

Results SemEval-2018 Task 5

Counting Events and Participants in the Long Tail

Marten Postma, Filip Ilievski, and Piek Vossen
{m.c.postma, f.ilievski, piek.vossen@vu.nl}

February 6, 2018

This document contains the official results for [SemEval-2018 Task 5 Counting Events and Participants in the Long Tail](#). For the explanation about each metric, we refer to the [Data and evaluation](#) tab on the Codalab website. Please note the following important details about the result tables:

normalized scores for the incident-level evaluation (*inc_accuracy* in metric name) and document-level evaluation (*doc_f1* in metric name), we report both the performance with respect to the subset of questions answered (e.g. *s1_doc_f1*) and the normalized score (e.g. *s1_doc_f1 normalized*). The normalized score indicates the performance on all questions of a subtask. Obviously, if a submission has answers for all questions, the normalized score will be the same as the non-normalized score.

RMSE For most metrics, a higher metric value indicates better system performance. For RMSE, a lower metric value indicates better system performance. In addition, the RMSE scores have not been normalized since it is not reasonable to set a default value for non-answered questions. Finally, the RMSE results are ordered first by number of answered questions, and then by the RMSE value.

event coreference The mention-level task (Event Coreference) was the same across all three subtasks, The results are hence only reported once.

1 Subtask 1

Rank	team	s1_doc_f1 normalized	s1_doc_f1	s1 % answered
1	superlyc	78.33	78.33	100.00
2	IDDE	36.67	82.99	44.19
3	CarlaAbreu	24.65	24.65	100.00
4	*Piek	23.82	46.20	51.55
5	baseline1	11.09	67.33	16.47

Table 1: results for evaluation metric: **s1_doc_f1**.

We mark explicitly with an asterisk the teams that had a task co-organizer as a team member

2 Subtask 2

Rank	team	s2_inc_accuracy normalized	s2_inc_accuracy	s2 % answered
1	CarlaAbreu	26.38	26.38	100.0
2	*Piek	21.87	21.87	100.0
3	baseline1	18.25	18.25	100.0
4	superlyc	17.35	17.35	100.0
5	IDDE	13.74	20.36	67.5

Table 2: results for evaluation metric: **s2_inc_accuracy**.

We mark explicitly with an asterisk the teams that had a task co-organizer as a team member

Rank	team	s2_inc_rmse	s2 % answered
1	CarlaAbreu	6.13	100.0
2	baseline1	8.50	100.0
3	superlyc	20.59	100.0
4	*Piek	43.96	100.0
5	IDDE	6.15	67.5

Table 3: results for evaluation metric: **s2_inc_rmse**.

We mark explicitly with an asterisk the teams that had a task co-organizer as a team member

Rank	team	s2_doc_f1 normalized	s2_doc_f1	s2 % answered
1	superlyc	50.52	50.52	100.0
2	IDDE	37.24	55.16	67.5
3	*Piek	36.91	36.91	100.0
4	CarlaAbreu	30.51	30.51	100.0
5	baseline1	26.38	26.38	100.0

Table 4: results for evaluation metric: **s2_doc_f1**.

We mark explicitly with an asterisk the teams that had a task co-organizer as a team member

3 Subtask 3

Rank	team	s3_inc_accuracy normalized	s3_inc_accuracy	s3 % answered
1	CarlaAbreu	30.42	30.42	100.00
2	*Piek	21.05	21.05	100.00
3	superlyc	20.20	20.20	100.00
4	IDDE	12.87	19.32	66.61

Table 5: results for evaluation metric: **s3_inc_accuracy**.

We mark explicitly with an asterisk the teams that had a task co-organizer as a team member

Rank	team	s3_inc_rmse	s3 % answered
1	superlyc	13.45	100.00
2	*Piek	296.45	100.00
3	CarlaAbreu	478.72	100.00
4	IDDE	7.87	66.61

Table 6: results for evaluation metric: **s3_inc_rmse**.

We mark explicitly with an asterisk the teams that had a task co-organizer as a team member

Rank	team	s3_doc_f1 normalized	s3_doc_f1	s3 % answered
1	superlyc	63.59	63.59	100.00
2	IDDE	46.33	69.56	66.61
3	*Piek	26.84	26.84	100.00
4	CarlaAbreu	26.79	26.79	100.00

Table 7: results for evaluation metric: **s3_doc_f1**.

We mark explicitly with an asterisk the teams that had a task co-organizer as a team member

4 Event Coreference

Rank	team	men_coref_avg
1	IDDE	42.80
2	*Piek	40.00
3	baseline1	6.47

Table 8: results for evaluation metric: **men_coref_avg**.

We mark explicitly with an asterisk the teams that had a task co-organizer as a team member