

SemEval-2016 Task 4: Sentiment Analysis in Twitter

Preslav Nakov♣, Alan Ritter◇, Sara Rosenthal♡, Fabrizio Sebastiani♣*, Veselin Stoyanov♠

♣Qatar Computing Research Institute, Hamad bin Khalifa University, Qatar

◇Department of Computer Science and Engineering, The Ohio State University, USA

♡Department of Computer Science, Columbia University, USA

♠Facebook, USA

Fabrizio Sebastiani is currently on leave from Consiglio Nazionale delle Ricerche, Italy.

#	System	2013		2014			2015	2016
		Tweet	SMS	Tweet	Tweet sarcasm	Live- Journal	Tweet	Tweet
1	SwissCheese	0.700 ₅	0.637 ₂	0.716 ₅	0.566 ₁	0.695 ₇	0.671 ₁	0.633₁
2	SENSEI-LIF	0.706 ₄	0.634 ₃	0.744 ₂	0.467 ₈	0.741 ₁	0.662 ₂	0.630₂
3	unimelb	0.687 ₇	0.593 ₁₀	0.706 ₇	0.449 ₁₁	0.683 ₉	0.650 ₄	0.617₃
4	INESC-ID	0.723 ₂	0.609 ₆	0.728 ₃	0.554 ₃	0.702 ₄	0.657 ₃	0.610₄
5	aueb.twitter.sentiment	0.666 ₈	0.618 ₅	0.708 ₆	0.410 ₁₇	0.695 ₇	0.623 ₇	0.605₅
6	SentiSys	0.714 ₃	0.633 ₄	0.723 ₄	0.515 ₅	0.726 ₂	0.644 ₅	0.598₆
7	I2RNTU	0.693 ₆	0.597 ₇	0.680 ₈	0.469 ₆	0.696 ₆	0.638 ₆	0.596₇
8	INSIGHT-1	0.602 ₁₆	0.582 ₁₂	0.644 ₁₆	0.391 ₂₃	0.559 ₂₃	0.595 ₁₆	0.593₈
9	twise	0.610 ₁₅	0.540 ₁₇	0.645 ₁₄	0.450 ₁₀	0.648 ₁₃	0.621 ₈	0.586₉
10	ECNU	0.643 ₁₀	0.593 ₉	0.662 ₉	0.425 ₁₄	0.663 ₁₀	0.606 ₁₁	0.585₁₀
11	NTNUSentEval	0.623 ₁₂	0.641 ₁	0.650 ₁₁	0.427 ₁₃	0.719 ₃	0.599 ₁₃	0.583₁₁
12	MDSENT	0.589 ₁₉	0.509 ₂₁	0.588 ₂₀	0.386 ₂₄	0.606 ₁₉	0.593 ₁₈	0.580₁₂
13	CUFE	0.642 ₁₁	0.596 ₈	0.661 ₁₀	0.466 ₉	0.697 ₅	0.598 ₁₄	0.580₁₃
14	THUIR	0.616 ₁₃	0.575 ₁₄	0.648 ₁₂	0.400 ₂₀	0.640 ₁₆	0.617 ₁₀	0.576₁₄
15	PUT	0.565 ₂₁	0.511 ₂₀	0.615 ₁₉	0.360 ₂₇	0.648 ₁₄	0.597 ₁₅	0.576₁₅
16	LYS	0.650 ₉	0.579 ₁₃	0.647 ₁₃	0.406 ₁₈	0.655 ₁₁	0.603 ₁₂	0.575₁₆
17	IIP	0.598 ₁₇	0.466 ₂₃	0.645 ₁₄	0.405 ₁₉	0.640 ₁₅	0.619 ₉	0.574₁₇
18	DIEGOLab16	0.813 ₁	0.543 ₁₆	0.759 ₁	0.562 ₂	0.615 ₁₈	0.595 ₁₇	0.571₁₈
19	UniPI	0.592 ₁₈	0.585 ₁₁	0.627 ₁₈	0.381 ₂₅	0.654 ₁₂	0.586 ₁₉	0.571₁₉
20	GTI	0.612 ₁₄	0.524 ₁₈	0.639 ₁₇	0.468 ₇	0.623 ₁₇	0.584 ₂₀	0.539₂₀
21	OPAL	0.567 ₂₀	0.562 ₁₅	0.556 ₂₃	0.395 ₂₁	0.594 ₂₁	0.531 ₂₂	0.505₂₁
22	DSIC-ELIRF	0.494 ₂₅	0.404 ₂₆	0.546 ₂₆	0.343 ₂₉	0.517 ₂₄	0.531 ₂₁	0.502₂₂
23	UofL	0.490 ₂₆	0.443 ₂₄	0.547 ₂₅	0.372 ₂₆	0.574 ₂₂	0.502 ₂₅	0.499₂₃
24	ELiRF	0.462 ₂₈	0.408 ₂₅	0.514 ₂₈	0.310 ₃₃	0.493 ₂₅	0.493 ₂₆	0.498₂₄
25	ISTI-CNR	0.538 ₂₂	0.492 ₂₂	0.572 ₂₁	0.327 ₃₀	0.598 ₂₀	0.508 ₂₄	0.494₂₅
26	SteM	0.518 ₂₃	0.315 ₂₉	0.571 ₂₂	0.320 ₃₂	0.405 ₂₈	0.517 ₂₃	0.478₂₆
27	Tweester	0.506 ₂₄	0.340 ₂₈	0.529 ₂₇	0.540 ₄	0.379 ₂₉	0.479 ₂₈	0.455₂₇
28	Minions	0.489 ₂₇	0.521 ₁₉	0.554 ₂₄	0.420 ₁₆	0.475 ₂₆	0.481 ₂₇	0.415₂₈
29	aicyber	0.418 ₂₉	0.361 ₂₇	0.457 ₂₉	0.326 ₃₁	0.440 ₂₇	0.432 ₂₉	0.402₂₉
30	mib	0.394 ₃₀	0.310 ₃₀	0.415 ₃₁	0.352 ₂₈	0.359 ₃₁	0.413 ₃₁	0.401₃₀
31	VCU-TSA	0.383 ₃₁	0.306 ₃₁	0.444 ₃₀	0.425 ₁₅	0.336 ₃₂	0.416 ₃₀	0.372₃₁
32	SentimentalITists	0.339 ₃₃	0.238 ₃₃	0.393 ₃₃	0.288 ₃₄	0.323 ₃₄	0.343 ₃₃	0.339₃₂
33	Wisers_Research	0.355 ₃₂	0.284 ₃₂	0.393 ₃₂	0.430 ₁₂	0.366 ₃₀	0.377 ₃₂	0.330₃₃
34	CICBUAPnlp	0.194 ₃₄	0.193 ₃₄	0.335 ₃₄	0.393 ₂₂	0.326 ₃₃	0.303 ₃₄	0.303₃₄
	baseline (positive)	0.292	0.190	0.346	0.277	0.272	0.303	0.255

Table 1: Results for Subtask A “Message Polarity Classification”. The systems are ordered by their score on the Twitter2016 test dataset; the rankings on the individual datasets are indicated with a subscript.

#	System	Score
1	Tweester	0.797
2	LYS	0.791
3	thecerealkiller	0.784
4	ECNU	0.768
5	INSIGHT-1	0.767
6	PUT	0.763
7	unimelb	0.758
8	twise	0.756
9	GTI	0.736
10	finki	0.720
11	pkudblab	0.689
12	CUFE	0.679
13	ISTI-CNR	0.671
14	SwissCheese	0.648
15	SentimentalITists	0.624
16	PotTS	0.618
17	OPAL	0.616
18	Wisers_Research	0.522
19	VCU-TSA	0.502
	baseline (positive)	0.500

Table 2: Results for Subtask B “Tweet classification according to a two-point scale”. The systems are ordered by their ρ^{PN} score (higher is better).

#	System	Score
1	twise	0.719
2	ECNU	0.806
3	PUT	0.860
4	LYS	0.864
5	finki	0.869
6	INSIGHT-1	1.006
7	ISTI-CNR	1.074
8	YZU-NLP	1.111
9	SentimentalITists	1.148
10	PotTS	1.237
11	pkudblab	1.697
	baseline (0)	1.200

Table 3: Results for Subtask C “Tweet classification according to a five-point scale”. The systems are ordered by their MAE^M score (lower is better).

#	System	Score
1	finki	0.034
2	LYS	0.053
	twise	0.053
4	INSIGHT-1	0.054
5	GTI	0.055
	QCRI	0.055
7	NRU-HSE	0.084
8	PotTS	0.094
9	pkudblab	0.099
10	ECNU	0.121
11	ISTI-CNR	0.127
12	SwissCheese	0.191
13	UDLAP	0.261
14	HSENN	0.399
	baseline (1 0)	0.887
	baseline (avg on train+dev+devtest)	0.175

Table 4: Results for Subtask D “Tweet quantification according to a two-point scale”. The systems are ordered by their *KLD* score (lower is better).

#	System	Score
1	QCRI	0.243
2	finki	0.316
3	pkudblab	0.331
4	NRU-HSE	0.334
5	ECNU	0.341
6	ISTI-CNR	0.358
7	LYS	0.360
8	INSIGHT-1	0.366
9	HSENN	0.545
10	PotTS	0.818
	baseline (0 0 0 1 0)	0.734
	baseline (avg on train+dev+devtest)	0.474

Table 5: Results for Subtask E “Tweet quantification according to a five-point scale”. The systems are ordered by their *EMD* score (lower is better).

Subtasks	Team ID	Affiliation	Nation	Paper
A	aicyber	aicyber.com	China	
A	aueb.twitter.sentiment	Department of Informatics, Athens University of Economics and Business	Greece	
A	CICBUAPnlp	Instituto Politecnico Nacional Benemrita Universidad Autonoma de Puebla	Mexico	
A B	CUFE	Cairo University	Egypt	
A	DIEGOLab16	Arizona State University	USA	
A	DSIC-ELIRF	Universitat Politcnica de Valncia	Spain	
A B C D E	ECNU	East China Normal University	China	
A	ELiRF	Universitat Politcnica de Valncia	Spain	
B C D E	finki	Saints Cyril and Methodius University, Skopje	Macedonia	
A B D	GTI	AtlantTIC Centre, University of Vigo	Spain	
D E	HSENN	National Research University Higher School of Economics	Russia	
A	I2RNTU	Institute for Infocomm Research, A*STAR School of Computer Engineering, Nanyang Technological University	Singapore	
A	IIP	Infosys Limited	India	
A	INESC-ID	INESC-ID, Lisboa Instituto Superior Tcnico, Universidade de Lisboa	Portugal	
A B C D E	INSIGHT-1	INSIGHT Research Centre, National University of Ireland, Galway AYLIEN Inc.	Ireland	
A B C D E	LYS	Universidade da Corua Universidade de Vigo	Spain	
A	MDSSENT	University of Maryland Baltimore County	USA	
A	mib	Istituto di Informatica e Telematica, Consiglio Nazionale delle Ricerche	Italy	
A	Minions	University of Iasi	Romania	
A B C D E	ISTI-CNR	Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche	Italy	
D E	NRU-HSE	National Research University Higher School of Economics	Russia	
A	NTNUSentEval	Norwegian University of Science and Technology	Norway	
A B	OPAL	European Commission Joint Research Centre	Italy	
B C D E	pkudblab	Peking University	China	
B C D E	PotTS	University of Potsdam Retresco GmbH	Germany	
A B C	PUT	Poznan University of Technology	Poland	
D E	QCRI (**)	Qatar Computing Research Institute	Qatar	(?)
A	SENSEI-LIF	Aix-Marseille University - CNRS - LIF	France	
A B C	SentimentalTists	University of Iasi	Romania	
A	SentiSys	Aix-Marseille University	France	
A	SteM	Sabanci University Marmara University Otto-von-Guericke University Magdeburg	Turkey Turkey Germany	
A B D	SwissCheese	ETH Zrich	Switzerland	
B	thecerealkiller	Amazon.com	USA	
A	THUIR	Tsinghua University	China	
A B	Tweester	School of ECE, Technical University of Crete Department of Informatics, University of Athens Signal Analysis and Interpretation Laboratory (SAIL) Institute for Language & Speech Processing - ILSP	Greece	
A B C D	twise	University of Grenoble-Alpes	France	
D	UDLAP	Universidad de las Amricas Puebla (UDLAP)	Mexico	
A B	unimelb	University of Melbourne	Australia	
A	UniPI	Universit di Pisa	Italy	
A	UofL	University of Louisville	USA	
A B	VCU-TSA	Virginia Commonwealth University	USA	
A B	Wisers_Research	Wisers Information Limited	Hong Kong SAR, China	
C	YZU-NLP	Yuan Ze University, Taoyuan Yunnan University, Kunming	Taiwan China	
34 19 11 14 10	Total			

Table 6: Participating teams (Column 2), their affiliation (Column 3) and nationality (Column 4), the sub-tasks they have participated in (Column 1), and the paper they have contributed (Column 5). Teams marked with a (**) include some of the SemEval 2016 Task 4 organizers.