

Overview of SIGHAN 2014 Bake-off for Chinese Spelling Check

Liang-Chih Yu^{1,2}, Lung-Hao Lee^{3,4}, Yuen-Hsien Tseng³, Hsin-Hsi Chen⁴

¹Dept. of Information Management, Yuen-Ze University

²Innovation Center for Big Data and Digital Convergence, Yuen-Ze University

³Information Technology Center, National Taiwan Normal University

⁴Dept. of Computer Science and Information Engineering, National Taiwan University

lcyu@saturn.yzu.edu.tw, lhlee@ntnu.edu.tw,
samtseng@ntnu.edu.tw, hhchen@ntu.edu.tw

Abstract

This paper introduces a Chinese Spelling Check campaign organized for the SIGHAN 2014 bake-off, including task description, data preparation, performance metrics, and evaluation results based on essays written by Chinese as a foreign language learners. The hope is that such evaluations can produce more advanced Chinese spelling check techniques.

1 Introduction

Chinese spelling errors frequently arise from confusion between multiple Chinese characters which are phonologically and visually similar, but semantically distinct (Liu et al., 2011). The SIGHAN 2013 Chinese Spelling Check Bake-off was the first campaign to provide data sets as benchmarks for the objective performance evaluation of Chinese spelling checkers (Wu et al. 2013). The collected data set is publicly available at <http://ir.itc.ntnu.edu.tw/lre/sighan7csc.htm>. The competition resulted in the integration of effective NLP techniques in the development of Chinese spelling checkers. Language modeling was used to glean extra semantic clues and collect web resources together to identify and correct spelling errors (Chen et al., 2013). A hybrid model was proposed to combine language models and statistical machine translation for spelling error correction (Liu et al. 2013). A linear regression model was trained using phonological and orthographic similarities to correct misspelled characters (Chang et al. 2013). Web-based measures were adopted to score candidates for Chinese spelling error correction (Yu et al., 2013). A graph model was used to represent the

sentence, using the single source shortest path algorithm for correcting spelling errors (Jia et al. 2013)

SIGHAN 2014 Bake-off, again features a Chinese Spelling Check task, providing an evaluation platform for the development and implementation of automatic Chinese spelling checkers. Given a passage composed of several sentences, the checker should identify all possible spelling errors, highlight their locations and suggest possible corrections. While previous tasks were based on essays written by native Chinese speakers, the current task is based on essays written by learners of Chinese as a Foreign Language (CFL), which should provide a greater challenge

The rest of this article is organized as follows. Section 2 provides an overview of the SIGHAN 2014 Bake-off Chinese Spelling Check task. Section 3 introduces the data sets used for evaluation. Section 4 proposes evaluation metrics. Section 5 compares results for the various contestants. Finally, we conclude this with findings and future research directions in Section 6.

2 Task Description

This task evaluates Chinese spelling checker performance based on Chinese text passages consisting of several sentences with and without spelling errors. The checker should identify incorrect characters in the passage and suggest corrections. Each character or punctuation mark occupies 1 spot for counting location. The input instance is given a unique passage number *PID*. If the sentence contains no spelling errors, the checker should return “*PID*, 0”. If an input passage contains at least one spelling error, the output format is “*PID* [, location, correction]+”, where the symbol “+” indicates there is one or more instance of the predicting element “[, location, correction]”. “Location” and “correction”

respectively denote the location of incorrect character and its correct version. Table 1 presents some examples. In Ex. 1, the 15th character “無” is wrong, and should be “舞”. There are 3 wrong characters in Ex. 2, and correct characters “生,” “直,” and “關” should be used in locations 3, 26, and 35, respectively. Location “0” denotes that there is no spelling error in Ex. 3

Example 1	
Input	(pid= A2-1051-1) 後天是小明的生日，我要開一個無會。
Output	A2-1051-1, 15, 舞
Example 2	
Input	(pid=B1-0201-1) 我一身中的貴人就是我姨媽，從我回來台灣的時候，她一只都很照顧我，也很觀心我。
Output	B1-0201-1, 3, 生, 26, 直, 35, 關
Example 3	
Input	(pid=C1-1849-1) 聯合國報告指出，至二零五零年，全球人口將達九十二億，新增人口幾乎全來自開發中國家。
Output	C1-1849-1, 0

Table 1. Some examples used in our task

3 Data Preparation

The learner corpus used in our task was collected from the essay section of the computer-based Test of Chinese as a Foreign Language (TOCFL), administered in Taiwan. The writing test is designed according to the six proficiency levels of the Common European Framework of Reference (CEFR). A total of 1714 essays were typed online (i.e., not hand-written), and then spelling errors were manually annotated by trained native Chinese speakers who also provided corrections corresponding to each error. The essays were then split into three sets as follows

- Training Set

This set included 1,301 selected essays with a total of 5,284 spelling errors. Each essay is represented in SGML format shown in Fig. 1. The title attribute is used to describe the essay topic.

Each passage is composed of several sentences, and each passage contains at least one spelling error, and the data indicates both the error’s location and corresponding correction. All essays in this set are used to train the developed spelling checker.

```

<ESSAY title= "寫給即將初次見面的筆友的一封信">
<TEXT>
<PASSAGE id="B1-0112-1">那一天我會穿牛仔褲和紅色的外套；頭會帶著藍色的帽子。如果你找不到我，可以打我的手機。
</PASSAGE>
<PASSAGE id="B1-0112-2">我記得你說你想試試看越南菜是有什麼味覺，午餐我會帶你去吃。我也想想試看那一家的越南菜；網路站說很多人喜歡那一家餐廳。</PASSAGE>
</TEXT>
<MISTAKE id="B1-0112-1" location="19">
<WRONG>帶著</WRONG>
<CORRECTION>戴著</CORRECTION>
</MISTAKE>
<MISTAKE id="B1-0112-2" location="46">
<WRONG>網路站</WRONG>
<CORRECTION>網路上</CORRECTION>
</MISTAKE>
</ESSAY>

```

Figure 1. An essay represented in SGML format

- Dryrun Set

A total of 20 passages were given to participants to familiarize themselves with the final testing process. Each participant can submit several runs generated using different models with different parameter settings. In addition to make sure the submitted results can be correctly evaluated, participants can fine-tune their developed models in the dryrun phase. The purpose of dryrun is output format validation only, and no dryrun outcomes were considered in the official evaluation

- Test Set

Table 2 shows a statistical summary of the prepared test set. The set consists of 1,062 testing passages, each with an average of 50 characters. Half of these passages contained no spelling errors, while the other half included at least one spelling error each for a total of 792 spelling errors used to evaluate the spelling checkers. The evaluation was conducted as an open test. In addition to the data sets provided, registered re-

search teams were allowed to employ any linguistic and computational resources to detect and correct spelling errors. Besides, passages written by CFL learners may suffer from grammatical errors, missing or redundant words, poor word selection, or word ordering problems. The task in question focuses exclusively on spelling error correction.

Test Set	Stat.
Number of essays	413
Number of passages	1,062
Number of characters	53,114
Average number of characters in all passages	50.01
Number of passages <i>with</i> errors	531
Total number of characters in the passages <i>with</i> errors	26,609
Number of erroneous characters	792
Average number of characters in passages <i>with</i> errors	50.11
Average number of spelling errors in passages <i>with</i> errors	1.49
Number of passages <i>without</i> errors	531
Total number of characters in the passages <i>without</i> errors	26,505
Average number of characters in passages <i>without</i> errors	49.92

Table 2. Descriptive statistics of the test set.

4 Performance Metrics

Table 3 shows the confusion matrix used for performance evaluation. In the matrix, TP (True Positive) is the number of passages with spelling errors that are correctly identified by the spelling checker; FP (False Positive) is the number of passages in which non-existent errors are identified; TN (True Negative) is the number of passages without spelling errors which are correctly identified as such; FN (False Negative) is the number of passages with spelling errors for which no errors are detected.

Correctness is determined at two levels. (1) Detection level: all locations of incorrect characters in a given passage should be completely identical with the gold standard. (2) Correction level: all locations and corresponding corrections

of incorrect characters should be completely identical with the gold standard. The following metrics are measured at both levels with the help of the confusion matrix.

- False Positive Rate (FPR) = $FP / (FP+TN)$
- Accuracy = $(TP+TN) / (TP+FP+TN+FN)$
- Precision = $TP / (TP+FP)$
- Recall = $TP / (TP+FN)$
- $F1 = 2 * Precision * Recall / (Precision + Recall)$

Confusion Matrix		System Result	
		Positive (Erroneous)	Negative (Correct)
Gold Standard	Positive	TP	FN
	Negative	FP	TN

Table 3. Confusion matrix for evaluation.

Take for example, 5 testing inputs with gold standards shown as “C1-1765-2, 0”, “C1-1833-2, 3, 再, 47, 反”, “B1-0176-3, 15, 棄, 22, 身”, “B1-0206-5, 0”, and “B1-2707-4, 48, 現”. The system may output the result shown as “C1-1765-2, 0”, “C1-1833-2, 3, 再, 47, 返”, “B1-0176-3, 7, 氣, 22, 身, 35, 德”, “B1-0206-5, 13, 的”, and “B1-2707-4, 48, 現”. The evaluation tool will yield the following performance.

- False Positive Rate (FPR) = 0.5 (=1/2)
Notes: {“B1-0206-5, 13, 的”} / {“C1-1765-2, 0”, “B1-0206-5, 0”}
- Detection-level
 - Acc.=0.6 (=3/5). Notes: {“C1-1765-2, 0”, “C1-1833-2, 3, 47”, “B1-2707-4, 48”} / {“C1-1765-2, 0”, “C1-1833-2, 3, 47”, “B1-0176-3, 15, 22”, “B1-0206-5, 0”, “B1-2707-4, 48”}
 - Pre.= 0.5 (=2/4). Notes: {“C1-1833-2, 3, 47”, “B1-2707-4, 48”} / {“C1-1833-2, 3, 47”, “B1-0176-3, 7, 22, 35”, “B1-0206-5, 13”, “B1-2707-4, 48”}
 - Rec.= 0.67 (=2/3). Notes: {“C1-1833-2, 3, 47”, “B1-2707-4, 48”} / {“C1-1833-2, 3, 47”, “B1-0176-3, 15, 22”, “B1-2707-4, 48”}
 - F1=0.57 (=2*0.5*0.67/(0.5+0.67))

- Correction-level
 - Acc.=0.4 (=2/5). Notes: {“C1-1765-2, 0”, “B1-2707-4, 48, 現”} / {“C1-1765-2, 0”, “C1-1833-2, 3, 再, 47, 反”, “B1-0176-3, 15, 棄, 22, 身”, “B1-0206-5, 0”, “B1-2707-4, 48, 現”}
 - Pre.= 0.25 (=1/4). Notes: {“B1-2707-4, 48, 現”} / {“C1-1833-2, 3, 再, 47, 返”, “B1-0176-3, 7, 氣, 22, 身, 35, 德”, “B1-0206-5, 13, 的”, and “B1-2707-4, 48, 現”}
 - Rec.= 0.33 (=1/3). Notes: {“B1-2707-4, 48, 現”} / {“C1-1833-2, 3, 再, 47, 反”, “B1-0176-3, 15, 棄, 22, 身”, “B1-2707-4, 48, 現”}
 - F1=0.28 (=2*0.25*0.33/(0.25+0.33))

5 Evaluation Results

Table 4 summarizes the submission statistics for 19 participant teams including 10 from universi-

ties and research institutions in China (BIT, CAS, CAU, LYFYU, NJUPT, PKU, SCAU, SJTU, SUDA, and ZJOU), 8 from Taiwan (ITRI, KUAS, NCTU & NTUT, NCYU, NTHU, NTOU, SinicaCKIP, and SinicaSLMP) and one private firm (Lingage). Among 19 registered teams, 13 teams submitted their testing results. In formal testing phase, each participant can submit at most three runs that adopt different models or parameter settings. In total, we had received 34 runs.

Table 5 summarizes the participants’ developed approaches and the usage of linguistic resources for this bake-off evaluation. Among 13 teams that participated the official testing, KUAS and PKU did not submit their reports of developed models. We can observe that most of participants adopt statistical approaches such as n-gram model, language model, and machine-learning model. In addition to the Bakeoff 2013 CSC Datasets, some linguistic resources are used popularly for this bake-off evaluation such as Sinica Corpus, Web as Corpus, Google Web 1T N-gram, and Chinese Gigaword Corpus.

Participant (Ordered by abbreviations of names)	#Runs
Beijing Institute of Technology (BIT)	2
Chinese Academy of Sciences (CAS)	2
China Agriculture University (CAU)	0
Industrial Technology Research Institute (ITRI)	0
National Kaohsiung University of Applied Sciences (KUAS)	3
Lingage Inc. (Lingage)	0
Luoyang Foreign Language University (LYFYU)	0
National Chiao Tung University & National Taipei University of Technology (NCTU & NTUT)	2
National Chiayi University (NCYU)	3
Nanjing University of Posts and Telecommunications (NJUPT)	3
National Tsing Hua University (NTHU)	3
National Taiwan Ocean University (NTOU)	2
Peking University (PKU)	3
South China Agriculture University (SCAU)	3
Chinese Knowledge and Information Processing Group, IIS, Academia Sinica (SinicaCKIP)	3
Speech, Language and Music Processing Lab, IIS, Academia Sinica (SinicaSLMP)	0
Shanghai Jiao Tong University (SJTU)	3
Soochow University (SUDA)	2
Zhejiang Ocean University (ZJOU)	0
Total	34

Table 4. Submission statistics for all participants

Participant	Approach	Linguistic Resources
BIT	<ul style="list-style-type: none"> N-gram Model Heuristic Rules Layer-Based Chinese Parsing 	<ul style="list-style-type: none"> Bakeoff 2013 CSC Datasets Chinese Penn Treebank HIT-CIR TongyiciCilin (Extended) OpenCC Sinica Corpus Tsai’s list of Chinese Words
CAS	<ul style="list-style-type: none"> Decision-Making Model 	<ul style="list-style-type: none"> Bakeoff 2013 CSC Datasets, Web as Corpus
NCTU & NTUT	<ul style="list-style-type: none"> CRF-based Word Segmentation Part-of-Speech Tagger Tri-gram Language Model 	<ul style="list-style-type: none"> Chinese Gigaword Corpus Chinese Information Retrieval Benchmark Sinica Corpus Taiwan Panorama Magazine Wikipedia (zh-tw version)
NCYU	<ul style="list-style-type: none"> Rule Induction 	<ul style="list-style-type: none"> Bakeoff 2013 CSC Datasets E-HowNet
NJUPT	<ul style="list-style-type: none"> 2-Chars & 3-Chars Model CRF Model 	<ul style="list-style-type: none"> Bakeoff 2013 CSC Datasets Web as Corpus
NTHU	<ul style="list-style-type: none"> Noisy Channel Model 	<ul style="list-style-type: none"> Bakeoff 2013 CSC Datasets Google Web 1T N-gram Sinica Corpus
NTOU	<ul style="list-style-type: none"> N-gram Model Language Model Rule-based Classifier SVM-based Classifier 	<ul style="list-style-type: none"> Bakeoff 2013 CSC Datasets Sinica Corpus
SCAU	<ul style="list-style-type: none"> N-gram Model Language Model 	<ul style="list-style-type: none"> Web as Corpus
SinicaCKIP	<ul style="list-style-type: none"> Error Template Rule Tri-gram Language Model 	<ul style="list-style-type: none"> Bakeoff 2013 CSC Datasets Google Web 1T N-gram
SJTU	<ul style="list-style-type: none"> Graph Model CRF Model Rule-Based System 	<ul style="list-style-type: none"> Bakeoff 2013 CSC Datasets OpenCC Sinica Corpus Sogou Chinese Dictionary
SUDA	<ul style="list-style-type: none"> 5-gram Language Model 	<ul style="list-style-type: none"> Chinese Gigaword Corpus

Table 5. A summary of participants’ developed systems

Table 6 shows the task testing results. In addition to accurate error detection and correction, another key performance criteria is reducing the rate of false positives, *i.e.*, the mistaken identification of errors where none exist. The research teams, KUAS, NCTU&NTUT, NCYU and SUDA, achieved very low false positive rates, *i.e.*, less than 0.05.

Detection-level evaluations are designed to identify spelling errors and highlight their locations in the input passages. Accuracy is a key performance criterion, but accuracy can be affected by the distribution of testing instances. A neutral baseline can be easily achieved by always reporting all testing errors are correct without errors. According to the test data distribution, the baseline system can achieve an accuracy level of

0.5. Some systems (*i.e.*, CAS, KUAS, and NCYU) achieved promising results exceeding 0.6. Each participating team was allowed submit up to three iterative runs based on the same input, and several teams sent different runs aimed at optimizing either recall or precision rates. We thus used the F1 score to reflect the tradeoff between precision and recall. In the testing results, KUAS provided the best error detection results, providing a high F1 score of 0.633.

For correction-level evaluations, the systems need to locate errors in the passages and indicate the corresponding correct characters. The correction accuracy provided by the KUAS submission (0.7081) significantly outperformed the other teams. However, in terms of correction precision, the spelling checker developed by KUAS and

NCYU outperforms the others at 0.8. Most systems were unable to effectively correct spelling errors, with the better systems (CAS, and KUAS) achieving a correction recall rate of slightly above 0.3. The system developed by KUAS provided the highest F1 score of 0.6125 for spelling error correction. It is difficult to correct all

spelling errors found in the input passages, since some sentences contain multiple errors and only correcting some of them are regarded as a wrong case. In summary, none of the submitted systems provided superior performance in all metrics, though those submitted by KUAS, NCYU, and CAS provided best overall performance.

Submission	FPR	Detection-Level				Correction-Level			
		Acc.	Pre.	Rec.	F1	Acc.	Pre.	Rec.	F1
BIT-Run1	0.3352	0.4313	0.371	0.1977	0.258	0.4115	0.3206	0.1582	0.2119
BIT-Run2	0.3277	0.4482	0.4061	0.2241	0.2888	0.4303	0.365	0.1883	0.2484
CAS-Run1	0.1525	0.6149	0.7148	0.3823	0.4982	0.5829	0.676	0.3183	0.4328
CAS-Run2	0.1563	0.613	0.7098	0.3823	0.4969	0.581	0.6706	0.3183	0.4317
KUAS-Run1	0.1073	0.6008	0.7421	0.3089	0.4362	0.5951	0.7349	0.2976	0.4236
KUAS-Run2	0.0452	0.7194	0.9146	0.484	0.633	0.7081	0.9108	0.4614	0.6125
KUAS-Run3	0.0452	0.6525	0.8857	0.3503	0.502	0.6488	0.8835	0.3427	0.4939
NCTU&NTUT-Run1	0.0377	0.5132	0.6296	0.064	0.1162	0.5094	0.6	0.0565	0.1033
NCTU&NTUT-Run2	0.0998	0.5028	0.5138	0.1055	0.175	0.4925	0.4592	0.0847	0.1431
NCYU-Run1	0.1827	0.4831	0.4489	0.1488	0.2235	0.467	0.3899	0.1168	0.1797
NCYU-Run2	0.0414	0.6008	0.8543	0.2429	0.3783	0.5885	0.8406	0.2185	0.3468
NCYU-Run3	0.0414	0.5913	0.844	0.2241	0.3542	0.5791	0.8281	0.1996	0.3217
NJUPT-Run1	0.3898	0.403	0.3344	0.1959	0.247	0.3964	0.3191	0.1827	0.2323
NJUPT-Run2	0.6026	0.275	0.202	0.1525	0.1738	0.258	0.1645	0.1186	0.1379
NJUPT-Run3	0.5593	0.2853	0.1885	0.1299	0.1538	0.2665	0.1416	0.0923	0.1117
NTHU-Run1	0.0829	0.5235	0.6106	0.1299	0.2143	0.5113	0.56	0.1055	0.1775
NTHU-Run2	0.1507	0.5047	0.5152	0.1601	0.2443	0.484	0.4406	0.1186	0.1869
NTHU-Run3	0.3691	0.4228	0.3677	0.2147	0.2711	0.3823	0.2659	0.1337	0.1779
NTOU-Run1	0.258	0.4652	0.4219	0.1883	0.2604	0.4557	0.3965	0.1695	0.2375
NTOU-Run2	0.9925	0.1045	0.1688	0.2015	0.1837	0.0678	0.1143	0.1281	0.1208
PKU-Run1	0.9454	0.0367	0.0195	0.0188	0.0192	0.0348	0.0157	0.0151	0.0154
PKU-Run2	0.1168	0.4915	0.4609	0.0998	0.1641	0.4783	0.3861	0.0734	0.1234
PKU-Run3	0.4087	0.3616	0.2439	0.1318	0.1711	0.3418	0.1842	0.0923	0.123
SCAU-Run1	0.2034	0.4821	0.4518	0.1676	0.2445	0.4774	0.4375	0.1582	0.2324
SCAU-Run2	0.6441	0.275	0.2315	0.194	0.2111	0.2627	0.2083	0.1695	0.1869
SCAU-Run3	0.5009	0.3522	0.2907	0.2053	0.2406	0.3427	0.2712	0.1864	0.221
SinicaCKIP-Run1	0.1149	0.5169	0.5643	0.1488	0.2355	0.516	0.5612	0.1469	0.2328
SinicaCKIP-Run2	0.1827	0.564	0.6298	0.3107	0.4161	0.5395	0.589	0.2618	0.3625
SinicaCKIP-Run3	0.2655	0.5367	0.5607	0.339	0.4225	0.5104	0.5188	0.2863	0.3689
SJTU-Run1	0.5951	0.3117	0.2685	0.2185	0.2409	0.2938	0.2349	0.1827	0.2055
SJTU-Run2	0.2279	0.5471	0.5856	0.322	0.4156	0.5377	0.5709	0.3032	0.3961
SJTU-Run3	0.1921	0.5367	0.5802	0.2655	0.3643	0.5311	0.5696	0.2542	0.3516
SUDA-Run1	0.2524	0.4539	0.3881	0.1601	0.2267	0.4426	0.3527	0.1375	0.1978
SUDA-Run2	0.032	0.5292	0.7385	0.0904	0.1611	0.5235	0.7119	0.0791	0.1424

Table 6. Testing results of our Chinese spelling check task.

6 Conclusions and Future Work

This paper provides an overview of SIGHAN 2014 Bake-off Chinese spelling check, including task design, data preparation, evaluation metrics, and performance evaluation results. The task also encourages the proposal of unorthodox and innovative approaches which could lead to a breakthrough. Regardless of actual performance, all submissions contribute to the common effort to produce an effective Chinese spell checker, and

the individual reports in the Bake-off proceedings provide useful insight into Chinese language processing.

We hope the data sets collected for this Bake-off can facilitate and expedite the development of effective Chinese spelling checkers. All data sets with gold standards and evaluation tool are publicly available for research purposes at <http://ir.itc.ntnu.edu.tw/lre/clp14csc.htm>.

Based on the results of this Bake-off, we plan to build new language resources to improve existing and develop new techniques for computer-

aided Chinese language learning. In addition, new data sets obtained from CFL learners will be investigated for the future enrichment of this research topic.

Acknowledgments

Associate research fellow Dr. Li-Ping Chang, the Vice Director of Mandarin Training Center in National Taiwan Normal University (NTNU), is appreciated for supporting the NTNU learner corpus used in our evaluation. We would like to thank Kuei-Ching Lee for developing the evaluation tool. Finally, we thank all the participants for taking part in our task.

This research was partially support by Ministry of Science and Technology, Taiwan, under the grant MOST 102-2221-E-002-103-MY3, MOST 102-2221-E-155-029-MY3, and MOST 103-2221-E-003-013-MY3, and the “Aim for the Top University Project” of National Taiwan Normal University, sponsored by the Ministry of Education, Taiwan.

Reference

- Tao-Hsing Chang, Hsueh-Chih Chen, Yuen-Hsien Tseng, and Jian-Liang Zheng. 2013. Automatic detection and correction for Chinese misspelled words using phonological and orthographic similarities. *Proceedings of the 7th SIGHAN Workshop on Chinese Language Processing (SIGHAN-13)*, pages 97-101.
- Kuan-Yu Chen, Hung-Shin Lee, Chung-Han Lee, Hsin-Min Wang, and Hsin-Hsi Chen. 2013. A study of language modeling for Chinese spelling check. *Proceedings of the 7th SIGHAN Workshop on Chinese Language Processing (SIGHAN-13)*, pages 79-83.
- Zongye Jia, Peilu Wang, and Hai Zhao. 2013. Graph model for Chinese spelling checking. *Proceedings of the 7th SIGHAN Workshop on Chinese Language Processing (SIGHAN-13)*, pages 88-92.
- Xiaodong Liu, Fei Cheng, Yanyan Luo, Kevin Duh, and Yuji Matsumoto. 2013. A hybrid Chinese spelling correction using language and statistical machine translation with reranking. *Proceedings of the 7th SIGHAN Workshop on Chinese Language Processing (SIGHAN-13)*, pages 54-58.
- Chao-Lin Liu, Min-Hua Lai, Kan-Wen Tien, Yi-Hsuan Chuang, Shih-Hung Wu, and Chia-Ying Lee. 2011. Visually and phonologically similar characters in incorrect Chinese words: Analyses, identification, and applications. *ACM Transaction on Asian Language Information Processing*, 10(2), Article 10, 39 pages.
- Shih-Hung Wu, Chao-Lin Liu, and Lung-Hao Lee. 2013. Chinese spelling check evaluation at SIGHAN bake-off 2013. *Proceedings of the 7th SIGHAN Workshop on Chinese Language Processing (SIGHAN-13)*, pages 35-42.
- Liang-Chih Yu, Chao-Hong Liu, and Chung-Hsien Wu. 2013. Candidate scoring using web-based measure for Chinese spelling error correction. *Proceedings of the 7th SIGHAN Workshop on Chinese Language Processing (SIGHAN-13)*, pages 108-112.