

*****KSU scrapy webpage statistics*****

average doc len: 404.743565300286

most frequent emails:

('ksugrad@kennesaw.edu', 33)

('lgoodnow@kennesaw.edu', 12)

('epeter54@kennesaw.edu', 12)

('colescss@kennesaw.edu', 11)

('wellstaradvising@kennesaw.edu', 10)

('dmarsh6@kennesaw.edu', 10)

('honors@kennesaw.edu', 9)

('ksuinfo@kennesaw.edu', 8)

('gradcollegeforms@kennesaw.edu', 8)

('fparham2@kennesaw.edu', 7)

percent webpages with emails: 0.2840800762631077

Top 30 Most Frequent Words from Raw Text

rank	term	freq.	perc.	rank	term	freq.	perc.
1	and	13979	0.033	16	State	2326	0.005
2	the	13511	0.032	17	Marietta	2257	0.005
3	of	11404	0.027	18	on	2147	0.005
4	to	9389	0.022	19	GA	2089	0.005
5	in	6295	0.015	20	you	1985	0.005
6	a	5434	0.013	21	470-KSU-INFO	1944	0.005
7	for	4411	0.010	22	(470-578-4636)	1942	0.005
8	Kennesaw	3470	0.008	23	are	1871	0.004
9	Resources	3104	0.007	24	at	1815	0.004
10	Campus	3094	0.007	25	will	1810	0.004
11	is	2975	0.007	26	that	1776	0.004
12	KSU	2684	0.006	27	students	1709	0.004
13	The	2606	0.006	28	as	1708	0.004
14	For	2431	0.006	29	be	1703	0.004
15	with	2397	0.006	30	or	1642	0.004

Top 30 Most Frequent Words After Stopword/Punctuation Removal

rank	term	freq.	perc.	rank	term	freq.	perc.
1	Kennesaw	4573	0.011	16	Education	1055	0.002
2	Marietta	3290	0.008	17	South	1037	0.002
3	Campus	3168	0.007	18	1000	1030	0.002
4	Resources	3110	0.007	19	1100	1021	0.002
5	KSU	2935	0.007	20	Contact	1020	0.002
6	State	2382	0.006	21	Media	1012	0.002
7	students	2252	0.005	22	Rights	1012	0.002
8	University	2217	0.005	23	30060	1009	0.002
9	GA	2111	0.005	24	Road	1004	0.002
10	470KSUINFO	1944	0.005	25	Chastain	1001	0.002
11	4705784636	1944	0.005	26	Reserved	995	0.002
12	Phone	1788	0.004	27	Pkwy	993	0.002
13	College	1494	0.004	28	Links	991	0.002
14	2022	1431	0.003	29	Follow	989	0.002
15	30144	1057	0.002	30	Info	980	0.002

