COMP 479 - Project 2

François LaBerge

October 20, 2020

	(distinct) terms			nonpositional postings			tokens		
	number	Δ%	T%	number	Δ%	T%	number	Δ%	T%
unfiltered	92,317			1,789,665			3,010,306		
no numbers	66,649	-39		1,664,299			1,664,299		
case folding	50,267	-33		1,546,832			1,546,832		
30 stop words	50,239			1,345,802			1,345,802		
150 stop words	50,132			1,222,350			1,222,350		
stemming	39,406			1,218,845			1,218,845		