The Fourth International Chinese Language Processing Bakeoff: Chinese Word Segmentation, Named Entity Recognition and Chinese POS Tagging

Guangjin Jin
Institute of Applied linguistics
M.O.E., P.R.C.
No.51, Chaonei Nanxiaojie
Dong Cheng District, Beijing, China
guangjin2000@163.com

Xiao Chen
Dept of Chinese, Translation & Linguistics
City University of Hong Kong
83 Tat Chee Avenue
Kowloon, Hong Kong, China
cxiao2@student.cityu.edu.hk

Abstract

The Fourth International Chinese Language Processing Bakeoff was held in 2007 to assess the state of the art in three important tasks: Chinese word segmentation, named entity recognition and Chinese POS tagging. Twenty-eight groups submitted result sets in the three tasks across two tracks and a total of seven corpora. Strong results have been found in all the tasks as well as continuing challenges.

1 Introduction

Chinese is a kind of language which does not use word delimiters in its writing system. Now a days, under the background of information explosion, many application oriented natural language processing task become more and more important, such as parsing and machine translation. Chinese tokenization, as the foundation of many downstream processing tasks, has attracted lots of research interest. However, it is still a significant challenge for all the researchers.

SIGHAN, the Special Interest Group for Chinese Language Processing of the Association for Computational Linguistics, conducted three prior word segmentation bakeoffs, in 2003, 2005 and 2006(Sproat and Emerson, 2003; Emerson, 2005; Levow, 2006), which established benchmarks for word segmentation and named entity recognition. The bakeoff presentations at SIGHAN workshops highlighted new approaches in this field.

The fourth bakeoff was jointly held with the First CIPS Chinese Language Processing Evaluation in the summer of 2007, and co-organized by SIGHAN, Chinese LDC, and the Verifying Center of Chinese Language and Character Standards of the State Language Commission of P.R.C. In this bakeoff, we continue the Chinese word segmentation and named entity recognition tasks. Furthermore, a new evaluation task has been augmented, the task for Chinese POS tagging. In this evaluation task, a participating system will take a given segmented corpus as the input, and only the POS tagging performance will be evaluated. Both closed and open track are available for this task.

2 Details of the Evaluation

2.1 Corpora

Seven corpora were provided for the evaluation: five in Simplified characters and two in traditional characters. The Simplified character corpora were provided by Microsoft Research Asia (MSRA) for NER, by University of Pennsylvania/University of Colorado (CTB) for WS and POS tagging, by Peking University for NER and POS tagging, by Shanxi University for WS. The Traditional character corpora were provided by City University of Hong Kong (CITYU) for WS, NER and POS tagging, by the Chinese Knowledge Information Processing Laboratory (CKIP) of the Academia Sinica, Taiwan for WS and POS tagging. Each data provider offered separate training and test corpora. Statistical information for each corpus appears in Table1. All

data providers were requested to supply the training and test corpora in both the standard local encoding and in Unicode (UTF-16). For all providers, missing encodings were transcoded by the organizers using the appropriate software. Primary training and truth data for word segmentation were generated by the organizers via a C++ program by uniforming sentence end tags and delimiters. For test data, all tags removed except sentence end tags.

Comparable XML format data was also provided for all corpora and all tasks. Except as noted above, no additional changes were made to the data furnished by the providers.

Table 1: Corpora for Bakeoff-4

Source	Encoding	CWS	NER	TAG[a]
CITYU	BIG5HKSCS/UTF-16	√	√	√
CKIP	BIG5/UTF-16	√		√
CTB	GB/UTF-16	√		√
MSRA	GB/UTF-16		√	
NCC	GB/UTF-16	√		√
PKU	GB/UTF-16			√
SXU	GB/UTF-16	√		

[a]TAG:Chinese POS tagging

2.2 Rules and Procedures

The fourth Bakeoff followed the structure of the former three word segmentation bakeoffs. The only difference is that participating groups ("sites") registered online and for those who could not access our web site, email registration is acceptable; On registration, all the groups are asked to identify the corpora and tasks of interest. Training data was released for download from the online registration system on August 25, 2007. Test data was released on September 25, 2007 and results were due 12:00 Beijing Time on September 28, 2007. Scores for all submitted runs were emailed to the individual groups on October 15, and were made available to all groups on a web page a few days later.

Groups could participate in either or both of two tracks for each task and corpus:

In the open track, participants could use any external data they chose in addition to the provided training data. Groups were required to specify this information in their system descriptions.

In the closed track, participants could only use information found in the provided training data. Groups were required to submit fully automatic runs and were prohibited from testing on corpora which they had previously used.

Scoring was performed automatically using a C++ program. In cases where naming errors or minor divergences from required file formats arose, a mix of manual intervention and automatic conversion was employed to enable scoring. The primary scoring program was made available to participants for follow up experiments.

3 Participating sites

A total of 42 sites registered, and 28 submitted results for scoring. A summary of participating groups with task and track information appears in Table 2. A total of 263 official runs were scored: 166 for word segmentation, 33 for named entity recognition and 64 for POS tagging.

4 Results and Discussion

4.1 Word Segmentation Results & Discussion

There are five corpus provided in the CWS track. The statistics for these corpora are in Table 3. We introduce a type-token ration(TTR) to indicate the vocabulary diversity in each corpus.

To provide a basis for comparison, we computed baseline and possible topline scores for each of the corpora. The baseline was constructed by left-to-right maximal match algorithm, using the training corpus vocabulary. The topline employed the same procedure, but instead used the test vocabulary. These results are shown in Tables 5 and 6. For the CWS task, we computed the following measures: recall (R), precision (P), equally weighted F-measure ($F = 2PR/(P + R)$), the recall, precision and F-measure on OOV (R_{OOV}, P_{OOV}, F_{OOV}), and recall, precision and F-measure on in vocabulary words (R_{IV}, P_{IV}, F_{IV}). In and out of vocabulary status are defined relative to the training corpus. Following previous bakeoffs, we employ the Central Limit Theorem for Bernoulli trials (Grinstead and

Site ID	Site Name	CITYU CWS	CKIP CWS	CTB CWS	NCC CWS	SXU CWS	CITYU NER	MSRA NER	CITYU TAG	CKIP TAG	CTB TAG	NCC TAG	PKU TAG
1	Institute of Automation, Chinese Academy of Sciences							O					
2	City University of Hong Kong	C	C	C	C	C	C(a-b) O	C O					
3	Computing Laboratory, University of Oxford	O	O	O	O	O							
5	Dept. of Decision Sciences, The Chinese University of Hong Kong	C	C O	C	C O(a-b)	C							
7	Nara Institute of Science and Technology, JAPAN	C(a-d)	C(a-d)	C(a-d)	C(a-d)	C(a-d)							
8	Nanjing Normal University	C	C(a-b)	C(a-b) O(a-c)		C(a-d) O(a-c)							
9	National Central University			C	C	C			C	C	C	C	C
11	Institute of software, Chinese Academy of Sciences							O					
14	School of Computer and Information Technology, Shanxi University				O								
15	Tungnan University	C	C										
16	Department of Chinese, Translation and Linguistics, City University of Hong Kong									C	C	C	C
18	Institute of Computational Linguistics, EECS,Peking University	C(a-b)	C	C	C(a-c)	C(a-c)	C(a-c)	C(a-b) O(a-b)					
19	NiCT/ATR	C	C	C	C	C			C	C	C	C	C
21	Faculty of Science and Technology of University of Macau, INESC Macau	C(a-d)	C(a-b)	C(a-d)	C(a-b)	C(a-b)	C(a-b)		C(a-b)				
22	Center of Intelligence Science and Technology Research of Beijing University of Posts and Telecommunications			O	O	O		O(a-b)			O	O	O
23	The Chinese University of Hong Kong						O	O					
24	France Telecom R&D Beijing Co. Ltd	C(a-b) O(a-b)	C(a-b) O(a-b)	C(a-b) O(a-b)	C(a-b)	C(a-b)	C O	C O	C	C	C	C	
26	Microsoft Research Asia and Northeastern University of China	C	C	C	C	C							
27	Simon Fraser University	C	C	C	C	C							
28	State Key Laboratory of Machine Perception, Center for Information Science, School of Electronics Engineering & Computer Science, Peking University	C O	C O	C O(a-b)	C O	C O	C O	C	C O	C O	C O	C O	C O
29	ITNLP Lab, Computer Science of Technology, Harbin Institute of Technology				C O			O				C	C O
30	Yahoo! Inc								C(a-b)	C(a-b)		C(a-b)	C(a-b)
31	Dalian University of Technology	O	O	C(a-d) O(a-b)	C(a-d) O(a-b)	C(a-d) O(a-b)		CO			C(a-b) O(a-b)	C(a-b) O(a-b)	C(a-c) O(a-b)
33	Pohang University of Science and Technology	C	C	C	C	C							
34	NOKIA(CHINA) INVESTMENT CO.,LTD. Nokia research center, Beijing				C O	C							
37	Fudan university			C	C	C		C					C
39	Language Computer Corporation	O	O	O	O	O	O	O	O	O	O	O	O

Table 2: Participating Sites by Corpus, Task, and Track

Snell, 1997) to compute 95% confidence interval as $\pm 2\sqrt{\frac{p(1-p)}{n}}$.

Chinese Word Segmentation results for all runs grouped by corpus and track appear in Tables 6-15; all tables are sorted by F-score.

Across all corpora, the best closed track F-score was achieved in the SXU corpus at 0.9623.In the open track, two systems that has exceeded the topline in the CTB corpus, and there are also three runs approaching the topline. This might because of the overlapping of testing data in this bakeoff and the training data in the last bakeoff.

According to the statistics on all the corpus for this bakeoff, there is no clear negative linear correlation between the OOV rate of a corpus and the highest score achieved on it, since the OOV words are not the only obstacle for segmentation systems to overcome.

There are some difference in the segmentation scoring system between this bakeoff and the former ones. The precision and F-measure for both IV and OOV are appended. It could be observed that, from the result tables in every corpus, the highest total F-measure is always coming up with the highest OOV and IV F-measure rather than the recall of them. So, we consider the F-measure of both IV and OOV words a more powerful indicator for the performance of the segmentation systems in some sense.

4.2 Named Entity Recognition Results & Discussion

There are only two corpus CITYU and MSRA for named entity recognition task in this bakeoff. For statistics, we compute the OOV rate of named entities for each corpus, which denotes the proportion of named entities in testing data that are not seen in training corpus.

For each submission for named entity recognition, like the former bakeoff, we compute overall phrase precision (P), recall(R), and F-measure (F), as well as the F-measure for each entity type (PER,ORG,LOC). The only difference is the recall and precision for each entity type is appended.

We compute a baseline for each corpus as in the bakeoff-3. A left-to-right maximum match algorithm was applied on the testing data with a named entity list generated from the training data. This algorithm only detects those named entities with one unique tag in training data, others are considered as incorrectly tagged. These scores for all NER corpora are found in Table 18.

Named entity recognition results for all runs grouped by corpus and track appear in Tables 19-22; all tables are sorted by F-score.

It is shown in the result table that the baseline and the system performance for MSRA corpus are better than those for CITYU corpus. However,the statistics is showing that the number of named entities in CITYU training corpus is twice as large as the number in MSRA corpus. The system performance for these two corpus are consist with the OOV rate for these two corpora. Therefore,it seems that OOV named entities is a principal challenge for named entity recognition systems. Furthermore, the F-measure of organization name recognition is the lowest one in every participant's result on every corpus. This phenomenon is potentially implying that the organization name is the most difficult one among the three categories of named entities.

There are several systems participating both the closed and open track on the same corpus. All of them perform better in the open track. This phenomenon is implying that proper external information can strongly affect the performance of named entity recognition system.

since the testing data MSRA is a subset of the training data for last bakeoff, two sites have achieved novelly high scores in the open track.

4.3 POS Tagging Result & Discussion

There are five corpora in the Chinese POS tagging task, each of them is built on different tag set and tagging standard. For statistics and evaluation, we define several terms for this task:

- Multi-tag words: the words that been assigned more than one POS-tag in either the training corpus or testing corpus. For instance, if an IV

Table 3: Chinese Word Segmentation Training and Truth data statistics

	Training			Truth				
Source	Token	WT[a]	TTR[b]	Token	WT	TTR	OOV[c]	$\mathcal{R}_{OOV}$[d]
CITYU	1092687	43639	0.0399	235631	23303	0.0989	19382	0.0823
CKIP	721549	48114	0.0667	90678	14662	0.1617	6718	0.0741
CTB	642246	42159	0.0656	80700	12188	0.1510	4480	0.0555
NCC	913466	58592	0.0641	152354	21352	0.1401	7218	0.0474
SXU	528238	32484	0.0614	113527	12428	0.1095	5815	0.0512

Table 4: Chinese Word Segmentation Baseline

Source	R	P	F	R_{OOV}	P_{OOV}	F_{OOV}	R_{IV}	P_{IV}	F_{IV}
CITYU	.9006	.8225	.8598	.0970	.2262	.1358	.9727	.8424	.9029
CKIP	.8978	.8232	.8589	.0208	.0678	.0319	.9680	.8393	.8990
CTB	.8864	.8427	.8640	.0283	.0769	.0414	.9369	.8579	.8956
NCC	.9200	.8716	.8951	.0273	.1858	.0476	.9644	.8761	.9181
SXU	.9238	.8679	.8949	.0251	.0867	.0389	.9723	.8789	.9232

Table 5: Chinese Word Segmentation Topline

Source	R	P	F	R_{OOV}	P_{OOV}	F_{OOV}	R_{IV}	P_{IV}	F_{IV}
CITYU	.9787	.9840	.9813	.9917	.9678	.9796	.9775	.9855	.9815
CKIP	.9823	.9880	.9852	.9932	.9642	.9784	.9815	.9900	.9857
CTB	.9710	.9825	.9767	.9920	.9707	.9812	.9698	.9832	.9764
NCC	.9735	.9817	.9776	.9933	.9203	.9554	.9725	.9850	.9787
SXU	.9820	.9867	.9844	.9942	.9480	.9705	.9813	.9890	.9851

Table 6: CITYU: Word Segmentation: Closed Track

ID	RunID	R	Cr	P	Cp	F	R_{OOV}	P_{OOV}	F_{OOV}	R_{IV}	P_{IV}	F_{IV}
2		.9526	.000875	.9493	.000903	.9510	.7495	.7912	.7698	.9708	.9626	.9667
5		.9513	.000887	.9430	.000955	.9471	.7339	.7752	.7540	.9707	.9570	.9638
8		.9465	.000927	.9443	.000945	.9454	.7721	.7244	.7475	.9621	.9653	.9637
24	a	.9450	.000939	.9437	.000949	.9443	.7716	.7099	.7395	.9605	.9666	.9636
26		.9490	.000906	.9372	.000999	.9430	.6780	.7591	.7163	.9733	.9511	.9621
18	b	.9421	.000962	.9339	.001023	.9380	.7074	.7050	.7062	.9631	.9543	.9587
28		.9367	.001003	.9377	.000996	.9372	.6295	.7394	.6800	.9642	.9526	.9584
27		.9386	.000988	.9325	.001033	.9355	.6708	.6840	.6773	.9626	.9541	.9584
18	a	.9296	.001054	.9290	.001058	.9293	.6862	.6541	.6698	.9514	.9549	.9532
33		.9285	.001061	.9261	.001077	.9273	.6866	.6326	.6585	.9502	.9548	.9525
7	c	.9237	.001093	.9234	.001095	.9236	.6830	.5934	.6350	.9453	.9579	.9516
7	b	.9237	.001093	.9234	.001095	.9236	.6830	.5934	.6350	.9453	.9579	.9516
7	a	.9238	.001093	.9234	.001095	.9236	.6830	.5934	.6351	.9453	.9579	.9516
7	d	.9197	.001119	.9169	.001137	.9183	.6558	.5690	.6093	.9434	.9532	.9483
15		.9191	.001123	.9014	.001228	.9102	.5466	.5588	.5527	.9525	.9308	.9415
21	b	.9219	.001105	.8951	.001262	.9083	.4703	.5899	.5234	.9624	.9159	.9386
21	a	.9221	.001104	.8947	.001264	.9082	.4697	.5891	.5227	.9627	.9155	.9385
21	d	.9120	.001167	.8974	.001250	.9047	.5263	.5333	.5297	.9466	.9290	.9377
19		.8884	.001296	.8817	.001330	.8850	.6114	.6030	.6072	.9133	.9069	.9101
21	c	.0155	.000509	.0155	.000508	.0155	.0047	.0049	.0048	.0165	.0164	.0165

Table 7: CITYU: Word Segmentation: Open Track

ID	RunID	R	Cr	P	Cp	F	R_{OOV}	P_{OOV}	F_{OOV}	R_{IV}	P_{IV}	F_{IV}
24	a	.9670	.000736	.9725	.000674	.9697	.8988	.8525	.8750	.9731	.9839	.9785
24	b	.9657	.000750	.9715	.000685	.9686	.8963	.8411	.8678	.9719	.9841	.9780
39		.9181	.001129	.9024	.001222	.9102	.6656	.5843	.6223	.9407	.9346	.9377
28		.8860	.001309	.9349	.001016	.9098	.6595	.5657	.6090	.9063	.9764	.9401
3		.0445	.000862	.0446	.000863	.0446	.0226	.0229	.0227	.0465	.0466	.0465

[a] WT: word type.
[b] TTR: type-token ratio = type count / token count.
[c] OOV: number of OOV.
[d] $\mathcal{R}_{OOV}$: OOV Rate

Table 8: CKIP: Word Segmentation: Closed Track

ID	RunID	R	Cr	P	Cp	F	R_{OOV}	P_{OOV}	F_{OOV}	R_{IV}	P_{IV}	F_{IV}
2		.9501	.001445	.9440	.001527	.9470	.7404	.7649	.7524	.9669	.9577	.9623
26		.9497	.001451	.9361	.001624	.9429	.6556	.7481	.6988	.9732	.9490	.9610
5		.9455	.001507	.9371	.001612	.9413	.7004	.7373	.7184	.9651	.9521	.9586
28		.9383	.001597	.9396	.001582	.9390	.6962	.6780	.6870	.9577	.9612	.9594
19		.9432	.001536	.9333	.001657	.9383	.6882	.6885	.6883	.9637	.9527	.9581
8	a	.9412	.001562	.9345	.001643	.9378	.7228	.6688	.6948	.9586	.9575	.9580
18		.9369	.001615	.9270	.001727	.9319	.6636	.6624	.6630	.9587	.9480	.9533
24	a	.9345	.001643	.9289	.001707	.9317	.7124	.6602	.6853	.9522	.9521	.9522
24	b	.9336	.001653	.9277	.001720	.9306	.7091	.6589	.6831	.9515	.9508	.9512
27		.9354	.001632	.9173	.001828	.9263	.5521	.6877	.6125	.9661	.9316	.9485
8	b	.9247	.001753	.9162	.001840	.9204	.6859	.5896	.6341	.9438	.9467	.9452
33		.9241	.001758	.9165	.001836	.9203	.6746	.6195	.6459	.9441	.9424	.9432
7	c	.9233	.001767	.9161	.001841	.9197	.6801	.5846	.6287	.9428	.9471	.9449
7	a	.9233	.001767	.9162	.001840	.9197	.6801	.5849	.6289	.9428	.9471	.9450
7	d	.9224	.001777	.9153	.001849	.9188	.6672	.5732	.6166	.9428	.9473	.9450
15		.9150	.001852	.9001	.001991	.9075	.4751	.5689	.5178	.9502	.9216	.9356
21	b	.9074	.001925	.8897	.002080	.8985	.4405	.5020	.4692	.9447	.9161	.9302
21	a	.9076	.001923	.8896	.002081	.8985	.4406	.5028	.4697	.9449	.9159	.9302
7	b	.8588	.002312	.8850	.002118	.8717	.6204	.4183	.4997	.8779	.9447	.9101

Table 9: CKIP: Word Segmentation: Open Track

ID	RunID	R	Cr	P	Cp	F	R_{OOV}	P_{OOV}	F_{OOV}	R_{IV}	P_{IV}	F_{IV}
5		.9586	.001323	.9541	.001389	.9563	.7804	.8050	.7925	.9728	.9656	.9692
28		.9507	.001438	.9503	.001443	.9505	.7391	.7704	.7544	.9676	.964	.9658
24	b	.9367	.001616	.9360	.001625	.9364	.7527	.6911	.7206	.9515	.9575	.9545
24	a	.9324	.001667	.9326	.001665	.9325	.7459	.6631	.7021	.9473	.9571	.9522
39		.9218	.001782	.8960	.002027	.9087	.6454	.5901	.6165	.944	.9221	.9329
3		.3977	.003245	.3944	.003240	.3961	.3405	.3359	.3382	.4025	.3994	.4009

Table 10: CTB: Word Segmentation: Closed Track

ID	RunID	R	Cr	P	Cp	F	R_{OOV}	P_{OOV}	F_{OOV}	R_{IV}	P_{IV}	F_{IV}
2		.9583	.001408	.9596	.001386	.9589	.7730	.7761	.7745	.9691	.9704	.9697
26		.9538	.001477	.9527	.001493	.9533	.7031	.7491	.7254	.9685	.9639	.9662
31	b	.9505	.001527	.9528	.001492	.9517	.7580	.6886	.7216	.9618	.9701	.9659
31	a	.9503	.001529	.9520	.001505	.9512	.7540	.6845	.7176	.9619	.9694	.9656
27		.9494	.001543	.9508	.001522	.9501	.7208	.7012	.7108	.9628	.9659	.9644
18		.9487	.001553	.9514	.001513	.9500	.7507	.6753	.7110	.9603	.9696	.9650
8	b	.9482	.001560	.9516	.001511	.9499	.7596	.6740	.7142	.9592	.9702	.9647
8	a	.9481	.001561	.9514	.001513	.9498	.7614	.6742	.7152	.9591	.9700	.9645
31	d	.9487	.001552	.9509	.001520	.9498	.7583	.6812	.7177	.9599	.9687	.9643
9		.9471	.001575	.9500	.001533	.9486	.7670	.6736	.7173	.9577	.9688	.9632
24	a	.9451	.001603	.9521	.001503	.9486	.7694	.6714	.7171	.9555	.9713	.9633
31	c	.9495	.001542	.9474	.001571	.9485	.6638	.7456	.7023	.9663	.9579	.9621
28		.9429	.001633	.9535	.001481	.9482	.7536	.6661	.7072	.954	.9730	.9634
24	b	.9456	.001596	.9492	.001545	.9474	.7565	.6613	.7057	.9567	.9688	.9627
5		.9434	.001626	.9459	.001592	.9447	.6911	.6883	.6897	.9582	.9612	.9597
37		.9459	.001592	.9418	.001648	.9439	.6589	.6698	.6643	.9628	.9574	.9601
33		.9402	.001669	.9433	.001628	.9417	.7317	.6517	.6894	.9524	.9628	.9576
7	c	.9350	.001736	.9378	.001700	.9364	.7132	.5796	.6395	.9480	.9641	.9560
7	a	.9350	.001735	.9379	.001699	.9364	.7132	.5800	.6397	.9480	.9642	.9560
7	d	.9342	.001745	.9366	.001715	.9354	.6998	.5706	.6286	.9480	.9634	.9556
7	b	.9099	.002015	.9250	.001854	.9174	.6911	.4834	.5689	.9227	.9638	.9428
21	b	.9077	.002037	.9078	.002037	.9077	.4728	.5603	.5128	.9333	.9248	.9290
21	a	.9078	.002037	.9073	.002041	.9075	.4703	.5583	.5105	.9335	.9244	.9289
21	d	.8992	.002119	.9063	.002051	.9027	.5301	.5029	.5161	.9209	.9316	.9262
21	c	.8992	.002119	.9062	.002052	.9027	.5299	.5029	.5160	.9210	.9315	.9262
19		.8773	.002310	.8788	.002297	.8780	.6714	.5886	.6273	.8894	.8985	.8939

Table 11: CTB: Word Segmentation: Open Track

ID	RunID	R	Cr	P	Cp	F	R_{OOV}	P_{OOV}	F_{OOV}	R_{IV}	P_{IV}	F_{IV}
28	a	.9914	.000648	.9926	.000602	.9920	.9685	.9623	.9654	.9928	.9944	.9936
24	a	.9760	.001077	.9826	.000920	.9793	.9420	.8655	.9021	.9780	.9902	.9840
31	a	.9766	.001065	.9721	.001158	.9743	.9089	.8553	.8813	.9805	.9794	.9800
24	b	.9702	.001196	.9753	.001092	.9728	.9145	.8361	.8736	.9735	.9844	.9789
28	b	.9665	.001266	.9738	.001123	.9702	.8821	.8857	.8839	.9715	.9790	.9753
31	b	.9589	.001397	.9612	.001359	.9601	.7922	.7902	.7912	.9687	.9713	.9700
3		.9485	.001556	.9498	.001536	.9491	.7261	.6769	.7006	.9615	.9672	.9643
39		.9461	.001590	.9372	.001707	.9416	.7223	.6764	.6986	.9592	.9535	.9563
8	a	.9370	.001710	.9321	.001770	.9346	.6556	.6139	.6341	.9535	.9521	.9528
8	b	.9270	.001831	.9319	.001773	.9294	.6576	.6099	.6329	.9428	.9525	.9476
22		.9251	.001853	.9261	.001841	.9256	.5967	.7337	.6581	.9444	.9352	.9398
8	c	.9089	.002025	.8346	.002615	.8702	.2011	.3336	.2509	.9505	.8505	.8977

Table 12: NCC: Word Segmentation: Closed Track

ID	RunID	R	Cr	P	Cp	F	R_{OOV}	P_{OOV}	F_{OOV}	R_{IV}	P_{IV}	F_{IV}
2		.9402	.001214	.9407	.001210	.9405	.6179	.5984	.6080	.9562	.9583	.9573
26		.9452	.001166	.9320	.001289	.9386	.4502	.6196	.5215	.9698	.9430	.9562
5		.9365	.001249	.9365	.001249	.9365	.6158	.5542	.5834	.9524	.9577	.9551
34		.9417	.001200	.9272	.001331	.9344	.4001	.6454	.4940	.9687	.9356	.9518
31	b	.9387	.001229	.9301	.001306	.9344	.5561	.5728	.5643	.9577	.9472	.9524
31	a	.9389	.001226	.9298	.001309	.9343	.5556	.5743	.5648	.9580	.9467	.9523
37		.9396	.001220	.9286	.001319	.9341	.5007	.5411	.5201	.9614	.9462	.9537
19		.9328	.001282	.9353	.001260	.9340	.5907	.5218	.5542	.9498	.9588	.9543
31	d	.9307	.001301	.9318	.001292	.9312	.6309	.5222	.5715	.9456	.9566	.9511
31	c	.9380	.001235	.9223	.001371	.9301	.4709	.6247	.5370	.9613	.9331	.947
24	a	.9251	.001348	.9347	.001266	.9299	.6577	.4968	.5660	.9384	.9643	.9512
27		.9300	.001307	.9291	.001314	.9296	.5459	.5138	.5294	.9491	.9511	.9501
24	b	.9246	.001352	.9332	.001279	.9289	.6524	.4932	.5617	.9381	.9629	.9503
28		.9193	.001395	.9378	.001237	.9285	.6516	.4833	.5549	.9326	.9695	.9507
18	b	.9278	.001326	.9250	.001349	.9264	.5529	.4966	.5232	.9464	.9488	.9476
29		.9268	.001334	.9260	.001341	.9264	.6094	.4948	.5462	.9426	.9527	.9476
18	a	.9278	.001326	.9249	.001350	.9263	.5486	.4940	.5199	.9466	.9488	.9477
18	c	.9264	.001338	.9241	.001356	.9253	.5707	.4977	.5317	.9441	.9486	.9463
9		.9236	.001361	.9269	.001333	.9252	.6474	.4941	.5604	.9373	.9556	.9464
7	c	.9086	.001476	.9110	.001459	.9098	.5957	.4080	.4843	.9241	.9485	.9361
7	d	.9071	.001487	.9106	.001461	.9088	.5907	.3987	.4761	.9228	.9494	.9359
21	a	.8997	.001539	.8992	.001542	.8995	.4232	.3710	.3954	.9234	.9294	.9264
21	b	.8995	.001540	.8992	.001542	.8994	.4224	.3702	.3946	.9233	.9295	.9264
7	a	.7804	.002121	.8581	.001788	.8174	.5409	.2134	.3060	.7924	.9561	.8666
7	b	.7747	.002140	.8513	.001823	.8112	.5405	.2014	.2935	.7864	.9568	.8633
33		.3082	.002367	.3073	.002365	.3078	.2217	.1678	.1910	.3125	.3166	.3145

Table 13: NCC: Word Segmentation: Open Track

ID	RunID	R	Cr	P	Cp	F	R_{OOV}	P_{OOV}	F_{OOV}	R_{IV}	P_{IV}	F_{IV}
34		.9735	.000823	.9779	.000753	.9757	.8893	.8867	.8880	.9777	.9824	.9800
22		.9568	.001041	.9616	.000984	.9592	.8264	.8144	.8204	.9633	.9691	.9662
31	b	.9620	.000980	.9496	.001120	.9557	.6337	.7673	.6941	.9783	.9569	.9675
31	a	.9528	.001086	.9478	.001139	.9503	.7109	.7619	.7355	.9648	.9563	.9606
5	a	.9440	.001177	.9517	.001098	.9478	.7305	.6381	.6812	.9547	.9698	.9622
5	b	.9376	.001239	.9521	.001093	.9448	.7826	.6110	.6862	.9453	.9745	.9597
14		.9446	.001171	.9263	.001339	.9354	.4643	.7160	.5633	.9685	.9328	.9503
3		.9324	.001286	.9349	.001263	.9337	.6070	.5296	.5657	.9486	.9583	.9534
28		.9191	.001396	.9380	.001235	.9285	.6543	.4840	.5564	.9323	.9697	.9506
29		.9268	.001334	.9279	.001325	.9273	.6265	.5032	.5581	.9417	.9546	.9481
39		.9323	.001287	.9134	.001440	.9228	.6075	.5820	.5945	.9485	.9303	.9393

Table 14: SXU: Word Segmentation: Closed Track

ID	RunID	R	Cr	P	Cp	F	R_{OOV}	P_{OOV}	F_{OOV}	R_{IV}	P_{IV}	F_{IV}
2		.9622	.001132	.9625	.001127	.9623	.7429	.7159	.7292	.974	.9764	.9752
26		.9623	.001131	.9554	.001225	.9588	.6454	.7022	.6726	.9794	.9678	.9736
28		.9549	.001231	.9611	.001148	.9580	.6626	.6639	.6632	.9707	.9772	.9739
18	b	.9543	.001239	.9568	.001206	.9556	.7273	.6232	.6712	.9666	.9781	.9723
5		.9558	.001219	.9552	.001228	.9555	.6922	.6638	.6777	.9701	.9716	.9708
24	a	.9523	.001264	.9569	.001205	.9546	.7506	.6129	.6748	.9632	.9801	.9716
18	c	.9528	.001258	.9560	.001217	.9544	.7369	.6164	.6713	.9645	.9782	.9713
31	a	.9594	.001171	.9493	.001302	.9543	.6653	.6694	.6674	.9753	.9642	.9697
8	a	.9534	.001250	.9544	.001238	.9539	.7395	.6275	.6789	.9650	.9754	.9702
8	b	.9536	.001248	.9541	.001242	.9538	.7352	.6287	.6778	.9654	.9748	.9701
31	d	.9535	.001249	.9532	.001253	.9533	.7305	.6257	.6741	.9656	.9740	.9698
31	b	.9593	.001173	.9474	.001324	.9533	.6463	.6749	.6603	.9762	.9613	.9687
18	a	.9518	.001270	.9547	.001234	.9533	.7020	.6020	.6481	.9653	.9772	.9712
8	d	.9512	.001278	.9553	.001226	.9532	.7462	.6275	.6817	.9623	.9767	.9694
8	c	.9509	.001282	.9544	.001238	.9526	.7396	.6281	.6793	.9623	.9754	.9688
24	b	.9499	.001295	.9536	.001249	.9517	.7271	.5966	.6554	.9619	.9774	.9696
27		.9514	.001276	.9511	.001279	.9512	.6834	.6202	.6502	.9658	.9709	.9684
9		.9505	.001287	.9515	.001275	.9510	.7326	.6106	.6660	.9623	.9738	.9680
37		.9554	.001224	.9459	.001342	.9507	.6206	.6113	.6159	.9735	.9641	.9688
34		.9558	.001220	.9442	.001362	.9500	.5176	.6966	.5939	.9794	.9539	.9665
31	c	.9558	.001219	.9441	.001363	.9499	.5788	.7154	.6399	.9762	.9539	.9649
33		.9387	.001423	.9392	.001418	.9390	.6741	.5627	.6134	.9530	.9638	.9584
7	a	.9378	.001434	.9390	.001420	.9384	.6731	.5110	.5810	.9520	.9701	.9610
7	b	.9376	.001435	.9391	.001419	.9383	.6729	.5107	.5807	.9519	.9701	.9609
7	c	.9377	.001434	.9389	.001421	.9383	.6731	.5110	.5810	.9520	.9699	.9609
7	d	.9360	.001452	.9369	.001443	.9365	.6550	.4949	.5638	.9512	.9691	.9600
21	b	.9185	.001624	.9107	.001692	.9146	.4898	.4423	.4648	.9416	.9386	.9401
21	a	.9185	.001624	.9106	.001693	.9145	.4886	.4414	.4638	.9417	.9386	.9401
19		.7820	.002450	.7793	.002460	.7807	.4969	.3538	.4133	.7976	.8125	.8050

Table 15: SXU: Word Segmentation: Open Track

ID	RunID	R	Cr	P	Cp	F	R_{OOV}	P_{OOV}	F_{OOV}	R_{IV}	P_{IV}	F_{IV}
31	a	.9768	.000894	.9703	.001007	.9735	.7825	.8415	.8109	.9872	.9767	.9820
31	b	.9738	.000948	.9620	.001134	.9679	.7089	.8040	.7534	.9881	.9694	.9786
28		.9547	.001233	.9622	.001132	.9584	.6705	.6628	.6666	.9701	.9787	.9744
8	a	.9545	.001236	.9572	.001201	.9559	.7543	.6400	.6925	.9654	.9776	.9714
8	b	.9639	.001108	.9479	.001319	.9558	.6103	.7089	.6559	.9829	.9587	.9707
8	c	.9586	.001182	.9467	.001333	.9526	.6126	.6967	.6519	.9773	.9583	.9677
39		.9575	.001197	.9461	.001339	.9518	.7274	.6920	.7093	.9699	.9604	.9652
3		.9516	.001273	.9515	.001275	.9516	.6843	.6174	.6491	.9661	.9716	.9688
22		.8777	.001945	.8705	.001993	.8741	.5621	.6371	.5972	.8947	.8815	.8880

word has only one POS-tag in the training corpus, but has other POS-tags in the testing corpus, it is a multi-tag word.

- OOV tag: If a tag of a word is found in the test corpus, but not in the training corpus, or the word itself is an OOV word, the corresponding word-tag pair is called OOV tag.

- IV tag: if the pair of word and tag does occur in the training corpus, the pair is called IV tag.

- IV multi-tag words: the multi-tag words that occurred in training data.

For each submission, we compute total accuracy (A_{Total}),IV recall (R_{IV}), OOV recall (R_{OOV}), and IV Multi-tag word recall ($R_{MT_{IV}}$) for evaluation. The formula for total accuracy is: $A_{Total} = \frac{N_{correct}}{N_{truth}}$, where $N_{correct}$ denotes the number of words that are correctly tagged, and N_{truth} denotes the number of words in the truth corpus.

The recall for IV, OOV and IV Multi-tag words are supposed to indicate participating system's performance on these three categories.

As Chinese word segmentation task, a baseline and a topline for each corpus are computed to reflect

Table 16: Named Entity Recognition Training and Truth data statistics

	Training				Truth			
Source	NE[a]	PER[b]	LOC[c]	ORG[d]	NE	PER	LOC	ORG
CITYU	66255	16552	36213	13490	13014	4940	4847	3227
MSRA	37811	9028	18522	10261	7707	1864	3658	2185

Table 17: Named Entity Recognition Truth data OOV statistics

	NE		PER		LOC		ORG	
Source	OOV	$\mathcal{R}_{OOV}$[e]	OOV	$\mathcal{R}_{OOV}$	OOV	$\mathcal{R}_{OOV}$	OOV	$\mathcal{R}_{OOV}$
CITYU	6354	0.4882	3878	0.7850	900	0.1857	1576	0.4884
MSRA	1651	0.2142	564	0.3026	315	0.0861	772	0.3533

Table 18: Named Entity Recognition Baseline

Source	R	P	F	R_{PER}	P_{PER}	F_{PER}	R_{LOC}	P_{LOC}	F_{LOC}	R_{ORG}	P_{ORG}	F_{ORG}
CITYU	.4912	.7562	.5955	.2130	.7056	.3272	.7681	.8438	.8042	.5011	.6341	.5598
MSRA	.5451	.6937	.6105	.6459	.9205	.7591	.4513	.7847	.5731	.6160	.5091	.5575

Table 19: CITYU: Named Entity Recognition: Closed Track

ID	RunID	R	P	F	R_{PER}	P_{PER}	F_{PER}	R_{LOC}	P_{LOC}	F_{LOC}	R_{ORG}	P_{ORG}	F_{ORG}
24		.8247	.8768	.8499	.8615	.9240	.8917	.9098	.8612	.8848	.6402	.8221	.7199
2	a	.7556	.8850	.8152	.7688	.9165	.8362	.8659	.8695	.8677	.5699	.8589	.6852
2	b	.7541	.8846	.8142	.7638	.9167	.8333	.8675	.8684	.8680	.5689	.8596	.6847
18	c	.7608	.8751	.8140	.7771	.9143	.8401	.8692	.8551	.8621	.5730	.8451	.6829
28		.7570	.8585	.8046	.7682	.8976	.8279	.8750	.8314	.8526	.5624	.8462	.6757
18	b	.7286	.8933	.8026	.7306	.9254	.8165	.8535	.8789	.8660	.5380	.8650	.6634
18	a	.7277	.8926	.8017	.7287	.9252	.8153	.8529	.8781	.8653	.5380	.8633	.6628
21	a	.0874	.1058	.0957	.0656	.0962	.0780	.1388	.1200	.1288	.0437	.0789	.0562
21	b	.0211	.0326	.0256	.0128	.0218	.0161	.0390	.0433	.0410	.0068	.0192	.0101

Table 20: CITYU: Named Entity Recognition: Open Track

ID	RunID	R	P	F	R_{PER}	P_{PER}	F_{PER}	R_{LOC}	P_{LOC}	F_{LOC}	R_{ORG}	P_{ORG}	F_{ORG}
23		.8743	.9342	.9033	.9526	.9721	.9623	.9342	.9235	.9288	.6644	.8805	.7573
2		.8579	.9179	.8869	.8822	.9449	.9125	.9336	.9099	.9216	.7072	.8852	.7862
28		.8826	.8826	.8826	.9168	.8947	.9056	.9329	.8942	.9132	.7546	.8411	.7955
24		.8975	.8616	.8792	.9474	.9153	.9311	.9389	.8966	.9173	.7589	.7274	.7428
39		.7163	.8000	.7559	.7180	.8194	.7653	.8389	.7845	.8108	.5296	.7986	.6369

Table 21: MSRA: Named Entity Recognition: Closed Track

ID	RunID	R	P	F	R_{PER}	P_{PER}	F_{PER}	R_{LOC}	P_{LOC}	F_{LOC}	R_{ORG}	P_{ORG}	F_{ORG}
24		.9186	.9377	.9281	.9437	.9665	.9549	.9423	.9428	.9426	.8577	.9036	.8800
18	b	.8862	.9304	.9078	.9195	.9651	.9418	.9043	.9379	.9208	.8275	.8871	.8563
2		.8779	.9274	.9020	.9029	.9628	.9319	.9101	.9341	.9219	.8027	.8841	.8414
18	a	.8752	.9255	.8996	.9040	.9618	.9320	.8991	.9346	.9165	.8105	.8780	.8429
28		.8822	.9156	.8986	.9126	.9461	.9290	.9079	.9248	.9163	.8133	.8724	.8418
31		.8058	.9107	.8550	.9029	.9519	.9268	.8185	.9278	.8697	.7016	.8405	.7648
37		.8331	.8730	.8526	.8557	.8084	.8314	.8576	.9138	.8848	.7730	.8666	.8171

[a] NE: Number of Named Entities.
[b] PER: Number of Person names.
[c] LOC: Number of Location names.
[d] ORG: Number of Organization names
[e] $\mathcal{R}_{OOV}$:OOV rate

Table 22: MSRA: Named Entity Recognition: Open Track

ID	RunID	R	P	F	R_{PER}	P_{PER}	F_{PER}	R_{LOC}	P_{LOC}	F_{LOC}	R_{ORG}	P_{ORG}	F_{ORG}
24		.9995	.9982	.9988	1	.9989	.9995	.9997	.9975	.9986	.9986	.9986	.9986
2		.9961	.9956	.9958	1	1	1	.9992	.9929	.9960	.9876	.9963	.9920
1		.9377	.9603	.9489	.9657	.9574	.9615	.9593	.9769	.9680	.8778	.9338	.9049
23		.9111	.9471	.9288	.9458	.9833	.9642	.9336	.9397	.9366	.8439	.9280	.8840
18	a	.9135	.9321	.9227	.9560	.9601	.9581	.9221	.9388	.9304	.8627	.8959	.8790
18	b	.9084	.9278	.9180	.9544	.9575	.9559	.9169	.9322	.9245	.8549	.8938	.8739
22	b	.8675	.9163	.8912	.9217	.9630	.9419	.8445	.9352	.8875	.8600	.8502	.8551
29		.8791	.9035	.8911	.9549	.9498	.9524	.9194	.9129	.9161	.7469	.8408	.7911
11		.8674	.9003	.8836	.9083	.9216	.9149	.8989	.9166	.9077	.7799	.8516	.8141
31		.8238	.9038	.8619	.9206	.9517	.9359	.8362	.9424	.8862	.7204	.7966	.7565
22	a	.8452	.8720	.8584	.8734	.9498	.9100	.8710	.8909	.8808	.7780	.7798	.7789
39		.7890	.8347	.8112	.8771	.9196	.8979	.8365	.8331	.8348	.6343	.7557	.6897

the different degree of difficulty of tagging individual corpora. The algorithm of baseline and topline is briefly described as follows: Baseline indicates the different degree of difficulty of tagging individual corpus.

The baseline of each corpus is calculated by generating a list of words and POS tags from the training corpus, then: 1. tagging those IV words in the testing corpus which have only one POS tag in the list. 2. for those IV words that have not only one tag in training corpus, the unique most frequent tag in training corpus will be assigned to them. 3. for each IV word that does not have a unique most frequent tag in training corpus, one of its tag which is most frequent in the overall phase is assigned to it; 4. for those words that do not fall into any of the former three categories are assigned with a overall most frequent tag.

The topline algorithm is similar to baseline, instead the list of words and POS tags is generated from testing corpus.

Chinese POS tagging results for all runs grouped by corpus and track appear in Tables 27-36; all tables are sorted by A_{Total}.

The baseline and topline has shown that, with preliminary knowledge and mechanical algorithm, it is easy to achieve an accuracy over approximately 0.85. When excluding the effect caused by OOV tags, the accuracy can even be over 0.93.

There are two kind of problem in POS tagging task we should cope with: multi tag disambiguation and unknown words guessing. We could consider that the value of (topline - baseline) is the accuracy drop caused by unknown words guessing, and the value of (1 - topline) is the accuracy drop caused by multi tag disambiguation. The average of these two value is 0.0628 and 0.0600, therefore these two kind of problem can equally affect the performance of POS tagging system.

For this reason, unlike the topline of Chinese word segmentation, the topline of Chinese POS tagging could be easily exceeded by tagging systems, because the algorithm of this topline just excludes the effect of OOV tags, which is not a dominant determinant in this task.

In closed track, the highest total accuracy is achieved in the NCC corpus which has the lowest OOV tag rate, and the lowest total accuracy is achieved in the CITYU corpus which has the highest OOV tag rate.

Most of the participants outperformed baseline, some have exceeded topline. When comparing the OOV recall and IV multi tag word recall with topline, participant's system can easily approaching or surpass the IV multi tag word recall, but none system could successfully approach the OOV recall. This might because participant's systems do better in solving the multi tag disambiguation problem than in coping with the unknown words guessing problem.

5 Conclusions & Future Directions

The Fourth SIGHAN Chinese Language Processing Bakeoff successfully brought together a collection of 28 strong research groups to assess the progress of research in three important tasks, Chinese word

segmentation, named entity recognition and Chinese POS tagging, that in turn enable other important language processing technologies. The individual group presentations at the SIGHAN workshop will detail the approaches that yielded strong performance for both tasks. Issues of out-of-vocabulary word handling, annotation consistency and unknown guessing all continue to challenge system designers and bakeoff organizers alike.

In future analysis, we hope to develop additional analysis tools to better assess progress in these fundamental tasks, in a more corpus independent fashion. Such developments will guide the planning of future evaluations.

Finally, while Chinese word segmentation, named entity recognition and Chinese POS tagging are important in themselves, these three enabling technologies are also the foundation of those upper level applications such as parsing, reference resolution or machine translation. To evaluate the impact of improvement in these three technologies on the subsequent applications is still the future work for this evaluation.

6 Acknowledgements

We gratefully acknowledge the generous assistance of the organizations listed below who provided the data for this bakeoff; without their support, it could not have taken place:

- City University of Hong Kong, Hong Kong;
- Chinese Knowledge Information Processing Group, Academia Sinica, Taiwan;
- Institute of Applied linguistics, M.O.E., China;
- Microsoft Research Asia, China;
- Peking University, China;
- Shanxi University, China;
- University of Colorado, USA;

We also thank Olivia Oi Yee Kwong, the co-organizers of the sixth SIGHAN workshop, in conjunction with which this bakeoff takes place, and Yongsheng Guo from Institute of Applied Linguistics, M.O.E., P.R.C. who has made great effort for this bakeoff.

Professor Changning Huang merits special thanks for his help in this bakeoff. Finally, we thank all the participating sites who enabled the success of this bakeoff.

Table 25: Chinese POS tagging Baseline

Source	A_{Total}[a]	R_{IV}[b]	R_{OOV}[c]	$R_{MT_{IV}}$[d]
CITYU	.8425	.9021	.2543	.8083
CKIP	.8861	.9451	.2814	.8740
CTB	.8609	.8967	.3313	.8057
NCC	.9159	.9543	.2242	.8636
PKU	.8809	.9237	.2038	.8296

Table 26: Chinese POS tagging Topline

Source	A_{Total}	R_{IV}	R_{OOV}	$R_{MT_{IV}}$
CITYU	.9310	.9330	.9107	.8727
CKIP	.9606	.9597	.9699	.9103
CTB	.9147	.9120	.9555	.8369
NCC	.9588	.9593	.9507	.8822
PKU	.9351	.9354	.9305	.8600

Table 27: CITYU:POS tagging Closed Track

ID	RunID	A_{Total}	R_{IV}	R_{OOV}	$R_{MT_{IV}}$
30	b	.8951	.9389	.4637	.8745
30	a	.8929	.9367	.4608	.8705
28		.8905	.9328	.4733	.8687
9		.8865	.9326	.4322	.8707
19		.8693	.9284	.2868	.8585
24		.8564	.9149	.2805	.8506
21	b	.2793	.2969	.1051	.2538
21	a	.1890	.2031	.0550	.1704

Table 28: CITYU:POS tagging Open Track

ID	RunID	A_{Total}	R_{IV}	R_{OOV}	$R_{MT_{IV}}$
28		.8900	.9329	.4670	.8695
39		.8669	.9089	.4537	.8495

Table 29: CKIP:POS tagging Closed Track

ID	RunID	A_{Total}	R_{IV}	R_{OOV}	$R_{MT_{IV}}$
30	b	.9295	.9629	.5869	.9123
30	a	.9286	.9618	.5875	.9099
28		.9220	.9556	.5772	.9088
9		.9160	.9504	.5631	.9065
16		.9124	.9549	.4756	.8953
19		.8994	.9561	.3169	.9001
24		.8793	.9334	.3247	.8943

[a] A_{Total}: total accuracy
[b] R_{IV}: IV recall
[c] R_{OOV}: OOV recall
[d] $R_{MT_{IV}}$: MT_{IV} recall

Table 23: Chinese POS tagging Training data statistics

Source	Token	WT	TT[a]	ATN[b]	MT_{IV}[c]	$\mathcal{R}_{MT_{IV}}$[d]
CITYU	1092687	43639	44	1.2588	585056	0.5354
CKIP	721551	48045	60	1.0851	335017	0.4643
CTB	642246	42133	37	1.1690	334317	0.5205
NCC	535023	45108	60	1.0673	178078	0.3328
PKU	1116754	55178	103	1.1194	490243	0.4390

Table 24: Chinese POS tagging Truth data statistics

Source	Token	WT	TT	ATN	OOV	$\mathcal{R}_{OOV}$[e]	MT_{IV}	$\mathcal{R}_{MT_{IV}}$
CITYU	184314	17827	43	1.1446	16977	0.0921	92934	0.5042
CKIP	91071	15331	63	1.0530	8085	0.0888	38640	0.4243
CTB	59955	9797	35	1.1227	3794	0.0633	30513	0.5089
NCC	102344	17493	55	1.0675	5392	0.0527	33853	0.3308
PKU	156407	17643	103	1.1270	9295	0.0594	68065	0.4352

[a] TT: number of tag type.
[b] ATN: Average Tag Number per word.
[c] MT_{IV}: number of IV Multi-Tag word
[d] $\mathcal{R}_{MT_{IV}}$: coverage rate of IV Multi-Tag words
[e] $\mathcal{R}_{OOV}$: OOV tag rate

Table 30: CKIP:POS tagging Open Track

ID	RunID	A_{Total}	R_{IV}	R_{OOV}	$R_{MT_{IV}}$
28		.9211	.9542	.5813	.9082
39		.9004	.9327	.5686	.8936

Table 31: CTB:POS tagging Closed Track

ID	RunID	A_{Total}	R_{IV}	R_{OOV}	$R_{MT_{IV}}$
28		.9428	.9557	.7522	.9197
9		.9401	.9554	.7135	.9183
16		.9234	.9507	.5200	.9051
24		.9203	.9460	.5390	.9055
19		.9133	.9438	.4620	.8983
31	a	.9088	.9374	.4866	.8805
31	b	.8065	.8608	.0040	.7395

Table 32: CTB:POS tagging Open Track

ID	RunID	A_{Total}	R_{IV}	R_{OOV}	$R_{MT_{IV}}$
22		.9689	.9767	.8537	.9554
28		.9646	.9714	.8648	.9495
39		.9271	.9400	.7354	.9016
31	a	.9120	.9374	.5361	.8805
31	b	.8076	.8608	.0206	.7396

Table 33: NCC:POS tagging Closed Track

ID	RunID	A_{Total}	R_{IV}	R_{OOV}	$R_{MT_{IV}}$
30	b	.9541	.9738	.5998	.9195
30	a	.9525	.9717	.6059	.9135
28		.9494	.9690	.5959	.9129
9		.9456	.9658	.5822	.9116
16		.9395	.9690	.4086	.9059
19		.9336	.9687	.3017	.9050
31	a	.9313	.9604	.4080	.8809
29		.9277	.9664	.2329	.9000
24		.9172	.9498	.3312	.8963
31	b	.8940	.9303	.2411	.7948

Table 34: NCC:POS tagging Open Track

ID	RunID	A_{Total}	R_{IV}	R_{OOV}	$R_{MT_{IV}}$
28		.9496	.9694	.5938	.9141
31	a	.9326	.9604	.4336	.8809
39		.9280	.9477	.5749	.8954
22		.9096	.9377	.4045	.8935
31	b	.8940	.9303	.2411	.7948
25		.0836	.0855	.0488	.0645

Table 35: PKU:POS tagging Closed Track

ID	RunID	A_{Total}	R_{IV}	R_{OOV}	$R_{MT_{IV}}$
30	b	.9450	.9679	.5818	.9252
30	a	.9420	.9648	.5813	.9184
28		.9396	.9608	.6036	.9173
9		.9368	.9591	.5832	.9173
16		.9266	.9574	.4386	.9079
29		.9113	.9518	.2708	.8958
37		.9065	.9269	.5836	.8903
31	a	.9053	.9451	.2751	.8758
19		.8815	.9158	.3386	.8897
31	b	.8527	.8936	.2043	.7646
31	c	.8450	.8855	.2039	.7471

Table 36: PKU:POS tagging Open Track

ID	RunID	A_{Total}	R_{IV}	R_{OOV}	$R_{MT_{IV}}$
28		.9411	.9622	.6057	.9200
31	a	.9329	.9518	.6332	.8972
29		.9197	.9512	.4222	.8990
39		.9134	.9341	.5862	.8894
31	b	.8427	.8935	.0398	.7643
22		.6649	.6796	.4308	.6495

References

Thomas Emerson. 2005. The second international Chinese word segmentation bakeoff. In *Proceedings of the Fourth SIGHAN Workshop on Chinese Language Processing*, pages 123–133, Jeju Island, Korea.

Charles Grinstead and J. Laurie Snell. 1997. *Introduction to Probability*. American Mathematical Society, Providence, RI.

Gina-Anne Levow. 2006. The third international chinese language processing bakeoff: Word segmentation and named entity recognition. In *Proceedings of the Fifth SIGHAN Workshop on Chinese Language Processing*, pages 108–117, Sydney, Australia, July. Association for Computational Linguistics.

Richard Sproat and Thomas Emerson. 2003. The first international Chinese word segmentation bakeoff. In *The Second SIGHAN Workshop on Chinese Language Processing*, pages 133–143, Sapporo, Japan.