

**Supplementary Material of the paper
“Retrieval of the Best Counterargument
without Prior Topic Knowledge”**

Corpus and Source Code

Corpus and source code of all experiments are available at: <http://www.arguana.com>

Validation Set Results

Found in Table 1 on the next page.

Test Set Results per Theme

Found in Tables 2 and 3 on the third page.

#	Baseline / Approach	Same Debate				Same Theme	
		Opp. Ctr.'s	Counters	Opposing	Arguments	Counters	Arguments
w	Word argument similarity	68.1	49.3	42.7	32.7	46.2	30.6
e	Embedding argument similarity	64.3	46.5	52.8	38.4	41.5	35.3
w_{\downarrow}	Word unit similarity minimum	55.9	40.6	48.6	36.8	31.3	28.7
w_{\uparrow}	Word unit similarity maximum	68.5	48.9	44.0	32.0	46.1	30.5
w_{\times}	Word unit similarity product	66.4	51.4	58.3	45.2	48.1	42.0
w_{+}	Word unit similarity sum	73.3	55.5	55.2	42.7	53.2	40.3
e_{\downarrow}	Embedding unit similarity minimum	64.2	46.9	47.0	35.5	41.2	30.8
e_{\uparrow}	Embedding unit similarity maximum	67.5	48.7	51.2	37.1	43.7	33.4
e_{\times}	Embedding unit similarity product	72.1	55.2	56.8	43.3	51.8	40.3
e_{+}	Embedding unit similarity sum	72.1	55.1	56.9	43.1	51.8	40.2
	$1.0 \cdot (w_{+} + e_{+})$	72.0	52.4	51.1	38.3	49.6	36.9
	$1.0 \cdot (w_{\downarrow} + e_{\downarrow})$	66.4	49.9	54.8	42.7	42.0	37.1
	$1.0 \cdot (w_{\uparrow} + e_{\uparrow})$	72.1	53.2	50.1	36.4	50.3	35.7
we	$1.0 \cdot (w_{\times} + e_{\times})$	72.8	57.7	62.1	48.5	53.7	44.9
	$1.0 \cdot (w_{+} + e_{+})$	75.4	58.2	58.4	44.8	56.5	43.5
	$0.9 \cdot (w_{\downarrow} + e_{\downarrow}) - 0.1 \cdot (w_{\uparrow} + e_{\uparrow})$	60.1	44.0	51.4	39.0	35.7	32.1
	$0.9 \cdot (w_{\downarrow} + e_{\downarrow}) - 0.1 \cdot (w_{\times} + e_{\times})$	64.6	47.8	53.8	41.1	40.0	35.3
	$0.9 \cdot (w_{\downarrow} + e_{\downarrow}) - 0.1 \cdot (w_{+} + e_{+})$	63.1	46.5	53.0	40.3	38.7	34.2
	$0.9 \cdot (w_{\uparrow} + e_{\uparrow}) - 0.1 \cdot (w_{\downarrow} + e_{\downarrow})$	71.3	51.6	49.0	35.2	48.8	34.0
	$0.9 \cdot (w_{\uparrow} + e_{\uparrow}) - 0.1 \cdot (w_{\times} + e_{\times})$	70.9	52.0	49.1	35.4	49.0	34.0
	$0.9 \cdot (w_{\uparrow} + e_{\uparrow}) - 0.1 \cdot (w_{+} + e_{+})$	71.6	52.1	49.6	36.1	49.3	35.0
we_{\downarrow}	$0.9 \cdot (w_{\times} + e_{\times}) - 0.1 \cdot (w_{\downarrow} + e_{\downarrow})$	73.0	57.7	61.9	49.0	54.3	45.2
	$0.9 \cdot (w_{\times} + e_{\times}) - 0.1 \cdot (w_{\uparrow} + e_{\uparrow})$	71.2	55.2	60.9	48.0	51.1	44.6
	$0.9 \cdot (w_{\times} + e_{\times}) - 0.1 \cdot (w_{+} + e_{+})$	72.3	56.6	61.9	48.9	52.4	44.5
we_{\uparrow}	$0.9 \cdot (w_{+} + e_{+}) - 0.1 \cdot (w_{\downarrow} + e_{\downarrow})$	75.5	58.0	57.3	43.4	56.1	42.1
	$0.9 \cdot (w_{+} + e_{+}) - 0.1 \cdot (w_{\uparrow} + e_{\uparrow})$	75.5	58.7	59.6	46.8	56.6	45.7
	$0.9 \cdot (w_{+} + e_{+}) - 0.1 \cdot (w_{\times} + e_{\times})$	75.1	58.0	57.7	43.9	56.2	42.4
	$0.8 \cdot (w_{\downarrow} + e_{\downarrow}) - 0.2 \cdot (w_{\uparrow} + e_{\uparrow})$	50.7	36.0	43.7	32.6	27.0	24.6
	$0.8 \cdot (w_{\downarrow} + e_{\downarrow}) - 0.2 \cdot (w_{\times} + e_{\times})$	60.8	44.8	51.1	39.0	36.4	31.9
	$0.8 \cdot (w_{\downarrow} + e_{\downarrow}) - 0.2 \cdot (w_{+} + e_{+})$	58.7	42.7	50.2	37.8	33.6	30.5
	$0.8 \cdot (w_{\uparrow} + e_{\uparrow}) - 0.2 \cdot (w_{\downarrow} + e_{\downarrow})$	68.7	49.7	46.6	34.8	46.5	32.3
	$0.8 \cdot (w_{\uparrow} + e_{\uparrow}) - 0.2 \cdot (w_{\times} + e_{\times})$	68.9	49.6	46.2	33.8	46.7	31.6
	$0.8 \cdot (w_{\uparrow} + e_{\uparrow}) - 0.2 \cdot (w_{+} + e_{+})$	71.2	51.4	48.4	35.0	48.6	33.6
	$0.8 \cdot (w_{\times} + e_{\times}) - 0.2 \cdot (w_{\downarrow} + e_{\downarrow})$	73.9	58.4	61.9	48.4	54.9	45.1
	$0.8 \cdot (w_{\times} + e_{\times}) - 0.2 \cdot (w_{\uparrow} + e_{\uparrow})$	66.4	50.7	57.8	45.1	44.7	40.4
	$0.8 \cdot (w_{\times} + e_{\times}) - 0.2 \cdot (w_{+} + e_{+})$	71.2	54.9	61.1	48.2	50.7	44.5
	$0.8 \cdot (w_{+} + e_{+}) - 0.2 \cdot (w_{\downarrow} + e_{\downarrow})$	74.4	57.3	55.2	42.2	54.6	40.6
	$0.8 \cdot (w_{+} + e_{+}) - 0.2 \cdot (w_{\uparrow} + e_{\uparrow})$	74.9	58.4	61.0	47.1	55.9	45.0
	$0.8 \cdot (w_{+} + e_{+}) - 0.2 \cdot (w_{\times} + e_{\times})$	74.8	57.6	56.2	42.2	54.9	40.9
r	Random baseline	26.4	13.4	13.4	7.2	0.7	0.3

Table 1: **Validation set accuracy** of ranking the best counterargument highest for several baselines and approaches (w , e , w_{\downarrow} , \dots , r) in the first six of the eight retrieval tasks (the last two were not evaluated on the validation set). The best accuracy values in each task are marked in bold. Aside from the ten baselines at the top and the random baseline at the bottom, we chose the three approaches (we , we_{\downarrow} , we_{\uparrow}) for the test set experiments that were best in the six tasks. We left out the second one that was best in the first task, because — as can be seen — it performed notably worse in most other tasks.

Theme	Same Debate: Opp. Ctr.'s				Same Debate: Counters				Same Debate: Opposing				Same Debate: Arguments			
	Acc'y@1		M. Rank		Acc'y@1		M. Rank		Acc'y@1		M. Rank		Acc'y@1		M. Rank	
	w_+	we_{\uparrow}	w_+	we_{\uparrow}	w_+	we_{\uparrow}	w_+	we_{\uparrow}	w_{\times}	we	w_{\times}	we	e_{\times}	we	e_{\times}	we
Culture	73.9	78.3	1	1	50.7	62.3	2	2	50.7	60.9	3	2	47.8	50.7	2	3
Digital freedoms	75.4	73.8	1	1	54.1	55.7	2	2	59.0	60.7	2	2	44.3	45.9	4	3
Economy	73.6	76.8	1	1	54.4	59.2	2	2	59.2	59.2	2	2	41.6	42.4	3	3
Education	75.3	77.8	1	1	59.3	64.2	2	2	66.7	64.2	2	2	39.5	49.4	3	3
Environment	65.2	67.4	2	1	52.2	50.0	2	2	60.9	69.6	3	2	39.1	47.8	3	2
Free speech debate	63.8	69.0	2	1	50.0	44.8	2	2	39.7	46.6	3	3	24.1	25.9	4	5
Health	67.5	66.2	2	1	49.4	54.6	2	2	48.1	61.0	3	2	42.9	45.5	4	3
International	70.9	75.3	1	1	52.4	57.6	2	2	56.1	60.9	2	2	41.7	45.0	3	3
Law	73.9	76.9	1	1	53.7	53.0	2	2	59.7	63.4	2	2	37.3	44.8	3	3
Philosophy	65.9	69.4	2	1	50.6	52.9	2	2	51.8	52.9	3	2	48.2	41.2	3	3
Politics	75.3	76.7	1	1	56.4	60.4	2	2	55.9	60.4	2	2	36.6	44.6	4	3
Religion	71.1	77.8	1	1	57.8	68.9	2	2	66.7	71.1	2	2	53.3	62.2	3	2
Science	73.7	75.4	1	1	57.9	61.4	2	2	59.7	64.9	2	2	49.1	45.6	3	3
Society	66.7	68.3	1	1	50.0	58.3	2	2	45.0	48.3	3	2	41.7	43.3	3	4
Sport	66.7	80.0	2	1	56.7	63.3	2	2	63.3	63.3	3	2	33.3	46.7	4	3
All themes	71.5	74.5	1	1	53.7	57.7	2	2	56.1	60.3	2	2	41.0	44.9	3	3

Table 2: Test set accuracy@1 and mean rank of the best baseline (w_+ , w_{\times} , or e_{\times}) compared to the best approach (we_{\uparrow} or we) for all rankings of each theme on the 1st to 4th retrieval task.

Theme	Same Theme: Counters				Same Theme: Arguments				Entire Portal: Counters				Entire Portal: Arguments			
	Acc'y@1		M. Rank		Acc'y@1		M. Rank		Acc'y@1		M. Rank		Acc'y@1		M. Rank	
	w_+	we_{\uparrow}	w_+	we_{\uparrow}	e_{\times}	we_{\downarrow}	e_{\times}	we_{\downarrow}	w_+	we_{\uparrow}	w_+	we_{\uparrow}	w_+	we_{\uparrow}	w_+	we_{\uparrow}
Culture	50.7	60.9	3	2	47.8	46.4	3	4	47.8	55.1	18	6	31.9	36.2	12	9
Digital freedoms	47.5	54.1	2	2	37.7	45.9	7	7	44.3	52.5	7	5	37.7	44.3	58	20
Economy	52.0	56.8	5	4	36.0	38.4	6	8	48.8	52.0	38	15	27.2	25.6	21	10
Education	53.1	58.0	3	3	37.0	46.9	7	7	49.4	55.6	11	5	38.3	39.5	36	17
Environment	47.8	47.8	4	3	37.0	47.8	4	3	41.3	52.2	18	9	17.4	21.7	22	7
Free speech debate	34.5	37.9	3	2	22.4	22.4	9	11	31.0	27.6	9	4	10.3	12.1	130	55
Health	48.1	53.3	5	4	40.3	36.4	5	6	39.0	45.5	63	27	28.6	36.4	26	14
International	44.7	50.6	3	3	35.1	36.9	9	11	40.6	46.9	14	8	25.8	31.4	31	19
Law	49.3	50.8	5	4	31.3	42.5	6	6	44.0	41.8	16	11	38.8	38.1	16	8
Philosophy	48.2	50.6	4	3	45.9	40.0	5	6	43.5	49.4	9	5	34.1	38.8	29	14
Politics	52.5	58.4	3	2	34.7	42.6	5	8	48.0	55.5	18	9	28.7	33.2	28	11
Religion	53.3	60.0	4	3	46.7	51.1	4	4	48.9	55.6	15	6	24.4	33.3	58	8
Science	56.1	63.2	3	2	47.4	47.4	3	3	54.4	61.4	29	5	19.3	28.1	6	5
Society	48.3	55.0	2	2	40.0	41.7	4	6	43.3	50.0	4	3	16.7	20.0	45	22
Sport	56.7	63.3	3	2	33.3	43.3	4	4	53.3	60.0	23	11	43.3	46.7	35	9
All themes	49.0	54.2	4	3	37.1	41.0	6	7	44.7	50.0	17	9	28.6	32.4	33	15

Table 3: Test set accuracy@1 and mean rank of the best baseline (w_+ or e_{\times}) compared to the best approach (we_{\uparrow} or we_{\downarrow}) for all rankings of each theme on the 5th to 8th retrieval task.