Difference in CPT compared to the tokenizer with alpha=1.0 Alpha 1.25 0.0 0.3 1.00 0.5 0.7 0.75 -1.0 0.50 -0.25 0.00 -0.25-0.50en vi ru fr de es th bg he zh el tr ar hi ta ka ur te mrsw Language

Characters per token