

<b>Datasets</b>	<b>TWEETS</b>	<b>GN</b>	<b>YELP</b>
# of msg	12.7M	2.2M	1.6M
Vocabulary size	1.7M	208K	85K
Avg. # of msg keywords	9	7	37
Size in GB	2.26	0.3	1.04