

lang	unique tokens	tokens
en	4.238	902.044
de	4.233	312.380
fr	1.749	247.492
ru	1.766	202.163
it	1.411	178.378
nl	2.021	105.224
pt	0.873	96.712
sv	1.730	82.376
es	0.829	76.587
ar	1.653	46.481
fa	0.925	32.557