



	Term	!term	
category	A	C	G
!category	B	D	H
	E	F	N

Table adapted from slide 58 of Module 2.

Preprocessing is conducted using PreprocessingMapper. E and F is computed using TermReducer, while G is computed using CategoryCountReducer. The main task lies within CategoryReducer. It receives category, [term] pairs and computes A. It then computes C, B and D using E, F and G. Then, chi square is computed for all terms and the top 200 terms are emitted alphabetically.