

Tamkang University

淡江大學



IMTKU Textual Entailment System for Recognizing Inference in Text at NTCIR-9 RITE

Demo

<http://rite.im.tku.edu.tw>

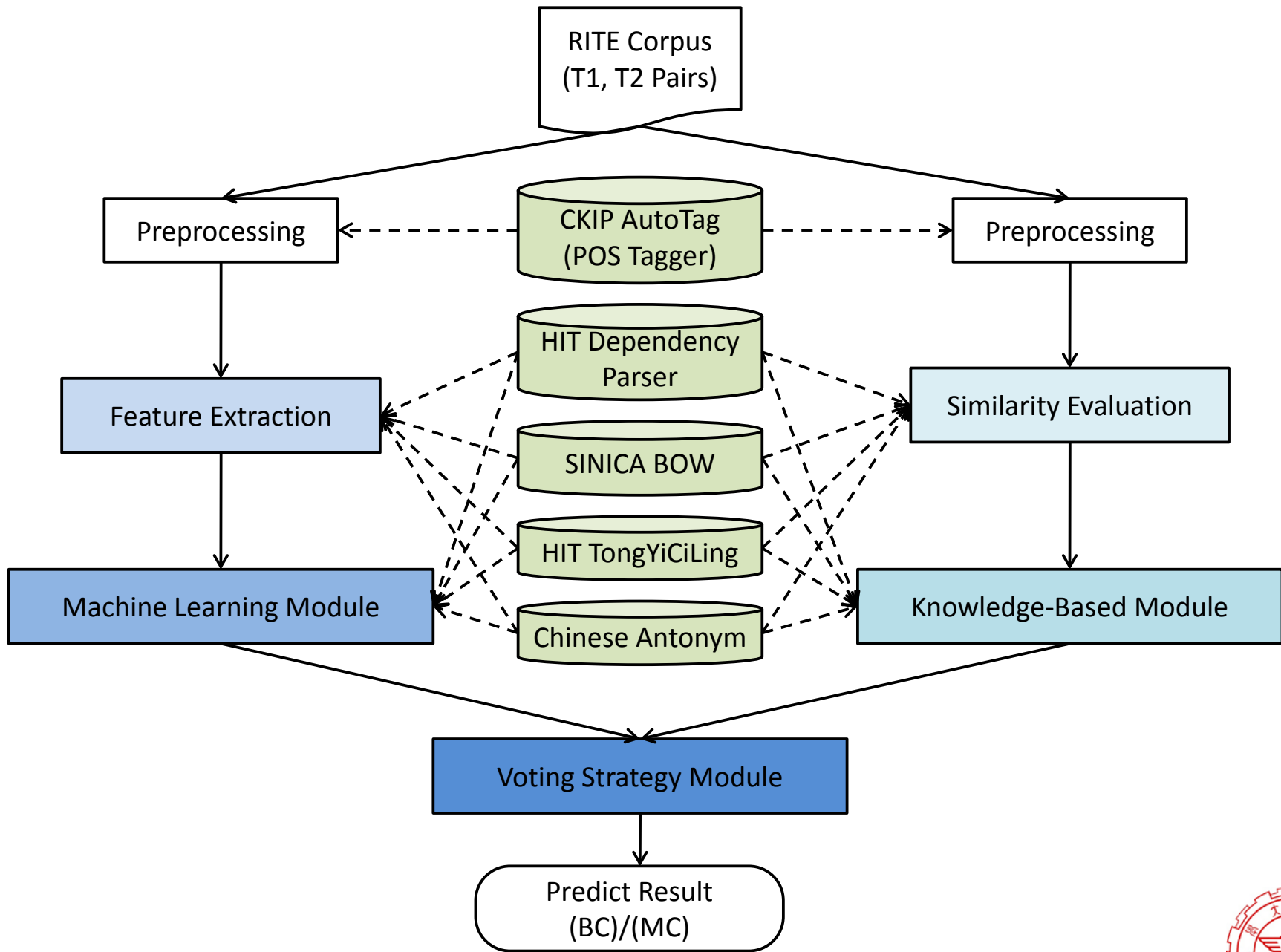


Min-Yuh Day *, Chun Tu
Tamkang University, Taiwan

myday@mail.tku.edu.tw

2011.12.08

NTCIR-9 Workshop, December 6-9, 2011, Tokyo, Japan





IMTKU Textual Entailment System

Department of Information Management, Tamkang University



- DEMO
- LINKS
- ABOUT US

Demo

- BC
- MC
- Rite4QA

-Example: ● Sample 1 ● Sample 2 ● Sample 3

-Textual (T1):

-Hypothesis (T2):

predict

Result:

Detail:



IMTKU Textual Entailment System

Department of Information Management, Tamkang University



- DEMO
- LINKS
- ABOUT US

Demo

BC

MC

Rite4QA

-Example: Sample 1 Sample 2 Sample 3

-Textual (T1):

一九九七年香港回歸中國

-Hypothesis (T2):

香港的主權和領土是在一九九七由英國歸還給中國的。

predict

Result:

Detail:

Empty text area for displaying the result details.



IMTKU Textual Entailment System

Department of Information Management, Tamkang University



- DEMO
- LINKS
- ABOUT US

Demo

BC

MC

Rite4QA

-Example: ● Sample 1 ● Sample 2 ● Sample 3

-Textual (T1):

一九九七年香港回歸中國

-Hypothesis (T2):

香港的主權和領土是在一九九七由英國歸還給中國的。

predict

Result: No 0.711081

Detail:

Longest Common subsequence	:7
T1 Token Length	:4
T2 Token Length	:15
Token Length Ratio	:0.2666666666666667
Token Length Difference	:-11
Word-Based Edit Distance	:18
Token-Based Edit Distance	:14
Difference of (T1-T2)Noun	:-2
Difference of (T1-T2)Verb	:-2



IMTKU Textual Entailment System

Department of Information Management, Tamkang University



- DEMO
- LINKS
- ABOUT US

Demo

BC

MC

Rite4QA

-Example: Sample 1 Sample 2 Sample 3

-Textual (T1):

一九九七年香港回歸中國

-Hypothesis (T2):

香港的主權和領土是在一九九七由英國歸還給中國的。

predict

Result: No 0.711081

Detail:

```

Token-Based Edit Distance      :14
Difference of (T1-T2)Noun      :-2
Difference of (T1-T2)Verb      :-2
T1: 一九九七年香港回歸中國
T2: 香港的主權和領土是在一九九七由英國歸還給中國的。
T1 CKIP: 一九九七年(N) 香港(N) 回歸(Vt) 中國(N)
T2 CKIP: 香港(N) 的(T) 主權(N) 和(C) 領土(N) 是
(Vt) 在(P) 一九九七(DET) 由(P) 英國(N) 歸還(Vt) 給
(Vt) 中國(N) 的(T) 。(PERIODCATEGORY)

```



IMTKU Textual Entailment System

Department of Information Management, Tamkang University



- DEMO
- LINKS
- ABOUT US

Demo

BC

MC

Rite4QA

-Example: ● Sample 1 ● Sample 2 ● Sample 3

-Textual (T1):

阿富汗的政權塔利班

-Hypothesis (T2):

塔利班在阿富汗建立全國性政權，正式名稱為阿富汗伊斯蘭酋長國

predict

Result: No 0.780922

Detail:

Longest Common subsequence	:5
T1 Token Length	:4
T2 Token Length	:13
Token Length Ratio	:0.307692307692308
Token Length Difference	:-9
Word-Based Edit Distance	:24
Token-Based Edit Distance	:10
Difference of (T1-T2)Noun	:-5
Difference of (T1-T2)Verb	:-2



IMTKU Textual Entailment System

Department of Information Management, Tamkang University



- DEMO
- LINKS
- ABOUT US

Demo

BC

MC

Rite4QA

-Example: ● Sample 1 ● Sample 2 ● Sample 3

-Textual (T1):

阿富汗的政權塔利班

-Hypothesis (T2):

塔利班在阿富汗建立全國性政權，正式名稱為阿富汗伊斯蘭酋長國

predict

Result: No 0.780922

Detail:

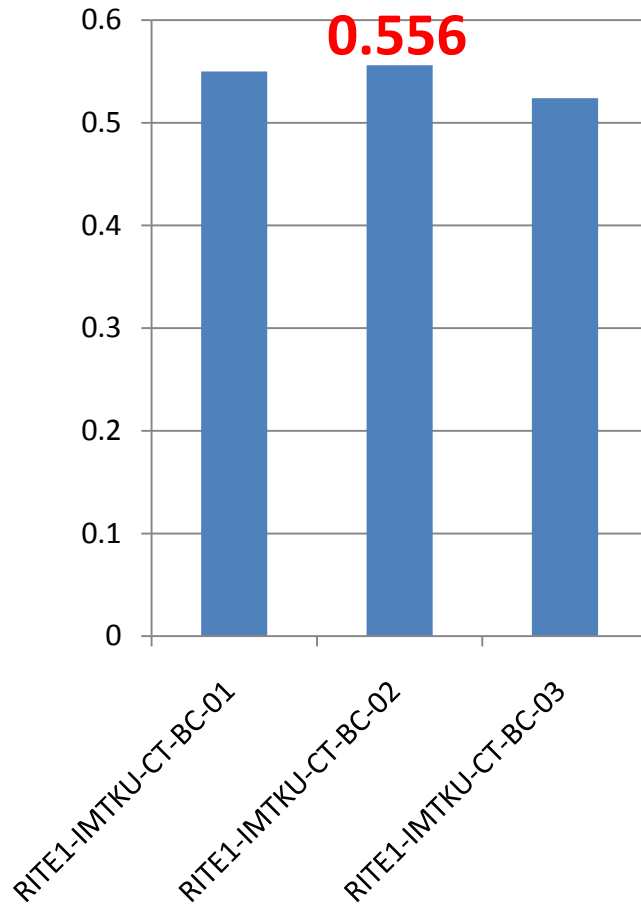
```

Token-based Edit Distance      :10
Difference of (T1-T2)Noun      : -5
Difference of (T1-T2)Verb      : -2
T1: 阿富汗的政權塔利班
T2: 塔利班在阿富汗建立全國性政權，正式名稱為阿富汗伊斯蘭酋長國
T1 CKIP:  阿富汗(N) 的(T) 政權(N) 塔利班(N)
T2 CKIP:  塔利班(N) 在(P) 阿富汗(N) 建立(Vt) 全國性(N) 政權(N) ，(COMMACATEGORY) 正式(Vi) 名稱(N) 為(P) 阿富汗(N) 伊斯蘭(N) 酋長國(N)

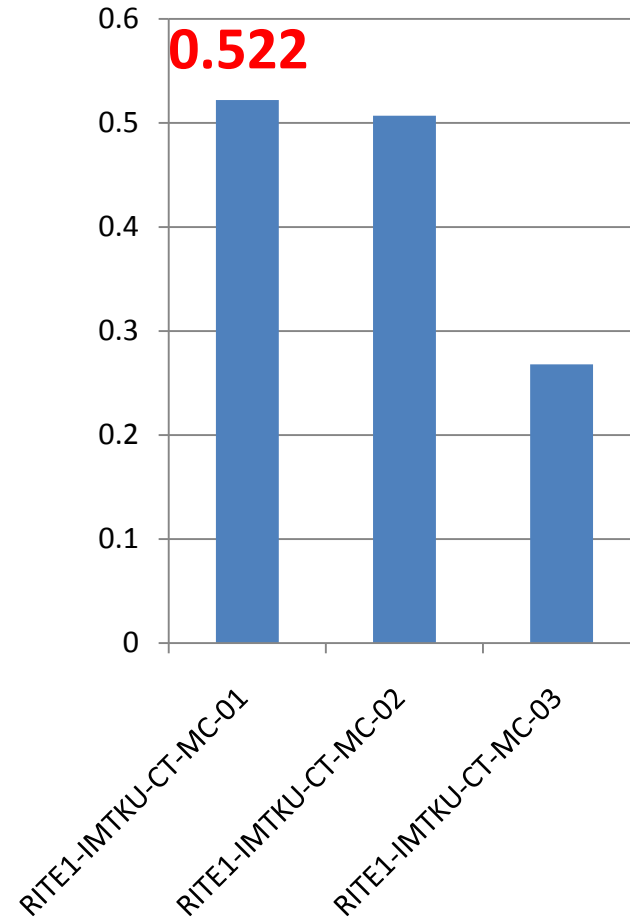
```


IMTKU at NTCIR-9 RITE Task Performance

CT-BC Subtask
Accuracy

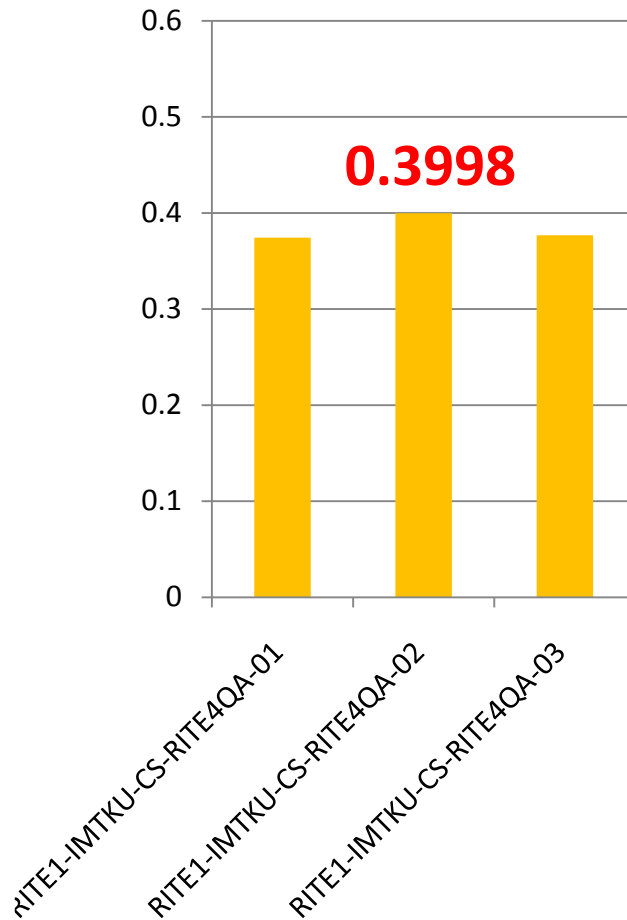


CT-MC Subtask
Accuracy



IMTKU at NTCIR-9 RITE Task Performance

CS-RITE4QA Subtask
MRR



CT-RITE4QA Subtask
MRR

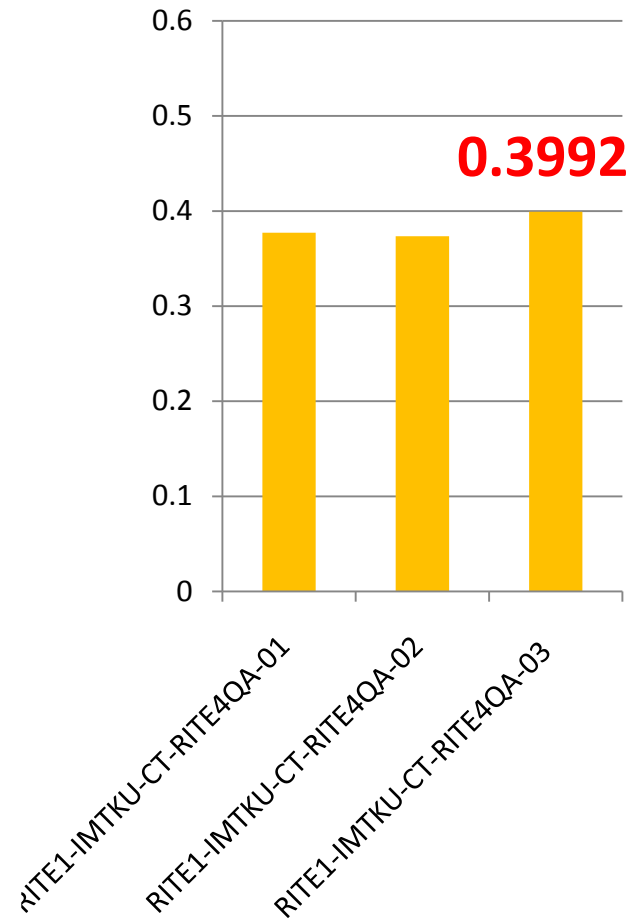


Table 9. Evaluation result on BC subtask (CS).

Run	Accuracy
UIOWA-CS-BC-01	0.9705
UIOWA-CS-BC-03	0.9631
UIOWA-CS-BC-02	0.9361
ICRC_HITSZ-CS-BC-03	0.7764
FudanNLP-CS-BC-02	0.7617
ICRC_HITSZ-CS-BC-02	0.7568
FudanNLP-CS-BC-01	0.7469
WHUTE-CS-BC-03	0.7371
NTU-CS-BC-01	0.7346
WHUTE-CS-BC-02	0.7322
WUST-CS-BC-01	0.7248
NTU-CS-BC-02	0.7224
NTU-CS-BC-03	0.7199
ZSWSL-CS-BC-01	0.7199
IASLD-CS-BC-01*	0.7150
ICL-CS-BC-01	0.7150
WHUTE-CS-BC-01	0.7125
ICL-CS-BC-02	0.7101
ICRC_HITSZ-CS-BC-01	0.7076
IASLD-CS-BC-02*	0.7052
IASLD-CS-BC-03*	0.6880
III_CYUT_NTHU-CS-BC-02	0.6830
NSNG-CS-BC-02	0.6683
ZSWSL-CS-BC-02	0.6658
NSNG-CS-BC-01	0.6536
Yuntech-CS-BC-01	0.6364
NSNG-CS-BC-03	0.5897
ZSWSL-CS-BC-03	0.5897
Yuntech-CS-BC-02	0.5602
III_CYUT_NTHU-CS-BC-01	0.5577
III_CYUT_NTHU-CS-BC-03	0.5577
<i>Baseline (char overlap)</i>	<i>0.7617</i>

Table 10. Evaluation result on BC subtask (CT).

Run	Accuracy
UIOWA-CT-BC-01	0.9078
UIOWA-CT-BC-02	0.8844
IASLD-CT-BC-03	0.6611
IASLD-CT-BC-02	0.6533
III_CYUT_NTHU-CT-BC-02	0.6500
IASLD-CT-BC-01	0.6478
NTOUA-CT-BC-02*	0.6422
ICRC_HITSZ-CT-BC-01	0.6133
NTOUA-CT-BC-01*	0.6133
NTU-CT-BC-01	0.6078
NTU-CT-BC-03	0.6067
NTOUA-CT-BC-03*	0.6022
ICRC_HITSZ-CT-BC-02	0.5967
NTU-CT-BC-02	0.5956
III_CYUT_NTHU-CT-BC-01	0.5733
III_CYUT_NTHU-CT-BC-03	0.5733
IMTKU-CT-BC-02	0.5556
MCU-CT-BC-01	0.5544
IMTKU-CT-BC-01	0.5500
Yuntech-CT-BC-01	0.5278
IMTKU-CT-BC-03	0.5244
Yuntech-CT-BC-02	0.5244
<i>Baseline (char overlap)</i>	<i>0.6667</i>

Table 12. Evaluation result on MC subtask (CS).

Run	Accuracy
UIOWA-CS-MC-01	0.8919
UIOWA-CS-MC-02	0.8919
UIOWA-CS-MC-03	0.8870
ICRC_HITSZ-CS-MC-03	0.6413
ICRC_HITSZ-CS-MC-02	0.6241
ZSWSL-CS-MC-02	0.6192
WHUTE-CS-MC-02	0.6093
III_CYUT_NTHU-CS-MC-02	0.5897
FudanNLP-CS-MC-02	0.5848
WHUTE-CS-MC-01	0.5823
WUST-CS-MC-01	0.5823
FudanNLP-CS-MC-01	0.5799
ICRC_HITSZ-CS-MC-01	0.5749
NTU-CS-MC-02	0.5749
NTU-CS-MC-03	0.5700
IASLD-CS-MC-01*	0.5651
NTU-CS-MC-01	0.5651
ZSWSL-CS-MC-03	0.5627
IASLD-CS-MC-03*	0.5553
ZSWSL-CS-MC-01	0.5455
IASLD-CS-MC-02*	0.5430
III_CYUT_NTHU-CS-MC-01	0.5332
III_CYUT_NTHU-CS-MC-03	0.5307
Yuntech-CS-MC-01	0.5283
ICL-CS-MC-01	0.5061
ICL-CS-MC-02	0.4840
Yuntech-CS-MC-02	0.3980
<i>Baseline (char overlap)</i>	<i>0.5315</i>

Table 13. Evaluation result on MC subtask (CT).

Run	Accuracy
UIOWA-CT-MC-01	0.7867
UIOWA-CT-MC-02	0.7744
UIOWA-CT-MC-03	0.7244
MCU-CT-MC-01	0.5356
IMTKU-CT-MC-01	0.5222
IMTKU-CT-MC-02	0.5067
IASLD-CT-MC-03	0.5011
IASLD-CT-MC-01	0.4989
ICRC_HITSZ-CT-MC-01	0.4967
III_CYUT_NTHU-CT-MC-02	0.4911
IASLD-CT-MC-02	0.4867
NTU-CT-MC-03	0.4833
Yuntech-CT-MC-01	0.4767
NTOUA-CT-MC-02*	0.4611
NTU-CT-MC-01	0.4589
NTU-CT-MC-02	0.4578
NTOUA-CT-MC-01*	0.4400
III_CYUT_NTHU-CT-MC-03	0.4333
III_CYUT_NTHU-CT-MC-01	0.4300
NTOUA-CT-MC-03*	0.4211
Yuntech-CT-MC-02	0.3878
IMTKU-CT-MC-03	0.2678
<i>Baseline (char overlap)</i>	<i>0.4885</i>

**Table 16. Evaluation result on RITE4QA subtask (CS).
See the table blow for the baseline scores.**

Run	Accuracy	Top1	MRR
UIOWA-CS-RITE4QA-01	0.9010	0.4559	0.4272
IMTKU-CS-RITE4QA-02	0.4090	0.2953	0.3998
WHUTE-CS-RITE4QA-02	0.4876	0.2852	0.3979
WHUTE-CS-RITE4QA-01	0.3886	0.2651	0.3773
IMTKU-CS-RITE4QA-03	0.4716	0.2550	0.3768
IMTKU-CS-RITE4QA-01	0.3319	0.2450	0.3744
ICL-CS-RITE4QA-01	0.3231	0.2931	0.3545
ICRC_HITSZ-CS-RITE4QA-01	0.6390	0.2479	0.3520
WHUTE-CS-RITE4QA-03	0.3275	0.2248	0.3494
ICRC_HITSZ-CS-RITE4QA-03	0.7293	0.2262	0.3398
IASLD-CS-RITE4QA-01*	0.4833	0.2274	0.3028
IASLD-CS-RITE4QA-02*	0.4803	0.2274	0.3028
III_CYUT_NTHU-CS-RITE4QA-01	0.7525	0.2585	0.2944
III_CYUT_NTHU-CS-RITE4QA-02	0.7162	0.2408	0.2908
ICRC_HITSZ-CS-RITE4QA-02	0.6128	0.2234	0.2705
IASLD-CS-RITE4QA-03*	0.4352	0.2310	0.2608
III_CYUT_NTHU-CS-RITE4QA-03	0.3377	0.2320	0.2527

Table 17. Evaluation result on RITE4QA subtask (CT).

Run	Accuracy	Top1	MRR
UIOWA-CT-RITE4QA-01	0.9010	0.4559	0.4272
IMTKU-CT-RITE4QA-03	0.4003	0.2953	0.3992
NTOUA-CT-RITE4QA-03*	0.6346	0.2813	0.3824
NTOUA-CT-RITE4QA-01*	0.5459	0.2746	0.3803
IMTKU-CT-RITE4QA-01	0.3246	0.2517	0.3772
IMTKU-CT-RITE4QA-02	0.3392	0.2517	0.3736
NTOUA-CT-RITE4QA-02*	0.5124	0.2282	0.3572
ICRC_HITSZ-CT-RITE4QA-01	0.6390	0.2479	0.3520
ICRC_HITSZ-CT-RITE4QA-03	0.7293	0.2262	0.3398
IASLD-CT-RITE4QA-01*	0.4760	0.2274	0.3016
IASLD-CT-RITE4QA-02*	0.4731	0.2274	0.3016
III_CYUT_NTHU-CT-RITE4QA-01	0.7525	0.2598	0.2947
III_CYUT_NTHU-CT-RITE4QA-02	0.7147	0.2408	0.2908
ICRC_HITSZ-CT-RITE4QA-02	0.6128	0.2234	0.2705
IASLD-CT-RITE4QA-03*	0.4279	0.2290	0.2619
III_CYUT_NTHU-CT-RITE4QA-03	0.3392	0.2320	0.2527
<i>Baseline1 (char overlap)</i>	<i>0.2317</i>	<i>0.2317</i>	<i>0.3844</i>
<i>Baseline2 (all yes)</i>	<i>0.1906</i>	<i>0.2243</i>	<i>0.2378</i>
<i>Baseline3 (random)</i>	<i>0.5000</i>	<i>0.2109</i>	<i>0.3454</i>
<i>Baseline4 (QA system)</i>	<i>0.1906</i>	<i>0.4200</i>	<i>0.4852</i>
<i>Oracle</i>	<i>1.0000</i>	<i>0.5906</i>	<i>0.5906</i>

IMTKU Experiments for NTCIR-9 RITE Datasets

Datasets	10 Fold CV Accuracy
RITE1_CT_dev_bc_g.txt (gold standard) (BC Development Dataset: 421 pairs)	76.48%
RITE1_CT_test_bc_g.txt (BC Test Dataset: 900 pairs)	66.33%
RITE1_CT_dev_test_bc_g.txt (BC Dev+Test Dataset: 421+900 =1321 pairs)	67.67%



References

- Min-Yuh Day, Re-Yuan Lee, Cheng-Tai Liu, Chun Tu, Chin-Sheng Tseng, Loong Tern Yap, Allen-Green C.L. Huang, Yu-Hsuan Chiu and Wei-Ze Hong (2011), "IMTKU Textual Entailment System for Recognizing Inference in Text at NTCIR-9 RITE," in Proceedings of the 9th NTCIR Workshop Meeting on Evaluation of Information Access Technologies (NTCIR-9), Tokyo, Japan, December 6-9, 2011, pp. 340-344 .
- Hideki Shima, Hiroshi Kanayama, Cheng-Wei Lee, ChuanJie Lin, Teruko Mitamura, Yusuke Miyao, Shuming Shi, and Koichi Takeda, "Overview of NTCIR-9 RITE: Recognizing Inference in TExt," in Proceedings of the 9th NTCIR Workshop Meeting on Evaluation of Information Access Technologies (NTCIR-9), Tokyo, Japan, December 6-9, 2011, pp. 291-301.