

=== TREC Dataset Results ===

	original_size_bytes	gzip_size_bytes	lzw_size_bytes	llm_bits_total	llm_bits_per_token
count	5452.000000	5452.000000	5452.000000	5452.000000	5452.000000
mean	50.632062	67.119222	50.448092	45.423331	4.155984
std	20.743208	16.499787	19.287510	20.399941	1.508103
min	13.000000	33.000000	11.375000	0.000000	0.000000
25%	36.000000	55.000000	37.125000	32.000000	3.076923
50%	48.000000	66.000000	48.375000	40.000000	4.000000
75%	61.000000	77.000000	60.750000	56.000000	5.142857
max	196.000000	155.000000	164.250000	192.000000	11.000000

=== MedQuAD Dataset Results ===

	original_size_bytes	gzip_size_bytes	lzw_size_bytes	llm_bits_total	llm_bits_per_token
count	16407.000000	16407.000000	16407.000000	16407.000000	16407.000000
mean	50.684952	67.827513	51.158339	48.875480	4.078711
std	16.926465	13.449138	16.027120	17.097808	1.500853
min	16.000000	36.000000	14.000000	8.000000	0.400000
25%	38.000000	58.000000	40.500000	36.000000	3.076923
50%	48.000000	67.000000	49.500000	48.000000	4.000000
75%	61.000000	77.000000	60.750000	60.000000	4.800000
max	191.000000	158.000000	166.500000	156.000000	11.428571

=== CaseHold Dataset Results ===

	original_size_bytes	gzip_size_bytes	lzw_size_bytes	llm_bits_total	llm_bits_per_token
count	30612.000000	30612.000000	30612.000000	30612.000000	30612.000000
mean	129.843558	109.467006	113.725002	86.782308	4.060649
std	38.623308	20.343574	28.676497	32.082202	1.349260
min	12.000000	32.000000	10.500000	12.000000	0.400000
25%	101.000000	94.000000	92.250000	64.000000	3.111111
50%	131.000000	111.000000	114.750000	84.000000	4.000000
75%	161.000000	126.000000	136.125000	104.000000	4.857143
max	199.000000	160.000000	181.125000	292.000000	12.000000