SemEval-2016 Task 3: Community Question Answering

Preslav Nakov   Lluís Màrquez   Walid Magdy   Alessandro Moschitti
ALT Research Group, Qatar Computing Research Institute

James Glass
MIT Computer Science and Artificial Intelligence Laboratory

Bilal Randeree
Qatar Living

	Submission	MAP	AvgRec	MRR	P	R	F1	Acc
1	**Kelp-primary**	79.19_1	88.82_1	86.42_1	76.96_1	55.30_8	64.36_5	75.11_2
	ConvKN-contrastive1	78.71	88.98	86.15	77.78	53.72	63.55	74.95
	SUper_team-contrastive1	77.68	88.06	84.76	75.59	55.00	63.68	74.50
2	**ConvKN-primary**	77.66_2	88.05_3	84.93_4	75.56_2	58.84_6	66.16_2	75.54_1
3	**SemanticZ-primary**	77.58_3	88.14_2	85.21_2	74.13_4	53.05_{10}	61.84_8	73.39_5
	ConvKN-contrastive2	77.29	87.77	85.03	74.74	59.67	66.36	75.41
4	**ECNU-primary**	77.28_4	87.52_5	84.09_6	70.46_6	63.36_4	66.72_1	74.31_4
	SemanticZ-contrastive1	77.16	87.73	84.08	75.29	53.20	62.35	73.88
5	**SUper_team-primary**	77.16_5	87.98_4	84.69_5	74.43_3	56.73_7	64.39_4	74.50_3
	MTE-NN-contrastive2	76.98	86.98	85.50	58.71	70.28	63.97	67.83
	SUper_team-contrastive2	76.97	87.89	84.58	74.31	56.36	64.10	74.34
	MTE-NN-contrastive1	76.86	87.03	84.36	55.84	77.35	64.86	65.93
	SLS-contrastive2	76.71	87.17	84.38	59.45	67.95	63.41	68.13
	SLS-contrastive1	76.46	87.47	83.27	60.09	69.68	64.53	68.87
6	**MTE-NN-primary**	76.44_6	86.74_7	84.97_3	56.28_9	76.22_1	64.75_3	66.27_8
7	**SLS-primary**	76.33_7	87.30_6	82.99_7	60.36_8	67.72_3	63.83_6	68.81_7
	ECNU-contrastive2	75.71	86.14	82.53	63.60	66.67	65.10	70.95
	SemanticZ-contrastive2	75.41	86.51	82.52	73.19	50.11	59.49	72.26
	ICRC-HIT-contrastive1	73.34	84.81	79.65	63.43	69.30	66.24	71.28
8	**ITNLP-AiKF-primary**	71.52_8	82.67_9	80.26_8	73.18_5	19.71_{12}	31.06_{12}	64.43_9
	ECNU-contrastive1	71.34	83.39	78.62	66.95	41.31	51.09	67.86
9	**ICRC-HIT-primary**	70.90_9	83.36_8	77.38_{10}	62.48_7	62.53_5	62.50_7	69.51_6
10	**PMI-cool-primary**	68.79_{10}	79.94_{10}	80.00_9	47.81_{12}	70.58_2	57.00_9	56.73_{12}
	UH-PRHLT-contrastive1	67.57	79.50	77.08	54.10	50.11	52.03	62.45
11	**UH-PRHLT-primary**	67.42_{11}	79.38_{11}	76.97_{11}	55.64_{10}	46.80_{11}	50.84_{11}	63.21_{10}
	UH-PRHLT-contrastive2	67.33	79.34	76.73	54.97	49.13	51.89	62.97
12	**QAIIIT-primary**	62.24_{12}	75.41_{12}	70.58_{12}	50.28_{11}	53.50_9	51.84_{10}	59.60_{11}
	QAIIIT-contrastive2	61.93	75.22	69.95	49.48	49.96	49.72	58.93
	QAIIIT-contrastive1	61.80	75.12	69.76	49.85	50.94	50.39	59.24
	Baseline 1 (IR)	**59.53**	**72.60**	**67.83**	—	—	—	—
	Baseline 2 (random)	52.80	66.52	58.71	40.56	74.57	52.55	45.26
	Baseline 3 (all 'true')	—	—	—	40.64	100.00	**57.80**	40.64
	Baseline 4 (all 'false')	—	—	—	—	—	—	59.36

Table 1: **Subtask A, English (Question-Comment Similarity):** results for all submissions. The first column shows the rank of the primary runs with respect to the official MAP score. The second column contains the team's name and its submission type (primary vs. contrastive). The following columns show the results for the primary, and then for other, unofficial evaluation measures. The subindices show the rank of the primary runs with respect to the evaluation measure in the respective column.

	Submission	MAP	AvgRec	MRR	P	R	F1	Acc
	UH-PRHLT-contrastive2	77.33	90.84	83.93	63.57	70.39	66.80	76.71
1	**UH-PRHLT-primary**	**76.70**$_1$	**90.31**$_4$	**83.02**$_4$	**63.53**$_7$	**69.53**$_3$	**66.39**$_3$	**76.57**$_4$
	UH-PRHLT-contrastive1	76.56	90.22	83.02	62.74	70.82	66.53	76.29
	Kelp-contrastive1	76.28	91.33	82.71	63.83	77.25	69.90	77.86
	Kelp-contrastive2	76.27	91.44	84.10	64.06	77.25	70.04	78.00
	SLS-contrastive1	76.17	90.55	85.48	74.39	52.36	61.46	78.14
	SLS-contrastive2	76.09	90.14	84.21	77.21	45.06	56.91	77.29
2	**ConvKN-primary**	**76.02**$_2$	**90.70**$_2$	**84.64**$_1$	**68.58**$_3$	**66.52**$_6$	**67.54**$_2$	**78.71**$_3$
3	**Kelp-primary**	**75.83**$_3$	**91.02**$_1$	**82.71**$_6$	**66.79**$_4$	**75.97**$_2$	**71.08**$_1$	**79.43**$_1$
	ConvKN-contrastive1	75.57	89.64	83.57	63.77	72.53	67.87	77.14
4	**SLS-primary**	**75.55**$_4$	**90.65**$_3$	**84.64**$_1$	**76.33**$_2$	**55.36**$_9$	**64.18**$_6$	**79.43**$_1$
	SUper_team-contrastive1	75.17	88.84	83.66	63.25	63.52	63.38	75.57
5	**ICL00-primary**	**75.11**$_5$	**89.33**$_5$	**83.02**$_4$	**33.29**$_{11}$	**100.00**$_1$	**49.95**$_9$	**33.29**$_{11}$
	ICL00-contrastive1	74.89	89.08	82.71	33.29	100.00	49.95	33.29
6	**SUper_team-primary**	**74.82**$_6$	**88.54**$_7$	**83.66**$_3$	**63.64**$_6$	**57.08**$_8$	**60.18**$_7$	**74.86**$_7$
	ICL00-contrastive2	74.05	89.11	82.79	33.29	100.00	49.95	33.29
7	**ECNU-primary**	**73.92**$_7$	**89.07**$_6$	**81.48**$_7$	**100.00**$_1$	**18.03**$_{11}$	**30.55**$_{11}$	**72.71**$_9$
	ECNU-contrastive1	73.25	88.55	80.81	100.00	18.03	30.55	72.71
	ECNU-contrastive2	71.62	86.55	80.88	54.61	71.24	61.82	70.71
8	**ITNLP-AiKF-primary**	**71.43**$_8$	**87.31**$_8$	**81.28**$_8$	**62.75**$_9$	**68.67**$_4$	**65.57**$_4$	**76.00**$_6$
9	**UniMelb-primary**	**70.20**$_9$	**86.21**$_9$	**78.58**$_{11}$	**63.96**$_5$	**54.08**$_{10}$	**58.60**$_8$	**74.57**$_8$
10	**overfitting-primary**	**69.68**$_{10}$	**85.10**$_{10}$	**80.18**$_9$	**63.20**$_8$	**67.81**$_5$	**65.42**$_5$	**76.14**$_5$
	QAIIIT-contrastive1	69.24	85.24	80.30	38.99	66.09	49.04	54.29
11	**QAIIIT-primary**	**69.04**$_{11}$	**84.53**$_{11}$	**79.55**$_{10}$	**39.53**$_{10}$	**64.81**$_7$	**49.11**$_{10}$	**55.29**$_{10}$
	QAIIIT-contrastive2	46.23	68.07	48.92	36.25	51.50	42.55	53.71
	Baseline 1 (IR)	**74.75**	**88.30**	**83.79**	—	—	—	—
	Baseline 2 (random)	46.98	67.92	50.96	32.58	73.82	45.20	40.43
	Baseline 3 (all 'true')	—	—	—	33.29	100.00	**49.95**	33.29
	Baseline 4 (all 'false')	—	—	—	—	—	—	**66.71**

Table 2: **Subtask B, English (Question-Question Similarity):** results for all submissions. The first column shows the rank of the primary runs with respect to the official MAP score. The second column contains the team's name and its submission type (primary vs. contrastive). The following columns show the results for the primary, and then for other, unofficial evaluation measures. The subindices show the rank of the primary runs with respect to the evaluation measure in the respective column.

	Submission	MAP	AvgRec	MRR	P	R	F1	Acc
	Kelp-contrastive2	55.58	63.36	61.19	32.21	70.18	44.16	83.41
1	**SUper_team-primary**	**55.41**$_1$	**60.66**$_1$	**61.48**$_1$	**18.03**$_7$	**63.15**$_4$	**28.05**$_4$	**69.73**$_8$
	SUper_team-contrastive2	53.48	59.40	59.09	18.42	66.97	28.89	69.20
2	**Kelp-primary**	**52.95**$_2$	**59.27**$_2$	**59.23**$_2$	**33.63**$_5$	**64.53**$_3$	**44.21**$_1$	**84.79**$_5$
	Kelp-contrastive1	52.95	59.34	58.06	34.08	65.29	44.78	84.96
3	**SemanticZ-primary**	**51.68**$_3$	**53.43**$_6$	**55.96**$_4$	**17.11**$_8$	**57.65**$_5$	**26.38**$_5$	**69.94**$_7$
	SemanticZ-contrastive1	51.46	52.69	55.75	16.94	57.49	26.17	69.69
	MTE-NN-contrastive2	49.49	55.78	51.80	15.68	73.09	25.82	60.76
4	**MTE-NN-primary**	**49.38**$_4$	**55.44**$_4$	**51.56**$_7$	**15.26**$_9$	**76.15**$_2$	**25.43**$_6$	**58.27**$_9$
5	**ICL00-primary**	**49.19**$_5$	**51.07**$_7$	**53.89**$_6$	**9.34**$_{10}$	**100.00**$_1$	**17.09**$_8$	**9.34**$_{10}$
6	**SLS-primary**	**49.09**$_6$	**56.04**$_3$	**55.98**$_3$	**47.85**$_2$	**13.61**$_8$	**21.19**$_7$	**90.54**$_2$
	SemanticZ-contrastive2	48.76	50.72	53.85	16.92	57.34	26.13	69.71
	MTE-NN-contrastive1	48.52	54.71	50.51	15.13	76.61	25.27	57.67
7	**ITNLP-AiKF-primary**	**48.49**$_7$	**55.16**$_5$	**55.21**$_5$	**30.05**$_6$	**50.92**$_6$	**37.80**$_2$	**84.34**$_6$
	ECNU-contrastive1	48.49	53.17	53.47	68.75	8.41	14.99	91.09
	SUper_team-contrastive1	48.23	54.93	54.85	22.81	36.70	28.14	82.49
	ECNU-contrastive2	47.24	53.21	51.89	70.27	7.95	14.29	91.09
	ICL00-contrastive1	47.23	49.71	50.28	9.34	100.00	17.09	9.34
8	**ConvKN-primary**	**47.15**$_8$	**47.46**$_{10}$	**51.43**$_8$	**45.97**$_3$	**8.72**$_{10}$	**14.65**$_{10}$	**90.51**$_3$
	SLS-contrastive1	46.48	53.31	52.53	16.24	85.93	27.32	57.29
9	**ECNU-primary**	**46.47**$_9$	**50.92**$_8$	**51.41**$_9$	**66.29**$_1$	**9.02**$_9$	**15.88**$_9$	**91.07**$_1$
	SLS-contrastive2	46.39	52.83	51.17	16.18	85.63	27.22	57.23
	UH-PRHLT-contrastive1	43.37	48.01	48.43	38.56	32.72	35.40	88.84
	UH-PRHLT-contrastive2	43.32	47.97	48.45	38.21	32.72	35.26	88.77
	ConvKN-contrastive1	43.31	44.19	48.89	30.00	3.21	5.80	90.26
10	**UH-PRHLT-primary**	**43.20**$_{10}$	**47.96**$_9$	**47.79**$_{10}$	**37.65**$_4$	**34.25**$_7$	**35.87**$_3$	**88.56**$_4$
	ICL00-contrastive2	41.32	44.56	43.55	9.34	100.00	17.09	9.34
	ConvKN-contrastive2	41.12	38.89	44.17	33.55	32.11	32.81	87.71
	Baseline 1 (IR)	**40.36**	**45.97**	**45.83**	—	—	—	—
	Baseline 2 (random)	15.01	11.44	15.19	9.40	75.69	16.73	29.59
	Baseline 3 (all 'true')	—	—	—	9.34	100.00	**17.09**	9.34
	Baseline 4 (all 'false')	—	—	—	—	—	—	90.66

Table 3: **Subtask C, English (Question-External Comment Similarity):** results for all submissions. The first column shows the rank of the primary runs with respect to the official MAP score. The second column contains the team's name and its submission type (primary vs. contrastive). The following columns show the results for the primary, and then for other, unofficial evaluation measures. The subindices show the rank of the primary runs with respect to the evaluation measure in the respective column.

	Submission	MAP	AvgRec	MRR	P	R	F1	Acc
1	**SLS-primary**	**45.83_1**	51.01_1	53.66_1	34.45_1	52.33_3	41.55_1	71.67_1
2	**ConvKN-primary**	**45.50_2**	50.13_2	52.55_2	28.55_2	64.53_2	39.58_2	62.10_3
	SLS-contrastive1	44.94	49.72	51.58	62.96	2.40	4.62	80.95
3	**RDI_team-primary**	**43.80_3**	47.45_3	49.21_3	19.24_5	100.00_1	32.27_4	19.24_5
	SLS-contrastive2	42.95	47.61	49.55	27.20	74.40	39.84	56.76
	RDI_team-contrastive1	42.18	47.03	47.93	19.24	100.00	32.27	19.24
	ConvKN-contrastive2	39.98	43.68	46.41	26.26	68.39	37.95	57.00
	QU-IR-contrastive2	39.07	42.72	44.14	24.90	44.08	31.82	63.66
	RDI_team-contrastive2	38.84	42.98	42.97	19.24	100.00	32.27	19.24
4	**QU-IR-primary**	**38.63_4**	44.10_4	46.27_4	25.50_3	45.13_5	32.59_4	64.07_2
	ConvKN-contrastive1	38.33	42.09	43.75	20.38	96.95	33.68	26.58
	QU-IR-contrastive1	37.80	40.96	44.39	23.54	41.89	30.14	62.64
5	**UPC_USMBA-primary**	**29.09_5**	30.04_5	34.05_5	20.14_4	51.69_4	28.99_5	51.27_4
	Baseline 1 (IR)	**28.88**	28.71	30.93	—	—	—	—
	Baseline 2 (random)	29.79	31.00	33.71	19.53	20.66	20.08	68.35
	Baseline 3 (all 'true')	—	—	—	19.24	100.00	**32.27**	19.24
	Baseline 4 (all 'false')	—	—	—	—	—	—	80.76

Table 4: **Subtask D, Arabic (Reranking the correct answers for a new question):** results for all submissions. The first column shows the rank of the primary runs with respect to the official MAP score. The second column contains the team's name and its submission type (primary vs. contrastive). The following columns show the results for the primary, and then for other, unofficial evaluation measures. The subindices show the rank of the primary runs with respect to the evaluation measure in the respective column.

Team ID	Team Affiliation
ConvKN	Qatar Computing Research Institute, HBKU, Qatar; University of Trento, Italy
ECNU	East China Normal University, China
ICL00	Institute of Computational Lingustics, Peking University, China
ICRC-HIT	Intelligence Computing Research Center, Harbin Institute of Technology, China
ITNLP-AiKF	Intelligence Technology and Natural Language Processing Laboratory, Harbin Institute of Technology, China
Kelp	University of Roma, Tor Vergata, Italy; Qatar Computing Research Institute, HBKU, Qatar
MTE-NN	Qatar Computing Research Institute, HBKU, Qatar
overfitting	University of Waterloo, Canada
PMI-cool	Sofia University, Bulgaria
QAIIIT	IIIT Hyderabad, India
QU-IR	Qatar University, Qatar
RDI_team	RDI Egypt, Cairo University, Egypt
SemanticZ	Sofia University, Bulgaria
SLS	MIT Computer Science and Artificial Intelligence Lab, USA
SUper_team	Sofia University, Bulgaria; Qatar Computing Research Institute, HBKU, Qatar
UH-PRHLT	Pattern Recognition and Human Language Technologies Research Center, Universitat Politècnica de València; University of Houston
UniMelb	The University of Melbourne, Australia
UPC_USMBA	Universitat Politècnica de Catalunya, Spain; Sidi Mohamed Ben Abdellah University, Morocco

Table 5: The participating teams and their affiliations.