (10)	0.0644 (13)
UALB	NEWS18	0.0644 (13)	0.7129 (17)	0.1035 (12)	0.0644 (13)
SINGA	**NEWS18**	**0.0949 (4)**	**0.7135 (16)**	**0.1560 (4)**	**0.0949 (4)**
SINGA	NEWS18	0.0949 (4)	0.7135 (16)	0.0949 (13)	0.0949 (4)
SINGA	NEWS18	0.0915 (5)	0.7339 (3)	0.0915 (14)	0.0915 (5)
SINGA	NEWS18	0.0915 (5)	0.6757 (27)	0.0915 (14)	0.0915 (5)
SINGA	NEWS18	0.0864 (7)	0.6730 (28)	0.0864 (15)	0.0864 (7)
SINGA	NEWS18	0.0678 (12)	0.7205 (12)	0.0678 (20)	0.0678 (12)
SINGA	NEWS18	0.0000 (17)	0.6819 (24)	0.0000 (25)	0.0000 (17)
UJUS	**NEWS18**	**0.0831 (8)**	**0.7157 (14)**	**0.0831 (16)**	**0.0831 (8)**
UJUS	NEWS18	0.0780 (9)	0.7147 (15)	0.0780 (17)	0.0780 (9)
UJUS	NEWS18	0.0746 (10)	0.7122 (19)	0.0746 (18)	0.0746 (10)
UJUS	NEWS18	0.0712 (11)	0.7026 (21)	0.0712 (19)	0.0712 (11)
UJUS	NEWS18	0.0678 (12)	0.7031 (20)	0.0678 (20)	0.0678 (12)
UJUS	NEWS18	0.0644 (13)	0.7006 (22)	0.0644 (21)	0.0644 (13)

Team	Test Set	ACC	F-score	MRR	MAP$_{ref}$
UJUS	NEWS18	0.0610 (14)	0.7129 (18)	0.0610 (22)	0.0610 (14)
UJUS	NEWS18	0.0610 (14)	0.6788 (26)	0.0610 (22)	0.0610 (14)
UJUS	NEWS18	0.0593 (15)	0.6945 (23)	0.0593 (23)	0.0593 (15)
UJUS	NEWS18	0.0508 (16)	0.6818 (25)	0.0508 (24)	0.0508 (16)

Table A17: Results for the Hebrew to English back-transliteration task (B-HeEn) on Evaluation Test. Numbers in parentheses refer to the ranking of the submitted system.

Team	Test Set	ACC	F-score	MRR	MAP$_{ref}$
EDI	**NEWS18**	**0.0000 (14)**	**0.1996 (15)**	**0.0008 (16)**	**0.0000 (15)**
EDI	NEWS18	0.0000 (14)	0.1950 (16)	0.0007 (17)	0.0000 (15)
UALB	**NEWS18**	**0.6350 (1)**	**0.9363 (3)**	**0.7601 (1)**	**0.6348 (1)**
UALB	NEWS18	0.6300 (2)	0.9369 (2)	0.7576 (2)	0.6298 (2)
UALB	NEWS18	0.6240 (3)	0.9373 (1)	0.7555 (3)	0.6232 (3)
UALB	NEWS18	0.6120 (4)	0.9323 (6)	0.7223 (8)	0.6122 (4)
UALB	NEWS18	0.6120 (4)	0.9303 (7)	0.7426 (4)	0.6115 (5)
UALB	NEWS18	0.5990 (6)	0.9336 (4)	0.7304 (5)	0.5982 (7)
UALB	NEWS18	0.5980 (7)	0.9335 (5)	0.7304 (6)	0.5972 (8)
UALB	NEWS18	0.5920 (8)	0.9296 (8)	0.7250 (7)	0.5915 (9)
UALB	NEWS18	0.5840 (9)	0.9286 (9)	0.7204 (9)	0.5835 (10)
UALB	NEWS18	0.5320 (11)	0.9199 (11)	0.5320 (13)	0.5315 (12)
UALB	NEWS18	0.5060 (12)	0.9142 (13)	0.5060 (14)	0.5055 (13)
SINGA	**NEWS18**	**0.6100 (5)**	**0.9286 (10)**	**0.6100 (10)**	**0.6095 (6)**
SINGA	NEWS18	0.5560 (10)	0.9183 (12)	0.5560 (11)	0.5555 (11)
SINGA	NEWS18	0.3520 (13)	0.8776 (14)	0.5502 (12)	0.3518 (14)
SINGA	NEWS18	0.3520 (13)	0.8776 (14)	0.3520 (15)	0.3518 (14)

Table A18: Results for the English to Persian back-transliteration task (B-EnPe) on Evaluation Test. Numbers in parentheses refer to the ranking of the submitted system.

Team	Test Set	ACC	F-score	MRR	MAP$_{ref}$
EDI	**NEWS18**	**0.2760 (2)**	**0.8300 (1)**	**0.3860 (1)**	**0.2760 (2)**
EDI	NEWS18	0.2530 (3)	0.8257 (2)	0.3570 (3)	0.2530 (3)
UALB	**NEWS18**	**0.3000 (1)**	**0.8011 (8)**	**0.3741 (2)**	**0.3002 (1)**
UALB	NEWS18	0.2100 (4)	0.8024 (4)	0.3002 (4)	0.2100 (4)
UALB	NEWS18	0.2100 (4)	0.8024 (4)	0.3002 (4)	0.2100 (4)
UALB	NEWS18	0.2090 (5)	0.8023 (6)	0.2968 (6)	0.2090 (5)
UALB	NEWS18	0.2090 (5)	0.8023 (6)	0.2968 (6)	0.2090 (5)
UALB	NEWS18	0.2080 (6)	0.8034 (3)	0.2991 (5)	0.2080 (6)
UALB	NEWS18	0.1920 (9)	0.8024 (5)	0.1920 (10)	0.1920 (9)
UALB	NEWS18	0.1160 (22)	0.7672 (20)	0.1900 (12)	0.1160 (22)
UALB	NEWS18	0.0940 (24)	0.6607 (25)	0.1444 (22)	0.0940 (24)
SINGA	**NEWS18**	**0.1960 (7)**	**0.7636 (23)**	**0.1960 (8)**	**0.1960 (7)**
SINGA	NEWS18	0.1950 (8)	0.7840 (18)	0.2889 (7)	0.1950 (8)
SINGA	NEWS18	0.1950 (8)	0.7840 (18)	0.1950 (9)	0.1950 (8)
SINGA	NEWS18	0.1820 (12)	0.7561 (24)	0.1820 (14)	0.1820 (12)
SINGA	NEWS18	0.1790 (13)	0.7917 (11)	0.1790 (15)	0.1790 (13)

SINGA	NEWS18	0.1750 (14)	0.7850 (16)	0.1750 (16)	0.1750 (14)
SINGA	NEWS18	0.0100 (25)	0.3440 (26)	0.0100 (26)	0.0100 (25)
UJUS	**NEWS18**	**0.1910 (10)**	**0.8003 (9)**	**0.1910 (11)**	**0.1910 (10)**
UJUS	NEWS18	0.1900 (11)	0.8014 (7)	0.1900 (13)	0.1900 (11)
UJUS	NEWS18	0.1820 (12)	0.7962 (10)	0.1820 (14)	0.1820 (12)
UJUS	NEWS18	0.1600 (15)	0.7849 (17)	0.1600 (17)	0.1600 (15)
UJUS	NEWS18	0.1540 (16)	0.7875 (13)	0.1540 (18)	0.1540 (16)
UJUS	NEWS18	0.1520 (17)	0.7850 (15)	0.1520 (19)	0.1520 (17)
UJUS	NEWS18	0.1490 (18)	0.7879 (12)	0.1490 (20)	0.1490 (18)
UJUS	NEWS18	0.1470 (19)	0.7857 (14)	0.1470 (21)	0.1470 (19)
UJUS	NEWS18	0.1330 (20)	0.7656 (22)	0.1330 (23)	0.1330 (20)
UJUS	NEWS18	0.1280 (21)	0.7670 (21)	0.1280 (24)	0.1280 (21)
UJUS	NEWS18	0.1120 (23)	0.7736 (19)	0.1120 (25)	0.1120 (23)

Table A19: Results for the Chinese to English back-transliteration task (B-ChEn) on Evaluation Test. Numbers in parentheses refer to the ranking of the submitted system.