Corpus Size by Category (Billion Tokens)

