

LEXICAL SYNTAX ANALYSIS FOR HOSTILE BEHAVIOUR IDENTIFICATION

Akshath Jain

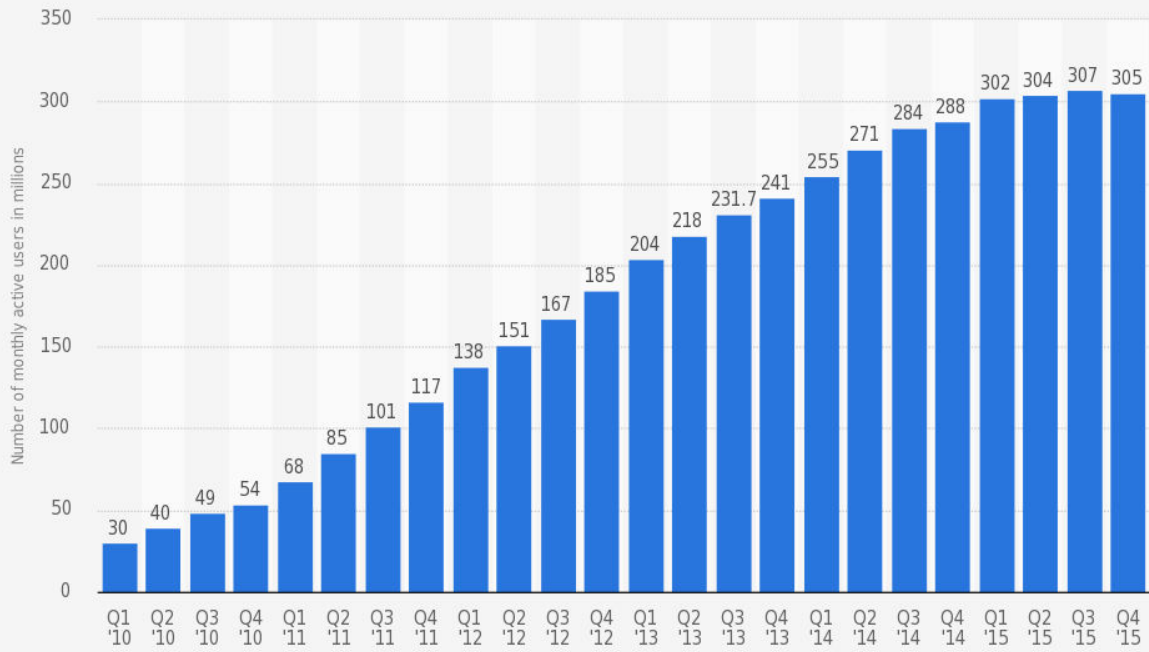
ISEF/PRSEF Data Book

April 1, 2016

## Table of Contents

Number of Monthly Active Twitter Users.....	3
Number of Fatalities Due to Terrorist Attacks Worldwide .....	4
Number of Common Word Matches.....	5
Percentage of Common Word Matches .....	6
Average Placement of Common Word Matches .....	7
Percent Affiliation.....	8
Training Set.....	9
Validation Set.....	10
Trial 1 Results .....	11
Trial 2 Results .....	13
Trial 3 Results.....	15

### Number of monthly active Twitter users worldwide from 1st quarter 2010 to 4th quarter 2015 (in millions)

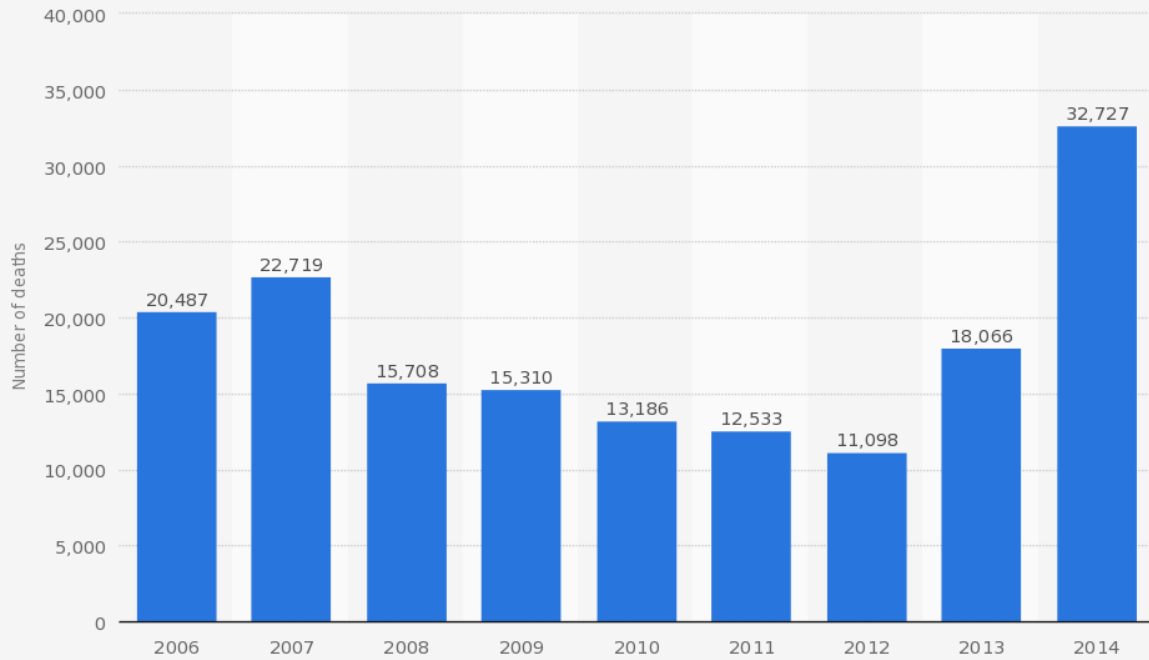


Source:  
Twitter  
© Statista 2016

Additional Information:  
Worldwide; Twitter, 1st quarter 2010 to 4th quarter 2015; excluding SMS  
fast followers



## Number of fatalities due to terrorist attacks worldwide between 2006 and 2014

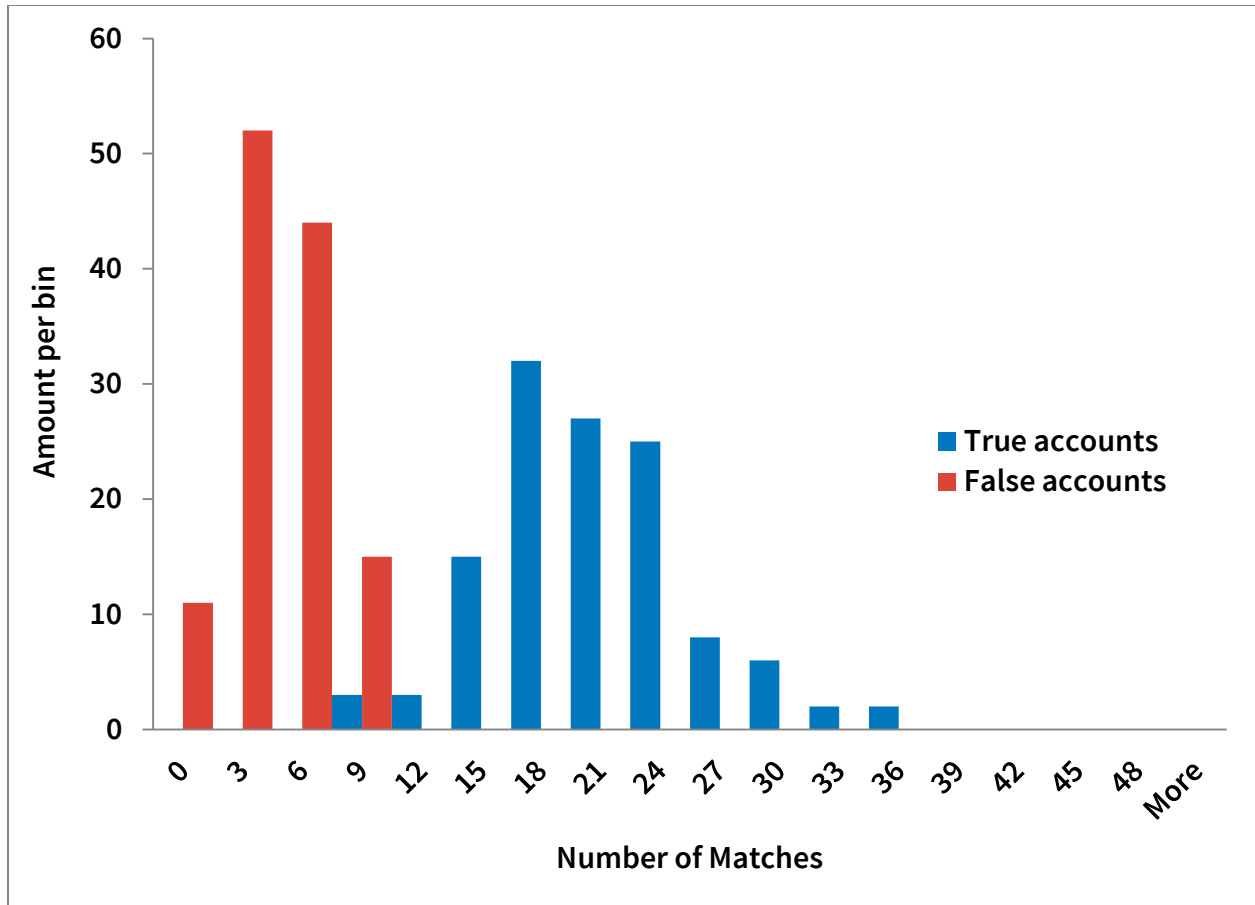


Source:  
NCTC  
© Statista 2015

Additional Information  
Worldwide; NCTC

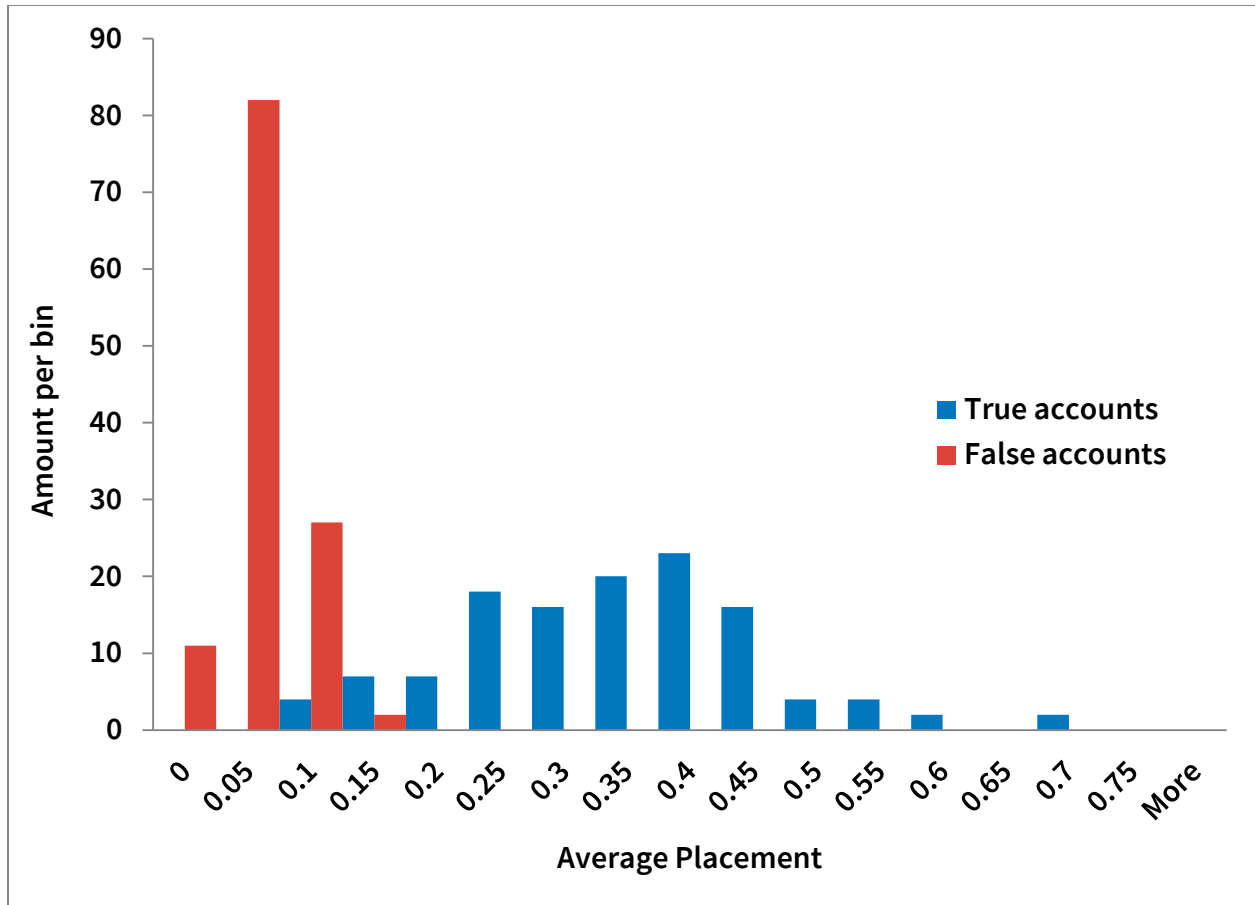
statista

### Number of Common Word Matches



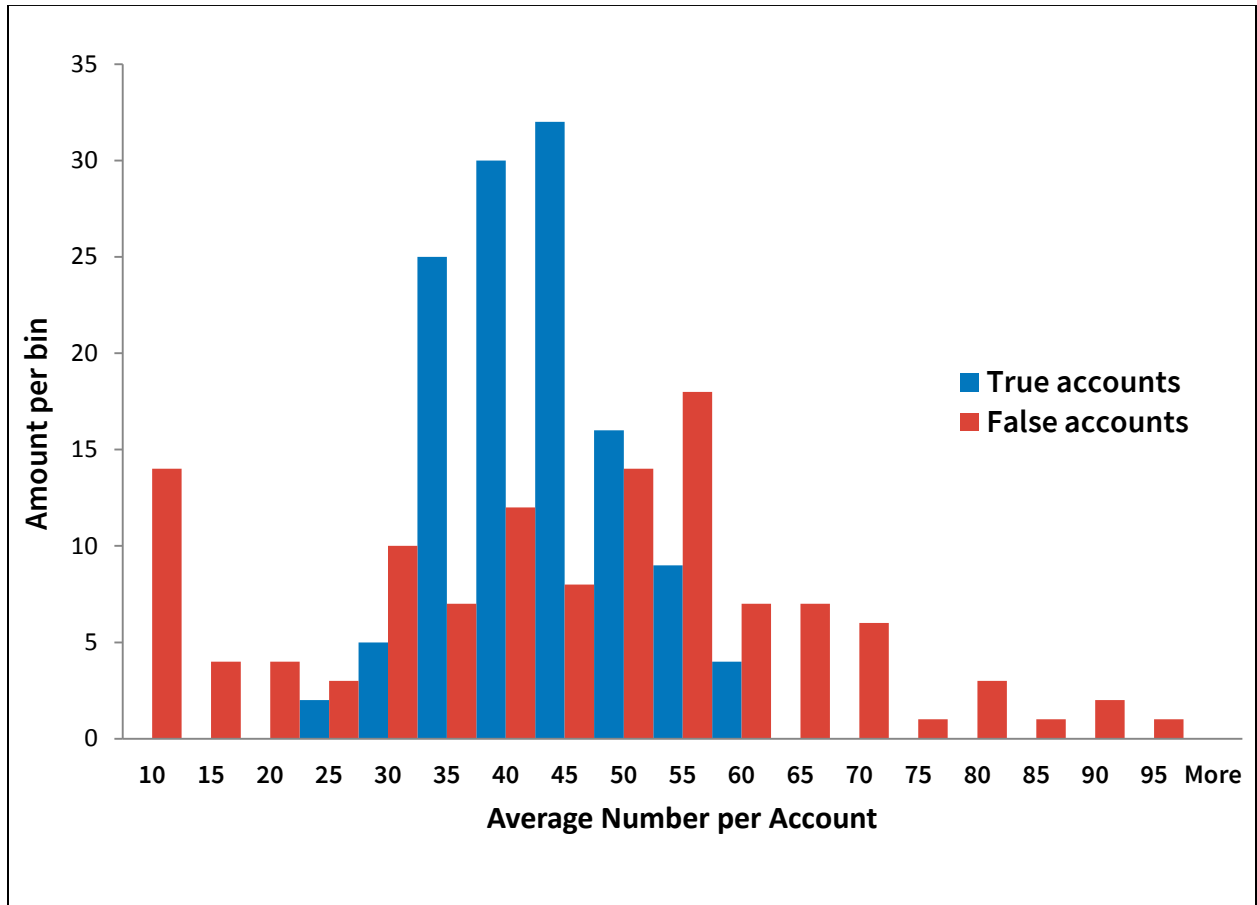
	True	False
<b>Minimum</b>	8	0
<b>Maximum</b>	35	9
<b>Average</b>	19.862	3.639

### Percentage of Common Word Matches



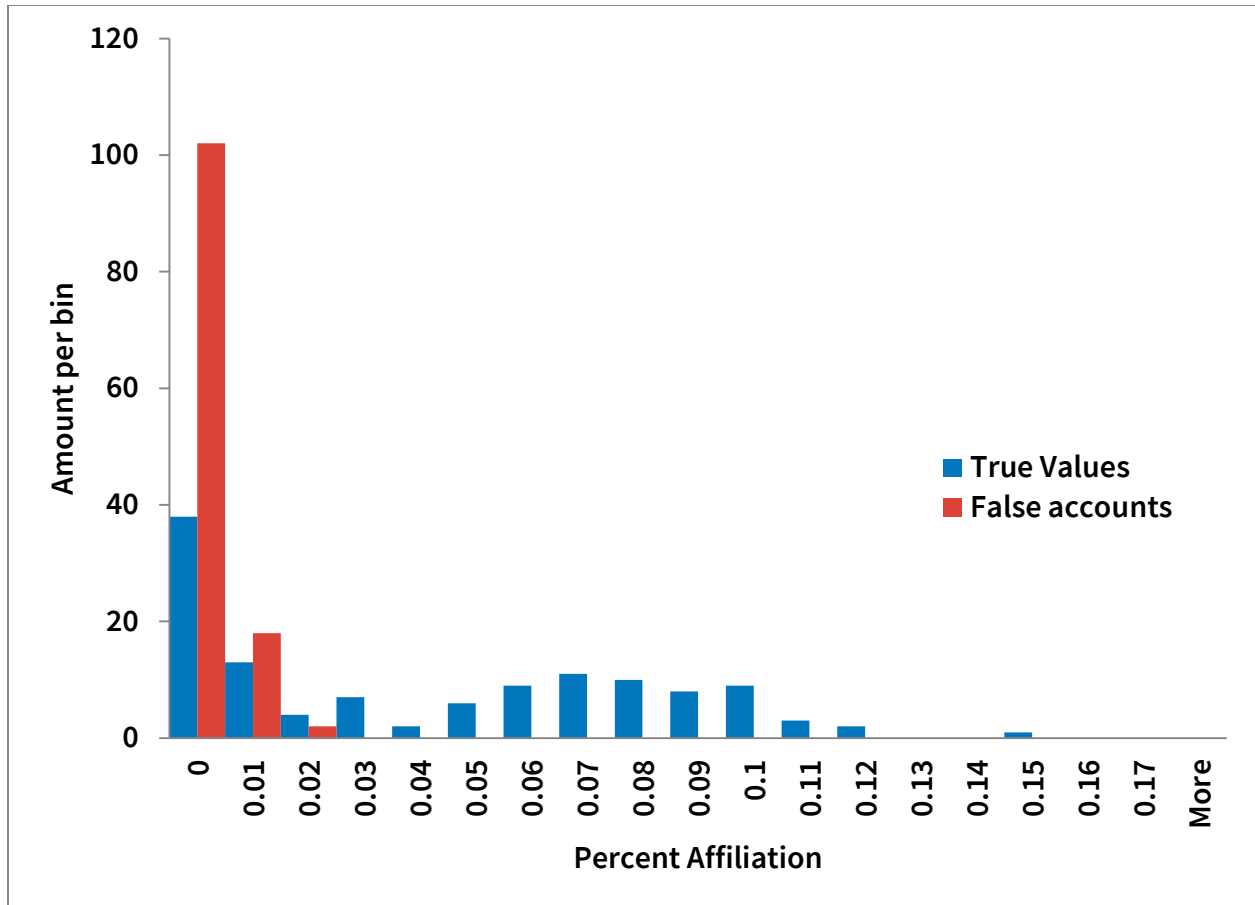
	True	False
<b>Minimum</b>	.057	0
<b>Maximum</b>	.674	.100
<b>Average</b>	.319	.037

### Average Placement of Common Word Matches



	True	False
<b>Minimum</b>	23.846	0
<b>Maximum</b>	59.333	95
<b>Average</b>	40.262	40.992

### Percent Affiliation



	True	False
<b>Minimum</b>	0	0
<b>Maximum</b>	.141	.011
<b>Average</b>	.039	.001



	<b>Training Set</b>	Affiliation	Number of Common Word Matches	Percentage of Common Word Matches	Placement of Common Words Matches
TRUE	AZCardinals	0.073818898	18	0.285714286	49.33333333
	Giants	0.069301848	20	0.334693878	32.6
	Colts	0.043859649	15	0.180555556	51.06666667
	Eagles	0.062642369	22	0.335008375	34.04545455
	Jaguars	0.068933824	17	0.188349515	42.88235294
	Panthers	0.002314815	20	0.355895197	32.95
	ChicagoBears	0.10019084	20	0.316981132	53.95
	Lions	0.025322284	23	0.37944664	41.69565217
	49ers	0.067849687	16	0.322429907	28.6875
	Browns	0.04509018	23	0.358139535	38.7826087
	br_lions	0.002252252	15	0.25093985	51.73333333
	br_49ers	0	14	0.291176471	40
	br_clebrowns	0	18	0.268072289	49.72222222
	br_miadolphins	0	18	0.208333333	35.11111111
	stillers	0	14	0.145181477	41.14285714
FALSE	aamir_khan	0	4	0.063793103	19
	beingsalmankhan	0	6	0.092391304	33.5
	adele	0	4	0.05090312	26
	mariahcarey	0.003453039	6	0.047348485	59.66666667
	xtina	0	9	0.077625571	54
	senbobcorker	0	5	0.035805627	63.6
	votemarsha	0	8	0.07816092	52.875
	3m	0	3	0.046099291	36.66666667
	rendezvousport	0	5	0.096091205	25.8
	tatacompanies	0	2	0.040880503	12.5
	generalelectric	0.002380952	2	0.037190083	16.5
	virgintrains	0	3	0.038366337	25
	unam	0	0	0	0
	preschoolers	0	4	0.024640657	83
	blakeshelton	0.010869565	6	0.074074074	51.16666667

	<b>Validation Set</b>	Affiliation	Number of Common Word Matches	Percentage of Common Word Matches	Placement of Common Words Matches
TRUE	Ravens	0.069852941	19	0.272012579	37.1052632
	MiamiDolphins	0.051070039	17	0.210526316	41
	steelers	0.089820359	17	0.223255814	36.2941176
	Titans	0.101941748	23	0.355072464	36.0869565
	Broncos	0.006830601	17	0.318	30.7058824
	HoustonTexans	0.073385519	23	0.234404537	48.3478261
	nyjets	0.073385519	23	0.234404537	48.3478261
	AtlantaFalcons	0.040944123	15	0.154855643	43.1333333
	RAIDERS	0.026595745	23	0.403474903	33.2173913
	dallascowboys	0.057581574	16	0.226691042	31.8125
	Chargers	0.017712551	22	0.232824427	42.2727273
	br_nfl	0.093555094	35	0.411764706	36.8857143
	br_nepatriots	0	15	0.289611752	44.6
	texansbuzztap	0.091575092	21	0.214285714	43.8571429
	br_azcardinals	0	24	0.473229706	29.5416667
FALSE	chrisbrown	0	4	0.032332564	46.75
	youtube	0	2	0.028391167	17
	kendalljenner	0.002753304	5	0.054347826	45
	agnezmo	0.002399232	3	0.027972028	41
	nasa	0	2	0.040841584	31
	crisantamx	0	0	0	0
	ivetesangalo	0	0	0	0
	beyonce	0	1	0.017857143	95
	ddlovato	0.002385496	5	0.064676617	24.4
	oprah	0.002192982	4	0.029668412	67.5
	drake	0.00246063	3	0.023060797	51.3333333
	realmadrid	0	0	0	0
	sismologicoMX	0	0	0	0
	google	0	3	0.034693878	50.6666667
	3mnewsroom	0.002777778	3	0.031434185	38.3333333

	<b>Trial 1 Results</b>	Affiliation	Number of Common Word Matches	Percentage of Common Word Matches	Placement of Common Words Matches	Correct Classification?
TRUE	br_nycgiants	0	18	0.285714286	49.61111111	1
	br_pittsteelers	0	25	0.393129771	42.4	1
	ramsbuzztap	0.024900398	21	0.20935412	46.38095238	1
	jaguarsbuzztap	0.081521739	22	0.23222061	49.09090909	1
	bengalsbuzztap	0.05	21	0.417413572	33.66666667	1
	cowboysbuzztap	0.072727273	19	0.447606727	31.73684211	1
	eaglesbuzztap	0.0625	21	0.480952381	36.71428571	1
	steelersbuzztap	0.094765343	24	0.422900763	40.75	1
	raidersbuzztap	0.063176895	19	0.211484594	39.47368421	1
	falconsbuzztap	0.082684825	18	0.227456258	42	1
	stillers	0	29	0.570324575	33.86206897	1
	nfl	0.046958855	24	0.312883436	39.41666667	1
	azcardsfanclub	0.002880184	29	0.480894688	34.48275862	1
	Redskins	0.017258383	24	0.393474088	39.70833333	1
	TBBuccaneers	0.032630522	24	0.373376623	33.54166667	1
	packers	0.109589041	24	0.424561404	36.20833333	1
	Chiefs	0.068044355	18	0.234234234	32.38888889	1
	Saints	0.022321429	16	0.259469697	53.5	1
	Patriots	0.057645631	25	0.398876404	36.44	1
	Vikings	0.092412451	26	0.401574803	40.30769231	1
	buffalobills	0.052521008	22	0.3655706	39.54545455	1
	Bengals	0.059255079	22	0.375987362	37.59090909	1
	RamsNFL	0.018621974	17	0.186500888	44.64705882	1
	Seahawks	0.008960573	18	0.271484375	51.44444444	1
	redskinsbuzztap	0.096525097	19	0.391737892	39.57894737	1
	bearsbuzztap	0.084519573	19	0.394242804	40.63157895	1
	49ersbuzztap	0.081730769	19	0.380597015	41.78947368	1
	giantsbuzztap	0.062260536	22	0.399759904	38.09090909	1
	VikingsFanClub	0.003056235	16	0.33649698	35.25	1
	buffalobills2	0	28	0.415492958	36.75	1
cinbengals	0.008680556	15	0.339597315	45.2	1	
dallascowboys	0.057581574	16	0.226691042	31.8125	1	
FALSE	harvardhealth	0	1	0.007598784	47	1
	cnn	0.002853881	6	0.045620438	68.66666667	1

harvad	0.003753754	0	0	0	1
ipn_mx	0	1	0.00811359	26	1
globaltimesnews	0	2	0.032751092	11	1
khanacademy	0	7	0.087102178	37.85714286	1
itbofficial	0	2	0.011520737	66.5	1
goldsilverclub	0	4	0.045454545	42.75	1
facebook	0	4	0.026706231	40	1
barackobama	0	7	0.049586777	63.71428571	1
repronpaul	0	1	0.011090573	28	1
robert_aderholt	0	4	0.043875686	50.25	1
judgetedpoe	0	5	0.055062167	33	1
scansnapit	0	0	0	0	1
cnn	0.002840909	7	0.051660517	74.85714286	1
real_liam_payne	0	4	0.037848606	48.5	1
actuallynph	0	8	0.082608696	52.875	1
pitbull	0	6	0.04496124	51.66666667	1
louis_tomlinson	0	7	0.05787037	57	1
danieltosh	0.005070994	5	0.077399381	27.2	1
neymarjr	0	2	0.013245033	58.5	1
emwatson	0	3	0.03057554	27.66666667	1
jrridinger	0	5	0.046683047	48.8	1
bbcbreaking	0	1	0.01686747	9	1
avrillavigne	0	4	0.029182879	78.25	1
conanobrien	0	6	0.075324675	40.5	1
davidguetta	0	6	0.048364154	47.5	1
srbachchan	0	2	0.019083969	31	1
liquidnewsroom	0	4	0.062761506	45.25	1
realmadrid	0	0	0	0	1
crisantamx	0	0	0	0	1
oprah	0.002192982	4	0.029668412	67.5	1

	<b>Trial 2 Results</b>	Affiliation	Number of Common Word Matches	Percentage of Common Word Matches	Placement of Common Words Matches	Correct Classification?
--	------------------------	-------------	-------------------------------	-----------------------------------	-----------------------------------	-------------------------

TRUE	lawrencetimmons	0.117728532	8	0.070278184	51.375	0
	steelers_fan	0.002983294	21	0.352647353	42.85714286	1
	Steelers__News	0	22	0.515151515	43	1
	btsteelcurtain	0.042325056	28	0.534328358	33.82142857	1
	pats_superbowl	0	16	0.238546603	42.75	1
	patriots__news	0	19	0.320193638	45	1
	pgpackersnews	0.037553648	21	0.367830424	42	1
	jetsbuzztap	0.072957198	22	0.291108404	34.77272727	1
	bucsbuzztap	0.057692308	15	0.341426404	24.26666667	0
	brownsbuzztap	0.058876812	19	0.316397229	42.78947368	1
	titansbuzztap	0.064484127	26	0.44625	45.73076923	1
	dolphinsbuzztap	0.09580292	18	0.177419355	50.77777778	1
	chargersbuzztap	0.061787072	17	0.202162162	35.76470588	1
	vikingsbuzztap	0.092936803	13	0.341059603	23.84615385	0
	Seahawks__News	0.004251701	21	0.329071332	56.0952381	1
	tenntitans_fans	0	18	0.591715976	30.61111111	1
	NYJetsSupporter	0	17	0.138790036	33.17647059	1
	spot_cardinals	0	24	0.407720145	45.41666667	1
	azcardinalscom	0.003125	26	0.436285097	29.26923077	1
	cardinalspress	0	32	0.673594132	33.625	1
	track_giants	0.059375	17	0.540731995	25.64705882	1
	eaglesinsider	0.083102493	16	0.269779507	41.375	1
	bdcpatriots	0	24	0.309462916	40.54166667	1
	dbackfan4lyfe	0.002129472	17	0.271304348	35.35294118	1
	panthersbuzztap	0.076992754	22	0.345120226	39.54545455	1
	packersbuzztap	0.072727273	20	0.351804124	35.05	1
	lionsbuzztap	0.096507353	19	0.352272727	37.47368421	1
	br_indcolts	0	13	0.08496732	53.46153846	1
	br_jaxjaguars	0	12	0.056807052	59	0
	NEPFCA	0.021008403	17	0.365942029	35.11764706	0
BroncosFanClub	0.024509804	10	0.226443769	34.5	0	
br_nepatriots	0	15	0.289611752	44.6	1	
FALSE	katyperry	0	2	0.017492711	47	1
	sony	0	9	0.10041841	54.44444444	1
	ladygaga	0	5	0.051823417	35.2	1
	theellenshow	0.005263158	6	0.062052506	47.83333333	1
	twitter	0	8	0.03030303	79.375	1
	jtimberlake	0.005530973	3	0.023529412	57	1
	csd	0	5	0.046511628	46.6	1
	britneyspears	0	9	0.073921971	51.77777778	1
	instagram	0	1	0.018867925	12	1

cristiano	0.010653409	8	0.083673469	47.375	1
selenagomez	0	7	0.070397112	45.85714286	1
shakira	0	1	0.008522727	31	1
arianagrande	0	1	0.005714286	87	1
jlo	0	6	0.057692308	39	1
jimmyfallon	0	5	0.042647059	47.6	1
cnnbrk	0	3	0.054098361	29	1
windows	0	3	0.033602151	40	1
gmail	0	2	0.011857708	62.5	1
skype	0	1	0.008547009	49	1
pink	0	6	0.06824147	41.83333333	1
billgates	0	2	0.034285714	17	1
mileycyrus	0	3	0.041602465	37	1
liltunechi	0.004770992	2	0.014423077	61	1
johnhering	0	3	0.03652968	34	1
globaltimesnews	0	2	0.031380753	11	1
nytimes	0	1	0.032482599	3	1
brunomars	0	3	0.03	51.66666667	1
nialloficial	0	6	0.043209877	58.83333333	1
aliciakeys	0	4	0.033766234	64.25	1
jonoringer	0	7	0.087078652	28	1
sismologicoMX	0	0	0	0	1
google	0	3	0.034693878	50.66666667	1

	<b>Trial 3 Results</b>	Affiliation	Number of Common Word	Percentage of Common Word Matches	Placement of Common Words Matches	Correct Clasification?
--	------------------------	-------------	-----------------------	-----------------------------------	-----------------------------------	------------------------

			Matches			
TRUE	AntonioBrown84	0.017301038	10	0.097686375	45.5	0
	Ike_SwagginU	0.079321663	9	0.100286533	35.22222222	0
	DolphinsFanSpot	0.003125	19	0.146701389	45.78947368	1
	chiefsbuzztap	0.05952381	18	0.181702668	43.77777778	1
	BR_Chiefs	0	15	0.122926829	41.2	1
	BR_BuffaloBills	0	25	0.401891253	37.68	1
	BR_CincyBengals	0	14	0.254995243	46.14285714	0
	BR_Vikings	0	20	0.324074074	38.9	1
	BR_Rams	0	13	0.123344371	34.23076923	0
	BR_Seahawks	0	17	0.350815851	38.88235294	1
	BR_Falcons	0	13	0.177172061	42.07692308	0
	tpolamalu	0.141402715	9	0.115334207	34.66666667	0
	broncosbuzztap	0.090648855	18	0.346036585	44.77777778	1
	AZCardsFanClub	0.002880184	29	0.480894688	34.48275862	1
	azcardsbuzztap	0.070754717	35	0.657030223	29.51428571	1
	BR_Packers	0	22	0.37037037	40.90909091	1
	BR_Redskins	0	18	0.25	55.5	1
	BR_Buccaneers	0	25	0.411684783	43	1
	BR_NOSaints	0	17	0.251008065	50.05882353	1
	br_panthers	0	19	0.356790123	33.21052632	1
	br_bears	0	24	0.425891182	31.45833333	1
	ravensbuzztap	0.086363636	19	0.2191582	41.36842105	1
	patriotsbuzztap	0.084519573	21	0.385255648	38.38095238	1
	seahawksbuzztap	0.076438849	21	0.341772152	46.33333333	1
	saintsbuzztap	0.11732852	15	0.310614525	33.06666667	1
	br_titans	0	18	0.279589935	59.33333333	1
	NFL_Sports_Talk	0	28	0.334328358	42.85714286	1
	EaglesNFLFans	0	25	0.437221728	49.32	1
	BrownsFanClub	0	32	0.503955696	42.125	1
	steelersdepot	0.021786492	21	0.409836066	34	1
	br_eagles	0	17	0.277237354	45.35294118	1
	HoustonTexans	0.073385519	23	0.234404537	48.34782609	1
FALSE	narendramodi	0	2	0.027372263	26.5	1
	iamsrk	0	5	0.048117155	35.4	1
	aplusk	0.002417795	5	0.079411765	23.6	1
	alex_butler	0	4	0.037037037	52.75	1
	edsheeran	0	4	0.040302267	53.25	1



coldplay	0	8	0.100486224	31	1
mariopicazo	0	0	0	0	1
firefox	0	2	0.009162304	86	1
sharethis	0	4	0.035346097	41.25	1
nagasaki	0	0	0	0	1
dropbox	0	3	0.02972973	25.33333333	1
googlemaps	0	2	0.022321429	53.5	1
photoshop	0	3	0.027088036	59.66666667	1
androiddev	0	6	0.047832586	49.66666667	1
googledrive	0	3	0.02200489	52.33333333	1
wordpressdotcom	0	1	0.007054674	64	1
geaviation	0	3	0.015469613	65.66666667	1
bayer	0	2	0.019017433	35.5	1
mahindraris	0	5	0.045655376	56	1
dsm	0	3	0.024934383	44.33333333	1
android	0	2	0.013618677	61	1
microsoft	0	5	0.045634921	51	1
googlechrome	0	6	0.046043165	51.33333333	1
leodicaprio	0	3	0.03220339	39	1
ryanseacrest	0.0025	3	0.020257827	67.66666667	1
skinnerliber8ed	0	1	0.023529412	5	1
parishilton	0	2	0.018907563	41.5	1
mtv	0	2	0.01497006	76	1
intel	0.009940358	7	0.047865459	51.28571429	1
razer	0	4	0.029239766	67.25	1
nvidia	0	4	0.042826552	36.5	1
hp	0	0	0	0	1