COMP 479 - Project 2

François LaBerge

October 22, 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	90,860	-2	-2	1,737,383	-3	-3	2,948,054	-2	-2
case folding	74,410	-22	-24	1,619,756	-7	-10	2,948,054	-0	-2
30 stop words	74,381	-0	-24	1,414,678	-14	-27	2,261,310	-30	-33
150 stop words	74,261	-0	-24	1,279,759	-11	-40	2,068,086	-9	-46
stemming	63,444	-17	-45	1,223,686	-5	-46	2,068,086	-0	-46

no numbers 66,649 -39   1,664,299   2,868,819