

WMT 2019 - Biomedical task

May 17, 2019

1 Results for Automatic Evaluation

BLEU scores were calculated using the multi-eval tool and tokenization as provided in Moses.

* indicates the primary run, as informed by the participants, in the case of multiple runs.

1.1 MEDLINE test sets

Teams	Runs	de/en	en/de	es/en	en/es	en/fr	fr/en
ARC	1	0.2871	0.2789	-	-	0.3995	0.3551
ARC	2	0.2879	0.2786	-	-	0.3667	0.3551
ARC	3	0.2882*	0.2785*	-	-	0.3619*	0.3556*
BSC	1	-	-	0.3769*	0.4421*	-	-
KU	1	-	-	-	-	0.3114*	-
MT-UOC-UPF	1	-	-	0.3659*	0.3974*	-	-
nrpu	1	-	-	-	-	0.1587*	0.1972*
Talp_upc	1	-	-	0.3941	0.4301*	-	-
Talp_upc	2	-	-	0.3792*	0.4340	-	-
Talp_upc	3	-	-	0.3721	0.4392	-	-
UCAM	1	0.2741	0.2651	0.4241	0.4492	-	-
UCAM	2	0.2863	0.2716	0.4303	0.4539	-	-
UCAM	3	0.2850*	0.2641*	0.4290*	0.4558*	-	-
UHH-DS	1	-	-	0.3561*	-	-	-
UHH-DS	2	-	-	0.3585	-	-	-
UHH-DS	3	-	-	0.3586	-	-	-

Table 1: BLEU results with all sentences

Teams	Runs	de/en	en/de	es/en	en/es	en/fr	fr/en
ARC	1	0.3866	0.3539	-	-	0.4241	0.3818
ARC	2	0.3880	0.3528	-	-	0.3889	0.3818
ARC	3	0.3884*	0.3526*	-	-	0.3829*	0.3824*
BSC	1	-	-	0.4356*	0.4701*	-	-
KU	1	-	-	-	-	0.3292*	-
MT-UOC-UPF	1	-	-	0.4159*	0.4219*	-	-
nrpu	1	-	-	-	-	0.1745*	0.2105*
Talp_upc	1	-	-	0.4509	0.4568*	-	-
Talp_upc	2	-	-	0.4355*	0.4609	-	-
Talp_upc	3	-	-	0.4270	0.4683	-	-
UCAM	1	0.3669	0.3328	0.4770	0.4834	-	-
UCAM	2	0.3807	0.3469	0.4833	0.4891	-	-
UCAM	3	0.3799*	0.3379*	0.4811*	0.4896*	-	-
UHH-DS	1	-	-	0.3969*	-	-	-
UHH-DS	2	-	-	0.3999	-	-	-
UHH-DS	3	-	-	0.3997	-	-	-

Table 2: BLEU results with OK sentences

Teams	Runs	pt/en	en/pt	zh/en	en/zh
ARC	1	-	-	0.3007	0.3547
ARC	2	-	-	0.3005	0.3547
ARC	3	-	-	0.3005*	0.3547*
BSC	1	0.3990*	0.4811*	-	-
KU	1	-	-	0.2489*	-
OOM	1	-	-	0.3413	0.4234
OOM	2	-	-	0.3413*	0.4234*
peace	1	-	-	0.2266*	0.3379*
Radiant	1	-	-	-	0.3266
Radiant	2	-	-	-	0.3265
Radiant	3	-	-	-	0.3294*

Table 3: BLEU results with all sentences

Teams	Runs	pt/en	en/pt	zh/en	en/zh
ARC	1	-	-	0.3215	0.3709
ARC	2	-	-	0.3216	0.3709
ARC	3	-	-	0.3216*	0.3709*
BSC	1	0.4617*	0.4951*	-	-
KU	1	-	-	0.2716*	-
OOM	1	-	-	0.3561	0.4392
OOM	2	-	-	0.3561*	0.4392*
peace	1	-	-	0.2518*	0.3508*
Radiant	1	-	-	-	0.3405
Radiant	2	-	-	-	0.3416
Radiant	3	-	-	-	0.3424*

Table 4: BLEU results with OK sentences

1.2 Terminology test set

Teams	Runs	en/es
MT-UOC-UPF	1	47.55

2 Results for Manual Validation

(coming soon)