

# **A text mining approach to determine the minimum vocabulary needed by food science professionals who use English as a second language**

Peechaya Suriyawong<sup>1</sup>, Orapin Nasawang<sup>2</sup> and Jeffrey Nash<sup>3\*</sup>

<sup>1</sup> Lecturer in English Language, the Language Center, Udon Thani Rajabhat University, Udon Thani 41000, Thailand

<sup>2</sup> Consultant and Lecturer in English Language, the Language Center, Udon Thani Rajabhat University, Udon Thani 41000, Thailand

<sup>3</sup> Researcher, Materials Science Program, Department of Physics, Khon Kaen University, Khon Kaen 40002, Thailand

\* Corresponding Author's Email: jcn32@cornell.edu

## **Abstract**

English is the *lingua franca* of scientific and engineering research. As such, researchers must develop good ability to read technical English to stay current in their field and develop their careers. The current study was undertaken to determine the most commonly used vocabulary in a journal reporting research findings in the field of food science and technology. The approach is based upon an earlier linguistics study that examined the vocabulary needed to enable readers to understand text in a narrow context. The current study employed various R statistical packages to conduct data text mining on more than one thousand two hundred research papers. The papers were downloaded in pdf format from an online open source food science journal and converted to plain text format using the textreadr package in R. Standard data cleaning practices were employed with the TM (text mining) package to remove non-data and low value tokens. From this, 3,240 of the most common single, as well as 288 multi-word terms were identified. Additionally, the research weaknesses of this approach are reported. To the best of our knowledge, this is the first report of the use of data text mining to identify core vocabulary.

## Appendices

Table S1. Stemmed unigrams (3240) from Volumes 1-8 of the open source journal, Foods. Superscripted terms are defined in table footnotes.

Term	Freq.	Term	Freq.	Term	Freq.	Term	Freq.	Term	Freq.
sampl	27321	reduc	5998	potenti	3600	depend	2816	assess	2375
food	23560	fruit	5891	form	3549	provid	2810	pressur	2358
acid	21133	condit	5821	raw	3525	mix	2802	honey	2357
product	18738	perform	5794	fresh	3517	volatil	2792	initi	2355
studi	18351	produc	5726	analyz	3492	presenc	2780	requir	2351
figur	17941	due	5606	strain	3484	industri	2759	chang	2349
content	17092	decreas	5605	affect	3470	age	2738	packag	2342
signific	15591	starch	5566	lipid	3462	bacteria	2724	interact	2335
extract	15472	properti	5506	treat	3447	limit	2710	white	2328
result	15421	evalu	5441	sourc	3433	main	2704	made	2327
effect	14430	qualiti	5404	experi	3422	separ	2698	set	2323
protein	14145	relat	5395	review	3421	consid	2685	ingredi	2268
increas	13925	similar	5390	sugar	3409	work	2671	dietari	2262
concentr	13385	low	5329	repres	3395	term	2652	period	2259
activ	12648	solut	5302	nutrit	3388	reaction	2651	intak	2256
tabl	12196	lower	5236	human	3386	combin	2650	induc	2229
time	11595	type	5229	contribut	3371	deriv	2645	ethanol	2225
water	11379	prepar	5088	variabl	3339	cultur	2644	volum	2217
analysi	11103	includ	4976	inhibit	3321	format	2626	commerci	2197
compound	11093	rang	4950	growth	3314	mean	2612	yeast	2197
high	10972	composit	4930	identifi	3267	capac	2612	predict	2190
min	10943	detect	4886	assay	3265	isol	2608	comparison	2181
total	10784	bread	4866	fraction	3247	amino	2607	letter	2179
method	10578	develop	4845	specif	3235	reduct	2604	grain	2172
show	10456	shown	4721	accept	3233	point	2603	peptid	2170
process	10431	import	4718	highest	3222	year	2600	green	2164
valu	10020	color	4706	investig	3219	textur	2596	solid	2162
temperatur	9657	rate	4700	organ	3210	carri	2589	flavonoid	2149
level	9396	meat	4667	select	3204	oliv	2580	major	2147
differ	9051	research	4631	order	3158	area	2563	fish	2141
higher	9025	fat	4564	speci	3153	general	2553	absorb	2131
oil	8897	system	4547	calcul	3150	suggest	2551	design	2128
cell	8887	sensori	4542	natur	3139	averag	2545	flavor	2126
report	8699	amount	4519	surfac	3131	essenti	2545	particip	2124
milk	8562	weight	4466	juic	3108	score	2526	radic	2123
determin	8550	materi	4434	attribut	3105	incub	2524	log	2122
treatment	8451	chees	4368	size	3088	mixtur	2521	indic	2103
day	8442	oxid	4328	characterist	3065	yield	2509	carotenoid	2103
control	8232	number	4268	consumpt	3059	analys	2506	sodium	2096
dri	8189	factor	4257	final	3047	mass	2503	acet	2087
obtain	7967	previous	4224	polyphenol	3043	solubl	2502	individu	2085
test	7891	heat	4186	influenc	3032	consist	2502	film	2077
data	7858	compon	4177	fatti	3030	glucos	2494	solvent	2075
compar	7402	paramet	4161	rice	3028	store	2483	line	2044
addit	7387	plant	4021	profil	2989	tast	2477	dough	2039
measur	7344	wine	3998	improv	2984	negat	2474	resist	2036
antioxid	7241	wheat	3975	phase	2982	liquid	2466	conduct	2035
present	7060	cook	3939	diet	2979	experiment	2459	life	2035
sampl	27321	express	3911	loss	2978	collect	2457	vitamin	2031
group	6839	chemic	3903	case	2966	inform	2452	anim	2031
observ	6747	statist	3883	ratio	2964	cultivar	2448	muscl	2026
model	6713	function	3828	red	2948	salt	2435	diseas	2021
base	6702	free	3811	appli	2928	energi	2434	medium	2019
consum	6681	respons	3728	technolog	2906	stabil	2433	impact	2018
flour	6636	digest	3673	powder	2904	moistur	2423	region	2017
standard	6539	structur	3658	univers	2882	veget	2411	scienc	2015
found	6537	posit	3654	origin	2861	column	2397	part	1992
storag	6510	seed	3626	varieti	2848	applic	2394	sweet	1992
ferment	6403	correl	3624	peak	2819	common	2383	leav	1985
phenol	5999	health	3609	enzym	2818	intens	2376	character	1972

Table S1. Stemmed unigrams (3240) from Volumes 1-8 of the open source journal, Foods (continued). Superscripted terms are defined in table footnotes.

Term	Freq.	Term	Freq.	Term	Freq.	Term	Freq.	Term	Freq.
state	1969	anthocyanin	1691	market	1482	trial	1306	abund	1154
gluten	1963	count	1684	whey	1478	manufactur	1306	brown	1152
bioactiv	1953	unit	1681	valid	1477	garlic	1304	long	1152
curv	1948	index	1680	beverag	1476	introduct	1303	record	1151
techniqu	1939	discuss	1678	popul	1475	constant	1302	peel	1150
supplement	1939	potato	1678	end	1469	microorgan	1294	replic	1148
cfu <sup>AA</sup>	1938	find	1677	explain	1469	meal	1294	wash	1147
receiv	1933	vitro	1673	strong	1467	salmonella	1294	intestin	1143
light	1932	residu	1666	support	1460	filter	1292	like	1141
bean	1932	chain	1662	caus	1457	freez	1291	lab	1138
demonstr	1921	larg	1661	power	1456	bar	1286	nitrogen	1137
deviat	1921	miner	1659	lowest	1452	examin	1277	skin	1136
percentag	1920	membran	1656	approxim	1452	physic	1273	pigment	1135
hplc	1919	abil	1647	casein	1443	normal	1273	imag	1134
lead	1900	coeffici	1644	sensit	1441	pre	1272	batch	1131
peer	1889	maximum	1644	complet	1439	literatur	1270	gelatin	1130
spectra	1878	countri	1643	chromatographi	1430	fact	1269	tomato	1128
good	1872	prevent	1640	exhibit	1425	tpc	1265	transfer	1127
refer	1867	thermal	1637	ripen	1425	degre	1260	promot	1125
purchas	1862	monocytoxygen	1622	conclus	1423	sucros	1258	biofilm	1125
scale	1860	beef	1621	yellow	1420	regul	1256	frequenc	1123
antimicrobi	1857	metabol	1618	greater	1420	load	1255	singl	1120
optim	1855	step	1614	eat	1417	expect	1254	corn	1119
direct	1853	coli	1614	hand	1416	note	1250	agent	1115
publish	1848	methanol	1612	least	1412	room	1247	densiti	1115
enhanc	1844	depart	1611	black	1411	anova	1245	damag	1114
recent	1844	methyl	1610	modifi	1407	reveal	1243	month	1112
small	1833	coat	1606	approach	1400	calcium	1240	air	1112
matrix	1831	cereal	1604	bake	1395	abstract	1234	ester	1108
anti	1829	aroma	1603	retent	1395	mpa	1233	subsequ	1108
tissu	1818	degrad	1581	stress	1394	nutrient	1229	cluster	1102
molecular	1810	role	1578	achiev	1393	laboratori	1226	regress	1100
fiber	1803	agricultur	1577	tradit	1393	multipl	1220	inflammatori	1099
equat	1801	feed	1566	centrifug	1391	blood	1217	independ	1094
contamin	1796	viscos	1566	remain	1390	pca <sup>AC</sup>	1215	scientif	1093
particl	1784	protect	1563	formul	1382	keyword	1212	homogen	1089
interest	1772	dairi	1562	biolog	1380	section	1211	oxygen	1088
mechan	1770	involv	1557	emuls	1379	environment	1210	target	1087
microbi	1765	gas	1556	error	1374	inocul	1209	preserv	1085
pasta	1765	carbohydr	1554	behavior	1372	identif	1208	institut	1084
safeti	1764	appl	1553	wast	1371	insect	1205	ethyl	1082
alcohol	1763	confirm	1552	grape	1370	relationship	1202	field	1080
distribut	1760	estim	1549	articl	1368	bodi	1202	fit	1080
plate	1755	soybean	1537	molecul	1352	percept	1196	substrat	1080
risk	1751	intern	1533	ion	1350	gene	1195	trend	1076
harvest	1744	rich	1531	varianc	1347	equival	1179	led	1075
dphph <sup>AB</sup>	1744	appear	1526	effici	1344	inhibitori	1176	carbon	1074
generat	1735	dilut	1524	slight	1343	pure	1175	proport	1072
linear	1730	remov	1520	dose	1335	colour	1174	kda	1070
stage	1728	plot	1518	softwar	1334	convent	1174	maintain	1068
make	1727	contain	1517	releas	1333	pattern	1174	princip	1062
procedur	1727	bacteri	1516	hydrolys	1326	distil	1167	mice	1062
fund	1725	post	1512	respect	1325	enrich	1166	current	1061
occur	1718	week	1512	absorpt	1324	discrimin	1164	caroten	1060
subject	1714	vari	1506	scaveng	1322	cut	1160	lactic	1057
yogurt	1703	pathogen	1505	hydrolysi	1321	hard	1159	environ	1056
analyt	1701	flow	1498	buffer	1321	triplic	1159	defin	1054
complex	1699	reach	1492	shelf	1315	incorpor	1158	benefit	1054
gel	1697	wide	1491	calibr	1310	band	1158	allow	1053
variat	1696	egg	1488	aim	1306	label	1158	cholesterol	1052

<sup>AA</sup> colony forming unit, <sup>AB</sup> (2,2-diphenyl-1-picryl-hydrazyl-hydrate) free radical method, <sup>AC</sup> principle components analysis

Table S1. Stemmed unigrams (3240) from Volumes 1-8 of the open source journal, Foods (continued). Superscripted terms are defined in table footnotes.

Term	Freq.	Term	Freq.	Term	Freq.	Term	Freq.	Term	Freq.
paper	1048	metabolit	943	matter	863	tool	779	sds <sup>AL</sup>	719
differenti	1047	refriger	942	panel	860	strawberri	777	diabet	717
bind	1043	blue	938	amin	860	evoo <sup>AJ</sup>	777	cycl	716
dna	1043	contrast	935	steril	859	ash	776	pls <sup>AM</sup>	716
altern	1041	sigma	932	bitter	859	tender	776	seafood	715
fluoresc	1041	account	932	diamet	858	quantit	773	spectrum	713
recommend	1039	squar	932	roast	856	surviv	770	lycopen	712
establish	1038	oper	928	kinet	856	fri	768	geograph	712
pork	1038	exposur	927	gum	856	extens	766	electron	711
liver	1035	nation	925	mode	854	fed	766	ensur	709
grow	1034	cow	925	soy	851	element	765	strategi	709
mol	1033	quercetin	922	crystal	848	integr	765	aldrich	707
cross	1028	chines	913	iron	842	date	764	spectroscopi	705
recoveri	1023	fragment	911	gae <sup>AH</sup>	842	avoid	761	spectrometri	704
chicken	1023	mushroom	911	spray	839	aggreg	760	challeng	702
typic	1020	perceiv	909	bond	837	desir	759	highlight	702
edibl	1018	thick	908	object	834	purpos	758	classif	700
modif	1017	farm	907	rpm <sup>AI</sup>	834	microbiolog	757	adult	698
sequenc	1017	hexan	902	mention	827	key	756	manuscript	696
blend	1013	european	901	access	824	frozen	756	detector	696
cancer	1010	puls	901	quantif	823	daili	754	display	696
continu	1003	serv	898	ground	820	gram	753	marker	696
prior	999	sensor	898	soil	820	lack	752	inject	695
citrus	999	employ	897	cost	819	fold	752	ultrasound	695
abt <sup>AD</sup>	998	manag	897	understand	818	lactobacillus	751	coagul	695
probiot	993	equip	897	suitabl	818	ascorb	749	relev	692
practic	992	replac	893	associ	816	root	749	medicin	691
propos	991	media	891	biscuit	816	dissolv	748	scan	688
project	989	quantifi	891	tube	813	secondari	745	amylos	686
spp <sup>AE</sup>	989	short	891	carrot	812	catechin	745	concern	684
grown	988	dispers	889	suppli	811	boil	745	crude	684
pef <sup>AF</sup>	988	partial	888	orang	810	correspond	744	grade	683
write	986	peopl	887	firm	807	peroxid	744	pomac	683
monitor	982	season	887	oat	805	beer	744	granul	682
layer	981	matur	886	cultiv	804	reflect	741	emot	682
dark	980	polysaccharid	886	success	801	chickpea	740	detail	679
reason	980	pea	886	bran	801	spectral	739	zone	679
methodolog	979	quantiti	883	accumul	800	physicochem	738	transport	677
cool	974	tran	883	start	799	center	737	per <sup>AN</sup>	677
focus	974	crop	881	odor	799	tannin	737	leaf	676
class	973	adjust	880	prefer	798	agreement	736	spoilag	675
link	967	plasma	880	nativ	798	extern	736	screen	674
forc	967	fibr	880	phosphat	797	code	735	minor	674
train	966	signal	879	featur	794	pectin	735	oven	673
past	965	consider	877	polar	793	vivo <sup>AK</sup>	734	tukey	672
basi	964	healthi	876	stabl	791	equal	732	global	672
gallic	964	agar	876	vacuum	791	classifi	731	pulp	672
local	963	supernat	875	action	789	serum	731	skim	671
locat	963	play	875	categori	789	descript	730	peanut	671
substanc	963	great	874	electr	788	chromatograph	728	item	670
instrument	957	coffe	874	random	787	formula	728	list	670
aqueous	955	encapsul	873	shape	785	acknowledg	726	primari	667
visual	953	tea	869	ice	785	pufa	726	genet	666
inactiv	952	maiz	869	util	785	issu	724	correct	665
toxic	951	NaCl <sup>AG</sup>	869	sustain	784	phytochem	723	carcass	665
cold	950	mill	867	clean	784	hot	723	need	664
close	949	knowledg	866	benefici	781	cocoa	722	alter	663
reagent	949	slice	865	world	780	sourdough	720	climat	662
enzymat	948	program	864	metal	780	exist	719	satur	662
rapid	944	evid	864	legum	780	hold	719	choic	662

<sup>AD</sup> (2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid)) which is used in a radical scavenging assay, <sup>AE</sup> an abbreviation for species (plural); sp is an abbreviation of the singular form, <sup>AF</sup> pulsed electric field, <sup>AG</sup> sodium chloride (common table salt), <sup>AH</sup> gallic acid equivalents, <sup>AI</sup> revolutions per minute, <sup>AJ</sup> extra virgin olive oil, <sup>AK</sup> in-vivo, i.e., in a living system, <sup>AL</sup> sodium dodecyl sulfate, <sup>AM</sup> partial least squares, <sup>AN</sup> polymerase chain reaction

Table S1. Stemmed unigrams (3240) from Volumes 1-8 of the open source journal, Foods (continued). Superscripted terms are defined in table footnotes.

Term	Freq.	Term	Freq.	Term	Freq.	Term	Freq.	Term	Freq.
minimum	662	MAP <sup>AQ</sup>	614	gastrointestin	571	edit	529	session	501
frap <sup>AO</sup>	661	extrud	614	chromatogram	571	revers	528	infrar	500
pasteur	661	ppm	613	buckwheat	570	question	527	take	500
minim	660	quinoa	613	expos	569	matric	527	precipit	500
possibl	660	drink	612	stimul	569	possess	527	glycat	500
side	660	versus	609	econom	568	flesh	526	cellular	499
clear	658	top	608	hydrophob	568	kind	526	infect	499
speed	658	inhibitor	608	protocol	565	fisher	526	ioniz	498
tocopherol	658	genotyp	607	assist	564	purifi	526	synthesi	498
male	657	fructos	606	south	563	wall	525	real	497
cream	656	administr	603	aspect	563	stir	524	smoke	497
fillet	653	antibacteri	603	rennet	562	page	524	optimum	496
untreat	653	viabil	602	plantarum	562	smaller	524	consequ	496
bag	650	reactiv	600	futur	560	biogen	522	belong	496
rheolog	650	domin	600	filtrat	560	batter	522	threshold	494
aldehyd	648	potassium	600	fast	559	convert	521	distinct	493
shear	648	hhp <sup>AR</sup>	600	algin	558	distanc	520	insolubl	493
femal	647	chitosan	599	absenc	557	conclud	519	cerevisia	493
lentil	646	dish	598	ace <sup>AT</sup>	557	superscript	519	illustr	492
portion	645	extrem	596	suspens	556	immobil	519	devic	492
hydrogen	645	rat	596	biomass	556	prebiot	518	intent	492
full	644	histamin	594	educ	554	mold	518	mark	491
neutral	643	contact	594	excess	551	pro	517	divid	491
glucosid	643	critic	593	supplementari	551	refin	517	inoculum	491
describ	642	cis <sup>AS</sup>	593	longer	550	modul	516	weigh	490
problem	641	safe	592	steam	550	frequent	515	glycerol	489
microwav	641	aromat	592	site	550	chosen	515	balanc	489
strength	640	read	592	demand	549	call	514	supervis	489
substitut	640	central	592	isom	549	innov	514	gain	488
adulter	640	colon	590	act	548	descriptor	513	aceton	488
left	639	yoghurt	590	starter	546	conjug	513	suppress	487
brine	637	length	589	plastic	545	grass	513	capabl	487
repeat	636	barley	588	aureus	543	raman	513	immun	486
adhes	635	physiolog	588	creat	541	evapor	512	extend	486
doubl	635	regard	588	larger	541	briefli	511	mango	485
advantag	634	oxal	587	soft	541	emulsifi	510	ginger	484
flavour	633	basic	586	constitu	540	hour	510	gradient	483
nir <sup>AP</sup>	633	pathway	586	denatur	540	immedi	509	half	482
version	632	chocol	586	glass	539	infant	509	excel	481
wavelength	632	extent	585	stain	539	esi <sup>AW</sup>	508	numer	481
moder	632	cover	584	draft	539	antibiot	508	piec	481
chlorid	631	elut	583	nanoparticl	539	trolox	507	famili	481
network	629	wild	583	pack	538	durat	507	attent	481
mmol	629	collagen	583	HCl <sup>AU</sup>	538	rear	507	interpret	481
bovin	629	place	581	berri	538	transit	506	difficult	480
compani	628	agil	581	diffus	537	row	506	space	480
crumb	627	public	580	rye	537	mic <sup>AX</sup>	506	trace	479
bound	627	resourc	580	obes	537	engin	505	coloni	479
amylas	626	coupl	580	recip	536	gut	505	respond	478
flower	626	microbiota	580	box	536	bath	505	dehydr	478
crossref	626	fortifi	577	SEM <sup>AV</sup>	536	adapt	505	oral	478
mobil	624	turkey	576	run	534	retail	504	ultra	478
poultri	624	divers	576	droplet	534	reliabl	503	ultrason	477
broth	622	spore	576	interv	534	cardiovascular	503	intervent	477
children	619	poor	575	tag	534	vis <sup>AY</sup>	503	smooth	476
patient	619	transform	575	muffin	534	semi	503	gastric	476
open	616	lutein	574	thermo	532	instanc	502	breed	475
clinic	616	survey	574	linol	531	synthet	502	earli	475
pastur	616	simul	573	cake	531	butter	502	convers	473
panelist	616	glycosid	573	foodborn	530	simpl	501	acetonitril	473

<sup>AO</sup> ferric reducing antioxidant power, <sup>AP</sup> near infrared, <sup>AQ</sup> modified atmosphere packaging, <sup>AR</sup> high hydrostatic pressure, <sup>AS</sup> a chemical term meaning adjacent (opposed to trans), <sup>AT</sup> could be several things such as the chemical prefix ace- (acetone, acetate, acetic), <sup>AU</sup> chemical formula for hydrochloric acid, <sup>AV</sup> scanning electron microscopy, <sup>AW</sup> electrospray ionization, <sup>AX</sup> minimum inhibitory concentration, <sup>AY</sup> visible, usually in UV-VIS

Table S1. Stemmed unigrams (3240) from Volumes 1-8 of the open source journal, Foods (continued). Superscripted terms are defined in table footnotes.

Term	Freq.								
adequ	472	compress	444	phytosterol	415	insid	390	spss <sup>BP</sup>	366
suscept	472	compris	444	press	414	colleg	390	pass	366
seri	471	polym	444	ether	414	permeat	390	drop	366
onion	470	view	444	qualit	414	mda <sup>BM</sup>	389	formal	366
recov	469	address	443	duplic	414	max	388	avail	365
spectrophotomet	469	bubbl	443	outbreak	414	spanish	388	contrari	365
special	468	larva	443	osmot	413	rank	387	exceed	364
chamber	468	explor	442	morpholog	411	charg	386	overnight	364
sens	467	share	442	substanti	411	pretreat	386	million	364
status	466	elimin	442	predomin	410	fix	385	hypertens	364
snack	466	drug	442	aliquot	410	suffici	385	herb	363
micro	465	irradi	442	appar	409	patti	385	brain	363
sour	465	sausag	441	inflamm	409	stem	385	declin	363
onlin	465	fluid	439	tbar <sup>BI</sup>	408	Maillard	385	foam	363
broccoli	465	proteas	439	march	408	insulin	385	citric	363
soak	464	bakeri	438	sector	408	break	384	algorithm	363
cooki	464	wet	438	approv	407	chronic	384	swell	362
lactos	463	toler	438	wpi <sup>BJ</sup>	407	mycotoxin	384	worldwid	361
glucosidas	463	viabl	437	progress	406	cherri	383	assum	361
kernel	462	artifici	437	folin	406	breadfruit	383	glucan	360
probe	461	optic	436	millet	406	flatbread	383	toxin	360
implement	461	ingest	436	tree	404	cytotox	382	hull	360
sum	461	gmp <sup>BE</sup>	436	vinegar	404	recogn	382	gave	359
lod <sup>AZ</sup>	461	hydroxyl	435	odour	403	trypsin	382	distinguish	358
titrat	461	lemon	435	marin	403	hydrat	380	iso	358
visibl	460	listeria	433	blueberri	402	construct	380	rest	358
young	460	atcc <sup>BF</sup>	433	block	401	italian	380	north	358
dynam	460	conceptu	433	bacillus	400	decis	379	multi	356
accuraci	459	pesticid	432	assign	400	spread	379	carrier	356
live	458	retain	431	elast	399	popular	377	fructan	356
western	458	unknown	430	bead	399	agre	377	minc	355
cohes	458	caffeic	430	percent	399	readi	377	proteolysi	355
salmon	458	loq <sup>BG</sup>	430	broad	398	ftir <sup>BN</sup>	377	chip	355
ident	457	offer	429	crust	398	sunflow	377	angl	354
database	454	regular	428	check	398	wood	377	triglycerid	354
electrod	454	turn	428	modulus	397	proxim	377	except	353
school	453	blank	427	typhimurium	397	terpen	376	precursor	353
begin	453	ferul	427	amaranth	397	irrig	375	seawe	353
precis	453	flaxse	426	grant	396	defat	375	seal	352
nmr <sup>BA</sup>	453	easili	425	introduc	396	radiat	375	biochem	352
noc <sup>BB</sup>	453	promis	425	cla <sup>BK</sup>	396	student	375	syring	352
bone	452	compos	424	durum	394	context	375	connect	352
attitud	451	accur	421	emerg	394	extra	374	technic	351
micell	451	germin	421	antibodi	394	purif	374	cellulos	351
accord	450	pah <sup>BH</sup>	421	valuabl	393	confid	374	curd	351
faculti	450	advanc	420	concept	393	spike	373	polymer	351
give	450	australian	420	facilit	393	cont	373	tap	351
ros <sup>BC</sup>	450	kimchi	420	authent	393	blanch	373	claim	350
simultan	450	spice	417	handl	393	agit	372	shift	350
breast	450	bio	417	xanthan	393	dad <sup>BO</sup>	372	behaviour	350
pepper	450	hydrolyz	417	lps <sup>BL</sup>	393	magnet	372	intermedi	350
trait	449	allergen	417	stretch	392	microscop	372	gender	350
campylobact	449	sprout	417	induct	392	headspac	371	ham	349
humid	448	option	417	person	391	fill	370	medic	348
vulgari	448	efficaci	416	barrier	391	care	370	aerob	348
epicatechin	448	flask	416	array	391	notic	369	price	348
atmospher	447	oleic	416	bioavail	391	caffein	368	slaughter	348
preval	447	brand	415	occurr	391	zinc	368	brew	348
NaOH <sup>BD</sup>	445	goat	415	ara	391	receptor	367	height	347
multivari	445	kit	415	acquisit	390	india	367	inclus	347

<sup>AZ</sup> limit of detection, <sup>BA</sup> nuclear magnetic resonance, <sup>BB</sup> normal operating conditions, <sup>BC</sup> reactive oxygen species, <sup>BD</sup> chemical formula for sodium hydroxide, <sup>BE</sup> good manufacturing practices, <sup>BF</sup> American Type Culture Collection (an organization that sells microbial cultures), <sup>BG</sup> limit of quantification, <sup>BH</sup> polyaromatic hydrocarbon, <sup>BI</sup> thiobarbituric acid reactive substance, <sup>BJ</sup> whey protein isolate, <sup>BK</sup> not an abbreviation, many different words begin this way, <sup>BL</sup> liposaccharide, <sup>BM</sup> malondialdehyde, <sup>BN</sup> Fourier transform infrared spectroscopy, <sup>BO</sup> diode array detector, <sup>BP</sup> Statistical Package for Social Sciences, an IBM statistics program that is widely used.

Table S1. Stemmed unigrams (3240) from Volumes 1-8 of the open source journal, Foods (continued). Superscripted terms are defined in table footnotes.

Term	Freq.	Term	Freq.	Term	Freq.	Term	Freq.	Term	Freq.
scent	347	chlorogen	332	capillari	317	match	302	mild	287
summar	346	formic	332	resolut	317	puriti	302	name	286
gmbh	345	microp	332	upper	317	narrow	301	largest	286
favor	345	rise	331	aglycon	317	sauc	301	folat	286
incid	345	proper	330	rutin	317	chemometr	301	exchang	285
brix	345	spectromet	330	minut	317	match	302	principl	285
bsa <sup>BQ</sup>	345	aoac <sup>BS</sup>	330	help	316	enter	300	vibrat	285
chemistri	344	background	330	hygien	316	resveratrol	300	sediment	285
uniform	344	deposit	330	pumpkin	316	put	299	disulfid	285
death	344	palat	330	lettuc	316	phenomenon	299	depict	285
biotechnolog	344	sterol	330	deform	316	summer	299	defici	285
acceler	344	react	329	pearson	316	govern	298	appetit	285
ask	344	delay	329	caco	315	linalool	298	yolk	283
mediat	344	escherichia	329	maltodextrin	315	needl	298	kcal	283
middl	344	comput	329	dha <sup>BV</sup>	315	voltag	297	output	283
extrus	344	mouth	328	competit	315	attract	297	macrophag	282
women	344	author	328	fingerprint	315	ferric	297	slow	282
baselin	344	cereus	327	juici	315	melt	297	kaempferol	282
microstructur	343	fungi	327	elev	314	poland	297	colorimetr	281
exert	343	oligosaccharid	327	astring	314	brazil	297	limonen	280
prolifer	343	spinach	327	interfer	314	american	296	coordin	280
entir	343	semolina	327	dataset	314	bottom	296	instruct	280
pepsin	343	Merck	326	easi	312	salti	296	plum	280
acquir	342	slope	326	advers	312	alkalin	296	acrylamid	280
uniqu	342	ciocalteu	326	home	312	flavonol	296	noodl	280
actual	341	bulk	326	thin	312	intact	296	mci <sup>BZ</sup>	280
stock	341	carbonyl	326	hydrocarbon	312	sfa <sup>BX</sup>	296	plain	279
expert	341	meet	325	oak	312	genistein	296	kidney	279
procyanidin	341	rose	325	dimens	311	hpp	296	glutam	279
prove	340	fine	325	diphenyl	311	permeabl	295	earlier	278
exclud	340	ring	325	pool	311	probabl	295	offici	278
unsatur	340	adher	325	sanit	310	zealand	294	ultrafiltr	278
emiss	340	gradual	325	manner	310	bht <sup>BY</sup>	294	definit	277
axi	340	polyunsatur	325	disord	310	tyramin	294	heavi	277
bottl	340	pomegran	325	africa	310	pump	294	tract	277
dimension	339	sensat	324	lysin	310	chewi	294	northern	277
enabl	339	pilot	324	meter	309	wrote	294	southern	277
salad	339	babi	324	nutraceut	309	winter	294	apigenin	277
eye	338	criteria	324	hydrophil	309	subtili	294	nitrit	277
weak	338	pubm	324	MUFA <sup>BW</sup>	309	situat	293	voc <sup>CA</sup>	277
fulli	338	preliminari	323	lyophil	308	palmit	293	mpod <sup>CB</sup>	277
input	338	pair	323	stronger	308	condens	293	breakfast	276
bas <sup>BR</sup>	338	hedon	323	excit	308	maxim	292	glu <sup>CC</sup>	276
restrict	338	pbs <sup>BT</sup>	322	shellfish	308	equilibrium	291	steel	275
evolut	338	ad hoc	321	ionic	307	lactat	291	waxi	275
sea	337	botan	321	meoh	306	pore	291	machin	274
centr	337	household	321	crucial	306	isotherm	291	rapese	274
questionnair	337	don	321	ill	306	core	290	immers	274
spme	337	tend	321	search	306	oxidas	290	glycem	274
denot	336	orac <sup>BU</sup>	321	silag	306	cracker	290	mrna	274
europ	336	car	321	keton	305	outcom	289	faba	274
sever	336	fungal	320	porcin	305	guidelin	289	honeydew	273
inulin	336	adsorpt	320	lane	305	trout	289	linolen	273
gallat	336	thyme	320	sound	304	attach	289	cure	272
luteolin	335	virgin	319	rais	304	salin	289	plan	272
constitut	334	coumar	319	hazard	304	amylopectin	289	rough	272
microscopi	333	scatter	319	finish	304	digit	288	hay	272
sulfat	333	verifi	319	dimethyl	304	disrupt	288	adopt	271
sonic	333	vial	318	cabbag	303	tropic	288	exclus	271
purpl	333	held	318	teac	303	apoptosi	288	maltos	271

<sup>BQ</sup> bovine serum albumin, <sup>BR</sup> biogenic amines, <sup>BS</sup> Association of Official Analytical Chemists, <sup>BT</sup> phosphate buffered saline, <sup>BU</sup> reduced antioxidant compounds, <sup>BV</sup> docosahexaenoic acid, <sup>BW</sup> monounsaturated fatty acids, <sup>BX</sup> saturated fatty acid, <sup>BY</sup> butylated hydroxytoluene, <sup>BZ</sup> multivariate curve resolution, <sup>CA</sup> volatile organic compounds, <sup>CB</sup> macular pigment optical density, <sup>CC</sup> several interpretations, glucose, glutamic acid

Table S1. Stemmed unigrams (3240) from Volumes 1-8 of the open source journal, Foods (continued). Superscripted terms are defined in table footnotes.

Term	Freq.	Term	Freq.	Term	Freq.	Term	Freq.	Term	Freq.
injuri	271	darmstadt	257	subsp	246	scheme	235	goal	224
implic	270	intracellular	257	facil	246	enterica	235	carboxyl	224
sweeten	270	member	257	restaur	246	graph	234	pan	224
breakdown	270	uht <sup>CL</sup>	257	albumin	246	screw	234	tail	224
revis	269	summar	257	malt	246	ambient	234	chelat	224
tendenc	269	mealworm	257	fortif	245	ole <sup>DA</sup>	234	sod <sup>DI</sup>	224
casei	269	propoli	257	trp <sup>CU</sup>	245	disinfect	234	warm	223
pepton	269	fir	256	umami	245	leg	234	thaw	223
fos <sup>CD</sup>	269	impli	256	effort	244	fractur	233	discard	222
hue	268	document	256	extracellular	244	clot	233	domest	222
communic	268	soup	256	symptom	244	provinc	232	vortex	222
shimadzu	268	pharmaceut	256	shell	244	expans	232	superior	222
iodin	268	proteolyt	256	channel	244	byproduct	232	raspberri	222
mediterranean	267	back	255	cyanidin	244	pdo <sup>DB</sup>	232	nps <sup>DJ</sup>	222
quick	267	incom	255	remark	243	pda <sup>DC</sup>	232	sheet	221
greatest	267	spot	255	porridg	243	chill	231	feel	221
electrophoresi	267	zeaxanthin	254	window	242	rosemari	231	viscous	221
organolept	267	usda	254	comprehens	242	bioaccess	231	academ	221
beet	267	lamb	254	gap	241	well	231	deterior	220
social	266	tuber	254	sulfur	241	stainless	231	mesophil	220
duncan	266	ATP <sup>CM</sup>	253	chew	240	faster	231	parallel	220
stomach	266	portug	253	stand	240	quadrat	231	virus	220
rte <sup>CE</sup>	266	alpha	253	pronounc	240	citrat	231	lost	219
spring	266	postharvest	253	cognit	240	secret	231	shorter	219
sulfit	266	tnf <sup>CN</sup>	253	chlorophyl	240	indirect	230	isoflavon	219
lsd <sup>CF</sup>	265	paw <sup>CO</sup>	253	round	239	macer	230	universidad	219
cap	265	hpmc <sup>CP</sup>	253	uptak	239	undergo	230	melon	219
notabl	264	curat	253	vapor	239	pollen	230	familiar	219
heart	264	crb <sup>CQ</sup>	253	biosensor	239	stream	230	mgo <sup>DK</sup>	219
eucalyptus	264	fall	252	biomark	239	traceabl	230	eugenol	219
absolut	264	siev	252	aerat	239	stone	229	staphylococcus	218
ppo <sup>CG</sup>	264	calori	252	sdc <sup>CV</sup>	239	outer	229	union	218
glucosinol	264	FAME <sup>CR</sup>	252	rins	238	phospholipid	229	adsorb	218
manual	263	jam	252	opposit	238	ige <sup>DD</sup>	229	serotyp	218
style	263	pleasant	252	MRS <sup>CW</sup>	238	butyr	229	wind	218
tumor	263	alga	251	hous	238	hybrid	229	silica	217
genus	263	pellet	251	flux	238	reproduc	229	regulatori	217
primarili	262	sleep	251	turmer	238	pseudomon	229	factori	217
tyr <sup>CH</sup>	262	microgreen	251	farmer	237	commiss	228	pet <sup>DL</sup>	217
dislik	262	habit	250	phe <sup>CX</sup>	237	interfac	228	word	217
stec <sup>CI</sup>	262	corpor	250	predictor	237	dioxid	228	CaCl	217
nut	261	synthes	250	melanoidin	237	pear	228	autoclav	217
tuna	261	flavon	250	sticki	236	crystallin	228	cytokin	216
pinen	261	increment	250	split	236	proso	228	width	216
lipas	260	uplc <sup>CS</sup>	250	exampl	236	awar	227	syrup	216
suspend	260	trade	249	mathemat	236	surround	227	chart	216
nuclear	260	sheep	249	antigen	236	beta	227	delbrueckii	216
hexosid	260	primer	249	undesir	236	magnesium	226	kale	216
conform	259	decad	248	host	236	hop	226	financi	215
var	259	genom	248	celiac	236	ministri	225	cation	215
cattl	259	tongu	248	spe <sup>CY</sup>	236	putrescin	225	shop	215
banana	259	DSC <sup>CT</sup>	248	palm	236	tfc <sup>DE</sup>	225	mous	215
communiti	258	satieti	248	hazelnut	236	prolong	225	conceiv	215
chloroform	258	servic	247	tca <sup>CZ</sup>	236	reconstitut	225	teff <sup>DM</sup>	215
penetr	258	atom	247	event	235	copper	225	declar	214
buy	258	hydroxi	247	foodstuff	235	epa <sup>DF</sup>	225	equilibr	214
dss <sup>CI</sup>	258	deioniz	247	lacti	235	crab	225	shake	214
hmf <sup>CK</sup>	258	ammonium	247	conveni	235	brassica	225	termin	214
microalga	258	shrimp	246	micronutri	235	wlp <sup>DG</sup>	225	annual	214
supermarket	257	enumer	246	cystein	235	fas <sup>DH</sup>	224	postprandi	214

<sup>CD</sup> fructooligosaccharides, <sup>CE</sup> ready-to-eat, <sup>CF</sup> least square difference, <sup>CG</sup> polyphenyloxidase, <sup>CH</sup> tyrosine, <sup>CI</sup> Shigella Toxin *E. coli*, <sup>CI</sup> dextran sodium sulfate, <sup>CK</sup> hydromethylfurfural, <sup>CL</sup> ultra-high temperature, <sup>CM</sup> adenosine triphosphate, <sup>CN</sup> tumor necrosis factor, <sup>CO</sup> plasma activated water, <sup>CP</sup> hydroxypropyl methylcellulose, <sup>CQ</sup> carbohydrate rich bixin, <sup>CX</sup> fatty acid methylester, <sup>CS</sup> ultra-performance liquid chromatography, <sup>CT</sup> differential scanning calorimeter, <sup>CU</sup> tryptophan, <sup>CY</sup> Shuidouchi (fermented soybeans), <sup>CW</sup> de Man, Rogosa and Sharpe (a recipe of a culture medium), <sup>CX</sup> can be phenol or phenylalanine, <sup>CY</sup> solid phase extraction, <sup>CZ</sup> trichloroacetic acid, <sup>DA</sup> oleuropein compounds, <sup>DB</sup> protected designations of origin, <sup>DC</sup> potato dextrose agar, <sup>DD</sup> immunoglobulin E, <sup>DE</sup> total flavonoid content, or, total fecal coliforms, DF US Environmental Protection Agency, <sup>DG</sup> while-life period (refers animal agriculture), <sup>DH</sup> fatty acids, <sup>DI</sup> superoxide dismutase, <sup>DJ</sup> nanoparticles, <sup>DK</sup> methylglyoxal, DL polyethylene terephthalate, <sup>DM</sup> a type of grain from Africa that has no gluten

Table S1. Stemmed unigrams (3240) from Volumes 1-8 of the open source journal, Foods (continued). Superscripted terms are defined in table footnotes.

Term	Freq.	Term	Freq.	Term	Freq.	Term	Freq.	Term	Freq.
capsul	214	FAO <sup>DT</sup>	204	face	196	spontan	188	greenhous	180
reformul	214	potent	204	smell	196	complianc	188	modal	180
goji	213	branch	204	suffer	195	variet	188	MIP <sup>EM</sup>	180
interview	213	expens	204	ultim	195	grill	188	wrap	179
antifung	213	tss <sup>DU</sup>	204	parboil	195	carrageenan	188	migrat	179
build	212	cassava	204	cinnamaldehyd	195	lignan	187	tight	179
khz	212	jar	203	adipocyt	194	bile	187	counterpart	179
physico	212	monom	203	impair	194	export	187	synergist	179
aspergillus	212	vector	203	gummi	194	hepat	187	eaten	179
adzuki	212	reson	203	robust	194	enterobacteriacea	187	linse	179
drip	212	appreci	202	rural	194	hen	187	align	179
retrograd	212	trigger	202	heifer	194	derivat	186	leaven	179
cga <sup>DN</sup>	212	diagram	202	grey	193	prolin	186	dual	179
proven	211	opportun	202	nitric	193	intend	186	mortem	179
sequenti	211	staff	202	persist	193	ldl <sup>EH</sup>	186	molitor	179
corp	211	represent	202	cavit	193	endogen	186	pour	178
exploit	211	syndrom	202	saccharomyc	193	council	185	mash	178
picyrylhydrazyl	211	whc <sup>DV</sup>	202	phage	193	exponenti	185	nose	178
monoterpen	211	aeruginosa	202	carcinogen	192	satisfact	185	monounsatur	178
chroma	210	follow	202	legisl	192	way	185	routin	177
sas <sup>DO</sup>	210	recognit	201	schemat	192	lowercas	185	milan	177
ton	210	head	201	clarifi	192	flake	185	land	177
overview	210	orient	201	abbrevi	192	leftov	185	arab	177
mhz	210	sorghum	201	likewis	192	springi	185	ala <sup>EN</sup>	177
clove	210	wbsf <sup>DW</sup>	201	lda <sup>EC</sup>	192	licoric	185	sale	177
par	210	secur	200	std <sup>ED</sup>	192	globulin	185	canola	177
inspect	209	terpenoid	200	noni	192	choos	184	relax	177
reader	209	insight	200	stevianna	192	automat	184	patron	177
hypothes	209	cup	200	captur	191	liquor	184	vanilla	177
ethylen	209	aba <sup>DX</sup>	200	nonan	191	rule	184	gliadin	176
perspect	209	butanol	199	allergi	191	foundat	184	thought	176
indian	209	ideal	199	platelet	191	laser	184	deliv	176
pizza	209	tris <sup>DY</sup>	199	EFAS <sup>EE</sup>	191	overlap	184	framework	176
electrochem	209	allerg	199	transcript	191	sex	184	grate	176
ergosterol	209	oxoid	199	gray	191	nowaday	184	soxhlet	176
galloyl	209	glutathion	199	editor	191	glut	184	fecal	176
assessor	208	lupin	199	sold	190	alcalas <sup>EI</sup>	184	flat	176
wang	208	butyl	198	grind	190	ptm <sup>EJ</sup>	184	sign	176
tms <sup>DP</sup>	208	spectrophotometr	198	expand	190	shuidouchi	184	static	176
affin	208	hormon	198	Polyethylen	190	fourier	183	deliveri	176
gell	208	blot	198	xylos	190	elabor	183	neural	176
curri	208	built	198	trap	190	mandarin	183	transmiss	175
african	207	bac <sup>DZ</sup>	198	modern	190	add	182	stimuli	175
wolfberri	207	acidif	197	ecg <sup>EF</sup>	190	inter	182	fisheri	175
overcom	206	chi <sup>EA</sup>	197	agenc	189	baker	182	stakehold	175
ripe	206	myricetin	197	aid	189	hydroxytyrosol	182	poison	175
oregano	206	met	197	pig	189	moieti	182	veal	175
phosphorus	206	failur	197	west	189	rad	182	chia	175
vessel	205	polynomi	197	nitrat	189	uncoat	182	law	174
rsm <sup>DQ</sup>	205	median	197	arginin	189	elucid	182	systemat	174
invers	205	anticanc	197	demograph	189	transduc	182	phytat	174
theoret	205	hexen	197	urea	189	rsd <sup>EK</sup>	182	move	174
cowpea	205	fruity	197	volunt	189	lot	181	viscoelast	174
guar	205	sitosterol	197	spectroscop	189	offspr	181	lethal	174
faa <sup>DR</sup>	205	cara	197	ref	189	difficulti	180	coliform	174
alkaloid	205	coq <sup>EB</sup>	197	molar	188	glycin	180	olfactori	174
flw <sup>DS</sup>	205	sesam	196	res <sup>EG</sup>	188	millipor	180	brazilian	174
obvious	204	destruct	196	categor	188	hydroxycinnam	180	spirulina	174
amid	204	respir	196	dye	188	cmc <sup>EL</sup>	180	carob	174
answer	204	netherland	196	lactobacilli	188	last	180	attempt	173

<sup>DN</sup> chlorogenic acid, <sup>DO</sup> Statistical Analysis System (a powerful statistics program), <sup>DQ</sup> root mean square, <sup>DR</sup> free amino acids, <sup>DS</sup> food lost and waste, <sup>DT</sup> UN Food and Agricultural Organization, <sup>DU</sup> total suspended solids, <sup>DV</sup> water holding capacity, <sup>DW</sup> Warner Bratzler shear force, <sup>DX</sup> aminobutyric acid, <sup>DY</sup> refers to a type of buffer, <sup>DZ</sup> biologically active compounds, <sup>EA</sup> refers to the Chi-square test in statistics, <sup>EB</sup> co-enzyme Q, <sup>EC</sup> linear discriminant analysis, <sup>ED</sup> an abbreviation for 'standard', <sup>EE</sup> European Food Safety Authority, <sup>EF</sup> epicatechin gallate also called epigallocatechin (egc), <sup>EG</sup> abbreviation for research, <sup>EH</sup> low density lipoprotein (cholesterol), <sup>EI</sup> alcalase, <sup>EJ</sup> transmembrane pressure, <sup>EK</sup> root square difference, <sup>EL</sup> carboxymethyl cellulose, <sup>EM</sup> molecularly imprinted polymer (sensor technology), <sup>EN</sup> alanine

Table S1. Stemmed unigrams (3240) from Volumes 1-8 of the open source journal, Foods (continued). Superscripted terms are defined in table footnotes.

Term	Freq.	Term	Freq.	Term	Freq.	Term	Freq.
linkag	173	rma	168	radish	162	pcs <sup>EY</sup>	156
circul	173	beetroot	168	uvb <sup>EU</sup>	162	microencapsul	156
forag	173	evidenc	167	suboptim	162	proanthocyanidin	156
galactos	173	late	167	clust	162	asc <sup>EZ</sup>	156
propag	173	worth	167	tagatos	162	saew <sup>FA</sup>	156
tryptophan	173	bis <sup>ER</sup>	167	feder	161	tof <sup>FB</sup>	155
translat	172	propanol	167	busi	161	ultraviolet	155
repetit	172	statement	167	fourth	161	guid	155
twenti	172	dash	167	aluminum	160	KCl	155
gold	172	friend	167	supercrit	160	pcs <sup>EY</sup>	156
topic	172	leafi	167	explan	160	tension	155
wound	172	thiol	167	lactoglobulin	160	ELISA	155
net	172	erythritol	167	avocado	160	resin	155
niger	172	dehul	167	winemak	160	chemotyp	155
sci	172	malatas	167	metanil	160	imped	155
esl <sup>EO</sup>	172	coval	166	campesterol	160	auc <sup>FC</sup>	155
injector	171	older	166	microcapsul	160	fbfb <sup>FD</sup>	155
veloc	171	therapeut	166	tan	159	shorten	154
phenotyp	171	learn	166	vanil	159	shaken	154
floral	171	vital	166	empti	159	tentat	154
circ1	171	mainten	166	pine	159	english	153
pakistan	171	chestnut	166	aspart	159	insert	153
tyrosin	171	furfur	166	candid	159	reject	153
compromis	171	shanghai	165	feasibl	159	freezer	153
syneresi	171	sulfon	165	dispos	159	hydroxid	153
contributor	171	pink	165	clara	159	spheric	153
bacteriocin	171	stearic	165	omega	159	accompani	153
amplitud	171	classic	165	sorption	159	methionin	153
judg	170	aforement	165	fixat	159	histori	153
spelt	170	furanon	165	swallow	159	cue	153
sharp	170	uhplc	165	sbp <sup>EU</sup>	159	broiler	153
stop	170	sun	165	bixin	159	wort	153
onset	170	chlorin	165	economi	158	rankl	153
loin	170	diod	164	idea	158	agro	152
arsenic	170	peroxidas	164	thymol	158	bite	152
kestos	170	asp <sup>ES</sup>	164	harm	158	fertil	152
fundament	169	intrins	164	lipoprotein	158	usag	152
cadaverin	169	xylanas	164	acidifi	158	steep <sup>FE</sup>	152
carvacrol	169	gelat	164	butano	158	forward	152
custom	169	acut	164	phenomena	158	esterifi	152
nano	169	lime	164	biosynthesi	158	silver	152
rotat	169	benzo	164	symbol	158	ligand	152
pancreat	169	subunit	163	opl <sup>EV</sup>	158	histidin	152
men	169	hunger	163	conserv	157	wait	152
shs <sup>EQ</sup>	169	cryptoxanthin	163	specimen	157	cod <sup>FF</sup>	152
acetyl	169	fli	163	motiv	157	agav	152
blind	169	depth	163	forest	157	complic	151
redox	169	imf <sup>ET</sup>	163	nois	157	encourag	151
grapefruit	169	hydrocolloid	163	aptam	157	benzoic	151
superoxid	169	guarante	163	gsh <sup>EW</sup>	157	golden	151
tempor	169	alberta	163	wholegrain	157	autom	151
strip	168	barcod	163	flavan	157	terpineol	151
barcelona	168	subunit	163	cse <sup>EX</sup>	157	blackberri	151
oleuropein	168	hunger	163	glia	157	dens	150
unlik	168	cryptoxanthin	163	Minolta	156	hydroxybenzo	150
island	168	readili	162	margin	156	dryer	150
america	168	decarboxylas	162	pictur	156	phenylalanin	150
shrinkag	168	kPa	162	phytic	156	uncertainti	150
thicken	168	asian	162	urban	156	disappear	150
serial	168	fd	162	indigen	156	angiotensin	150

<sup>EO</sup> extended shelf life, <sup>EQ</sup> static headspace (refers to a type of chromatography), <sup>ER</sup> refers to a chemical structure, <sup>ES</sup> aspartic acid, <sup>ET</sup> infant milk formula, <sup>EU</sup> systolic blood pressure, <sup>EV</sup> orthogonal partial least squares, <sup>EW</sup> glutathione, EX conventional solid extraction, <sup>EY</sup> principle components, <sup>EZ</sup> autochthonous starter (strain) culture, <sup>FA</sup> slightly acidic electrolyzed water, <sup>FB</sup> time of flight (deals with a form of HPLC), <sup>FC</sup> area under curve, <sup>FD</sup> fortified blended food base, <sup>FE</sup> refers to a soaking process, e.g., corn steep liquor, <sup>FF</sup> chemical oxygen demand, or a type of fish

Table S1. Stemmed unigrams (3240) from Volumes 1-8 of the open source journal, Foods (continued). Superscripted terms are defined in table footnotes.

Term	Freq.	Term	Freq.	Term	Freq.	Term	Freq.	Term	Freq.
oyster	150	east	144	officinal	140	destroy	135	trimethyl	131
clone	150	drive	144	ellag	140	realiz	135	gte <sup>GB</sup>	131
preterm	150	aris	144	rrna	140	pastri	135	billion	130
kiwifruit	149	fluctuat	144	safflow	140	glutamin	135	dramat	130
veterinari	149	mint	144	cadmium	140	bulgaricus	135	loaf	130
bias	149	octano	144	drawn	139	assumpt	135	collis	130
som	149	confer	144	flame	139	mouthfeel	135	hepatocyt	130
infus	149	exp	144	monolay	139	epitop	135	magnif	130
spent	149	anaerob	144	compact	139	moist	135	lignin	130
decomposit	149	petroleum	144	campus	139	peach	135	triacylglycerol	130
fid <sup>FG</sup>	149	unfer	144	ancient	139	pekmez	135	culinari	130
cps <sup>FH</sup>	149	deep	144	ray	139	ghrelin	135	lifestyl	130
inl	149	bari	144	pay	139	tfs <sup>FW</sup>	135	aflatoxin	130
spirit	149	legal	143	can	139	jcp <sup>FX</sup>	135	cantaloup	130
dnaase	149	metabolom	143	propion	139	creami	134	mbf <sup>GC</sup>	130
tcata <sup>FI</sup>	149	brought	143	platform	139	assur	134	sesquiterpen	129
uvic <sup>FJ</sup>	149	heptan	143	thermophilus	139	rotari	134	ver	129
biodegrad	148	dessert	143	rapa <sup>FO</sup>	139	attenu	134	memori	129
lysozym	148	solubil	143	cinnamon	139	icp <sup>FY</sup>	134	hierarch	129
regist	148	apricot	143	tartar	139	big	134	donat	129
wave	148	segment	143	bark	139	bull	134	parent	129
mid	148	crave	143	sorbitol	139	prick	134	phenyl	129
disc	148	complementari	143	wax	138	uhph <sup>FZ</sup>	134	furan	129
caryophyllen	148	bulb	143	theori	138	macular	134	broken	129
citral	148	graminearum	143	kyoto	138	prolamin	134	drain	129
polici	147	zacateca	143	magnitud	138	matsutak	134	hepg (2) <sup>GD</sup>	129
val <sup>FK</sup>	147	apparatus	142	lag	138	cricket	134	bifidobacterium	129
stationari	147	herbal	142	tsb <sup>FR</sup>	138	kluyveri	134	pharmacolog	129
compet	147	mountain	142	edta	138	conflict	133	caramel	129
rare	147	catalyz	142	exact	138	cylind	133	steer	129
clover	147	hamburg	142	almond	138	arrang	133	anchot	129
poros	146	fbs <sup>FP</sup>	142	rgsls <sup>FS</sup>	138	bring	133	mitochondri	128
stapl	146	neuron	142	dielectr	137	today	133	skelet	128
librari	146	stigmasterol	142	cosmet	137	portugues	133	blender	128
epitheli	146	bright	141	graze	137	neglig	133	pain	128
ruptur	146	officinali	141	instant	137	refract	133	societi	128
donor	146	apoptot	141	attain	137	esterif	133	dip	128
leu	146	convect	141	characteris	137	criterion	133	cartridg	128
leucin	146	born	141	coars	137	gamma	133	minitab	128
tray	146	petri	141	bud	137	vanillin	133	BMI <sup>GF</sup>	128
dosag	146	processor	141	placebo	137	willing	133	encod	128
pancreatin	146	eventu	141	fatten	137	gls <sup>GA</sup>	133	tac	128
lean	146	electrospray	141	acyl	137	furosin	133	alfalfa	128
lri <sup>FL</sup>	146	heme	141	synbiot	137	camera	132	inin	128
propan	146	acetaldehyd	141	lac	137	cinnam	132	dhpea	128
french	146	decay	141	sakazakii	137	ferrous	132	sucralos	128
pseudoder	146	triangl	141	reli	136	capit	132	sls <sup>GG</sup>	128
ddsg <sup>FM</sup>	146	widespread	140	first	136	glycaem	132	microfluid	127
task	145	satisfactori	140	outlin	136	anion	132	track	127
certifi	145	hundr	140	aliphat	136	owe	132	believ	127
terpinen	145	slowli	140	asept	136	cas	132	benzaldehyd	127
surpris	145	leach	140	Carlo <sup>FT</sup>	136	younger	132	flavanon	127
slurri	145	epidemiolog	140	steak	136	allium	132	blade	127
park	145	wholem	140	thi	136	koji	132	tough	127
analog	145	insuffici	140	hfd <sup>FU</sup>	136	nutriti	131	movement	127
cdc <sup>FN</sup>	145	save	140	pmin	136	transpar	131	gos <sup>GH</sup>	127
universit	145	heterogen	140	cinerea	136	pose	131	stevia	127
deal	145	widespread	140	bend	135	aquacultur	131	outlier	127
graphic	145	charact	140	hexano	135	strongest	131	turbid	127
gaba <sup>FO</sup>	145	mutat	140	ANN <sup>FV</sup>	135	enteritidi	131	melatonin	127

<sup>FG</sup> flame ionization detector, <sup>FH</sup> refers to units of viscosity and a variety of maize, <sup>FI</sup> a sensory analysis method, <sup>FJ</sup> a component of UV light, <sup>FK</sup> valine, an amino acid, <sup>FL</sup> linear retention index, <sup>FM</sup> dried distillers spent grains, <sup>FN</sup> US Center for Disease Control, <sup>FO</sup> gamma aminobutyric acid, <sup>FP</sup> fetal bovine serum, <sup>FQ</sup> a species of vegetables, *Brassica rapa*, <sup>FR</sup> trypticase soy broth, <sup>FS</sup> rice glycosphingolipids, <sup>FT</sup> from Monte Carlo, <sup>FU</sup> high fat diet, <sup>FV</sup> artificial neural network, <sup>FW</sup> total flavonoids, <sup>FX</sup> JCPDS, <sup>FY</sup> inductively coupled plasma (atomic emission spectroscopy), <sup>FZ</sup> ultra high pressure homogenization, <sup>GA</sup> glucosinolates, <sup>GB</sup> green tea extract, <sup>GC</sup> milk based fruit beverages, <sup>GD</sup> a strain of human liver cells, <sup>GF</sup> body mass index, <sup>GG</sup> static light scattering, <sup>GH</sup> galacto-oligosaccharides

Table S1. Stemmed unigrams (3240) from Volumes 1-8 of the open source journal, Foods (continued). Superscripted terms are defined in table footnotes.

Term	Freq.	Term	Freq.	Term	Freq.	Term	Freq.	Term	Freq.
bergamot	127	FST <sup>GN</sup>	124	sera	120	fodmap <sup>GY</sup>	117	octen	114
gelatinis	127	biophenol	124	ethylbenzothiazolin	120	spm <sup>GZ</sup>	117	violet	114
tds <sup>GI</sup>	127	certif	123	allyl	120	mel	117	tortilla	114
mag	127	thiobarbitur	123	arilaitensi	120	diethyl	116	adolesc	114
pwp <sup>GJ</sup>	127	disadvantag	123	feijo	120	illumin	116	pyren	114
pulcherrima	127	norway	123	return	119	coil	116	curtail	114
mono	126	cleavag	123	crush	119	emmer	116	mask	114
japanes	126	absent	123	whatman	119	consecut	116	silymarin	114
maldi	126	configur	123	kefir	119	<i>lactococcus</i>	116	hioxmap	114
fece	126	burn	123	kinas	119	resid	116	elicit	113
cotyledon	126	defrost	123	closer	119	german	116	malondialdehyd	113
slower	126	lacton	123	glycol	119	polypeptid	116	spermidin	113
spi	126	steadi	123	domain	119	manur	116	supplier	113
hydrostat	126	soya	123	administ	119	consent	116	astaxanthin	113
dextran	126	mayonnais	123	antagonist	119	user	116	partit	113
rainfal	126	recombin	123	profession	119	reactor	116	barrel	113
burger	126	hexos	123	oval	119	rumen	116	divis	113
chardonnay	126	moment	123	polymorph	119	cytoplasm	116	riboflavin	113
owc <sup>GK</sup>	126	itc	123	fmwc	119	antihypertens	116	optimis	113
scientist	125	afb <sup>GO</sup>	123	ova	119	malic	116	korean	113
experienc	125	iss	123	bsg <sup>GU</sup>	119	aga	116	amplif	113
panellist	125	eeml	123	oviductus	119	hca <sup>HA</sup>	116	fsis	113
permit	125	committe	122	resili	118	usfda	116	norovirus	113
cymen	125	basal	122	toxicolog	118	polysorb	116	domesticus	113
quadrupol	125	excret	122	milli	118	edb <sup>HB</sup>	116	coliti	113
genotox	125	emphas	122	trisulfid	118	cielab <sup>HC</sup>	115	tlc <sup>HE</sup>	113
vertic	125	volumetr	122	sulphat	118	pollut	115	intercept	113
lipophil	125	rehydr	122	edg	118	anhydr	115	allevi	113
betaein	125	ewe	122	tank	118	livestock	115	flavanol	113
protocatechu	125	intra	122	multipli	118	inorgan	115	yam	113
destabil	125	ehec	122	walnut	118	unstabl	115	psc <sup>HF</sup>	113
intoler	125	pva <sup>GP</sup>	122	tandem	118	macronutri	115	Rhus <sup>HG</sup>	113
hsd <sup>GL</sup>	125	microb	122	savori	118	invas	115	dssf <sup>HH</sup>	113
recruit	125	roll	122	span	118	streptococcus	115	nist	112
diminish	125	psychrotroph	122	healthier	118	dextros	115	eas	112
kitchen	125	brevi	122	int	118	comment	115	inth	112
mortal	125	organis	122	dmso	118	hydroperoxid	115	collaps	112
adipos	125	diallyl	122	lifespan	118	autumn	115	raffinos	112
strict	125	rana	122	oeo <sup>GV</sup>	118	colloid	115	arabinoxylan	112
nonetheless	125	mal	122	rom	118	colleagu	115	stepwis	112
belgium	125	satisfi	121	xylitol	118	acacia	115	joint	112
parsley	125	submit	121	conglutin	118	genera	115	dead	112
landrac	125	alanin	121	bioscienc	117	fusarium	115	cooper	112
artisan	125	cata	121	ecolog	117	bdl <sup>HD</sup>	115	decarboxyl	112
shiraz	125	impregn	121	hydrochlor	117	adipogenesi	115	candida	112
vip	125	octan	121	oryza	117	apt	115	bcm <sup>HI</sup>	112
calera	125	touch	121	liposom	117	myrtl	115	bsfl <sup>HJ</sup>	112
durian	125	mother	121	conscious	117	graduat	114	orientali	112
lesser	124	macro	121	vendor	117	centuri	114	helium	111
ethic	124	salvia	121	Polydextros	117	irrespect	114	tyrosol	111
xanthophyl	124	ffa <sup>GQ</sup>	121	xico	117	disintegr	114	coalesc	111
promin	124	ddgs <sup>GR</sup>	121	desorpt	117	donkey	114	defect	111
sinap	124	pyralida	121	uncook	117	keep	114	scarc	111
cie	124	tsf <sup>GS</sup>	121	alloc	117	leakag	114	rmse	111
pod	124	commod	120	mar	117	mannitol	114	lys	111
fewer	124	cite	120	tocotrienol	117	mesh	114	unidentifi	111
propyl	124	sole	120	synthas	117	proof	114	ethnic	111
listeriosi	124	EtOH <sup>GT</sup>	120	cheesemak	117	longissimus	114	manganes	111
galangin	124	come	120	ase <sup>GW</sup>	117	loop	114	rancid	111
feo <sup>GM</sup>	124	pie	120	pif <sup>GX</sup>	117	actin	114	moduli	111

<sup>GI</sup> total dissolved solids, <sup>GJ</sup> polymerized whey protein, <sup>GK</sup> several meanings, oak wood chips, original water content, <sup>GL</sup> Tukey honest significant difference, <sup>GM</sup> forsythia essential oil, <sup>GN</sup> food science and technology, <sup>GO</sup> aflatoxin B (there are several of them), <sup>GP</sup> polyvinyl alcohol, <sup>GQ</sup> free fatty acids, <sup>GR</sup> distillers dried grains solubles, <sup>GS</sup> tomato seed flour, <sup>GT</sup> a common abbreviation for ethyl alcohol, <sup>GU</sup> brewer's spent grains, <sup>GV</sup> oregano essential oils, <sup>GW</sup> accelerated solvent extraction in addition to other terms such as, aseptically, <sup>GX</sup> powdered infant formula, <sup>GY</sup> a type of a diet low in fermentable carbohydrates that may be used by people who suffer from irritable bowel syndrome, <sup>GZ</sup> solid phase microextraction, <sup>HA</sup> hierarchical cluster analysis, <sup>HB</sup> elderberry, <sup>HC</sup> a way of classifying colors, <sup>HD</sup> a type of meat, Brazilian dried loin, <sup>HE</sup> thin layer chromatography, <sup>HF</sup> peroxyl radical scavenging capacity, <sup>HG</sup> a genus of plants in the cashew family, <sup>HH</sup> defatted sunflower seed flour, <sup>HI</sup> beta casomorphin, <sup>HJ</sup> black soldier fly larvae

Table S2. Bigrams from Volumes 1-8 of the open source journal, Foods (frequency &gt; 99)

Term	Freq.	Term	Freq.	Term	Freq.	Term	Freq.
fatty acids	2764	experimental design	278	food matrices	160	cross contamination	124
standard deviation	1326	cross validation	270	gram positive	157	electrical conductivity	122
amino acids	1271	lipid peroxidation	267	gastrointestinal tract	153	food security	122
shelf life	1188	titratable acidity	267	specific volume	153	xanthan gum	119
freeze drying	947	linear regression	248	lactobacillus plantarum	151	kinetic parameters	118
food safety	830	functional foods	245	negative control	151	salmonella enterica	118
essential oils	750	free radicals	239	staphylococcus aureus	151	consumer acceptability	117
skim milk	655	protein digestibility	236	positive effect	148	starch gelatinization	117
ascorbic acid	648	functional properties	228	textural properties	146	chi square	114
anti inflammatory	621	antibacterial activity	227	food ingredients	145	clinical trial	114
mass spectrometry	610	independent variables	221	detection limit	143	culture medium	114
sensory evaluation	496	soluble solids	214	solvent extraction	142	nir spectroscopy	114
retention time(s)	473	stainless steel	213	correlation coefficients	140	saccharomyces cerevisiae	113
calibration curve	463	shear force	210	heavy metals	140	total solids	110
spray dry	439	relative humidity	204	proximate composition	139	celiac disease	108
antimicrobial activity	390	pre treatment	200	uht milk	134	hydrochloric acid	108
cell lines	360	positive control	198	animal feed	133	ambient temperature	107
antioxidant properties	344	supply chain	197	carrier gas	133	hedonic scale	107
inhibitory activity	308	blood pressure	192	climate change	133	negative effect	107
<i>Escherichia coli</i>	304	small intestine	191	analytical grade	132	pilot scale	106
Maillard reaction	301	physical properties	185	starter culture	132	refractive index	106
post hoc	300	pearson correlation	174	immune system	131	food grade	105
raw material	293	color parameters	170	negative correlation	131	microbial counts	104
listeria monocytogenes	286	standard curve	168	pasting properties	127	guar gum	102
reducing sugars	284	foodborne pathogens	163	pasteurized milk	126	gum arabic	101
internal standard	280	secondary metabolites	163	glycemic index	125	sourdough bread	100

Table S3. Trigrams from Volumes 1-8 of the open source journal, Foods (frequency &gt; 99)

Term	Freq.	Term	Freq.	Term	Freq.
limit of detection (lod)	563	descriptive sensory analysis	137	Polyunsaturated fatty acids	113
limit of quantification (loq)	468	activation energy (ea)	133	virgin olive oil	112
analysis of variance (anova)	413	moisture sorption isotherm	132	water holding capacity	112
elastic modulus (pa)	384	rte meat	132	free radical scavenging	111
viscous modulus (pa)	371	total phenolic content	126	gallic acid equivalents	110
sea urchin roe	191	lactic acid bacteria	122	essential amino acids	107
uv colorimetric method	181	radical scavenging activity	121	reactive oxygen species	103
constant flow rate	180	principal component analysis	121	folin ciocalteu reagent	250
particle size distribution	149	fatty acid composition	118	narrow banded uvb	100
randomized double blind	139	free amino acids	113		

Table S4. Four-grams from Volumes 1-8 of the open source journal, Foods (frequency &gt; 19)

Term	Freq.	Term	Freq.
principal component analysis (pca)	221	Tukey multiple range test	32
total phenolic content (tpc)	182	long chain fatty acids	31
high performance liquid chromatography	168	texture profile analysis (tpa)	31
free radical scavenging	110	tumor necrosis factor (tnf)	31
duncan multiple range test	85	body mass index (bmi)	30
partial least squares discriminant	83	colony forming units (cfu)	30
reactive oxygen species (ros)	74	bovine serum albumin (bsa)	29
gallic acid equivalents (gae)	65	lactobacillus delbrueckii subsp bulgaricus	27
least significant difference (lsd)	63	modified atmosphere packaging (MAP)	27
oxygen radical absorbance capacity	59	pulsed electric field (pef)	27
scanning electron microscopy (sem)	58	conjugated linoleic acid (cla)	25
short chain fatty acids	53	linear discriminant analysis (lda)	24
trolox equivalent antioxidant capacity	52	saturated fatty acids (sfa)	24
food and drug administration (fda)	50	total soluble solids (tss)	24
lactococcus lactis subsp lactis	48	whey protein isolate (wpi)	24
least squares discriminant analysis	43	low molecular weight peptides	23
water holding capacity (whc)	40	plate count agar (pca)	23
broad banded uv (uvb)	39	sodium dodecyl sulfate (sds)	23
differential scanning calorimetry (dsc)	39	acidic amino acid residues	22
high hydrostatic pressure (hhp)	39	defatted sunflower seed flour	22
response surface methodology (rsm)	36	low density lipoprotein (ldl)	22
polymerase chain reaction (pcr)	35	multiple reaction monitoring (mrm)	22
solid phase extraction (spe)	34	tomato based food products	21
phosphate buffered saline (pbs)	33	medium grain white rice	20
medium chain fatty acids	32	modified maize starch (mms)	20
relative standard deviation (rsd)	32	polycyclic aromatic hydrocarbons (pah)	20

Table S5. Five through nine-grams from Volumes 1-8 of the open source journal, Foods (frequency &gt; 4)

Term	Freq.	Term	Freq.
lactic acid bacteria (lab)	115	solid phase microextraction gas chromatography mass spectrometry (spme ms)	12
european food safety authority (efsa)	80	inducible nitric oxide synthase (inos)	12
gas chromatography mass spectrometry (gc ms)	79	half maximal inhibitory concentration values	12
fatty acid methyl esters (fame)	78	oxygen radical absorbance capacity (orac) assay	11
ferric reducing antioxidant power assay (frap)	77	general linear model (glm) procedure	11
fourier transform infrared spectroscopy (ftir)	66	ferrie reducing antioxidant power assay (frap)	10
warner bratzler shear force (wbsf)	51	shiga toxin producing escherichia coli (stec)	10
thiobarbituric acid reactive substances (tbars)	46	gas chromatography flame ionization detector (gc fid)	9
narrow banded uv (nvb)	39	american association cereal chemists (aacc)	9
oxygen radical absorbance capacity (orac)	34	high performance liquid chromatography diode array detector (hplc dad)	8
trolox equivalent antioxidant capacity (teac)	33	ultra high performance liquid chromatography (uhplc)	8
extra virgin olive oil (evoo)	33	anova response surface quadratic model	8
sodium dodecyl sulfate polyacrylamide gel electrophoresis (sds page)	25	hydrophilic interaction liquid chromatography (hilic)	8
food agriculture organization united nations (fao)	25	partial least squares regression (plsrr)	8
united states department of agriculture (usda)	25	orthogonal partial least squares discriminant analysis (opls da)	7
total free amino acids (tfaa)	25	anova post hoc fisher lsd	7
total phenolic content antioxidant activity	24	high performance size exclusion chromatography	7
partial least squares discriminant analysis (pls da)	23	laser desorption ionization mass spectrometry	7
hplc high performance liquid chromatography	23	real time polymerase chain reaction	7
national institute standards technology (nist)	23	ultra high pressure homogenization (uhph)	7
reversed phase high performance liquid chromatography (hplc)	21	inductively coupled plasma optical emission spectrometry (icp oes)	6
high performance liquid chromatography (hplc)	21	liquid chromatography electrospray ionization tandem mass spectrometry	6
azino-bis-ethylbenzothiazoline sulphonate (abts)	18	fatty acid methyl esters (fame)	6
multivariate statistical process control (mspc)	16	field emission scanning electron microscopy	6
multivariate statistical process control (mspc)	16	brain heart infusion (bhi) broth	6
virgin olive oil extraction process	15	high performance liquid chromatography ultraviolet detection (hplc uv)	5
inductively coupled plasma mass spectrometry (icp-ms)	14	ultra high performance liquid chromatography mass spectrometry	5
high density lipoprotein cholesterol (hdl)	14	gas chromatography mass spectorscopy (gc ms)	5
hazard analysis critical control point (haccp)	13	high temperature short time htst pasteurization	5
inductively coupled plasma atomic emission spectroscopy (icp-aes)	13	american type culture collection (atcc)	5
enzyme linked immunosorbent assay (elisa)	13	high oxygen modified atmosphere packaging	5
fortified blended food base (fbfb)	13	human cervical cancer (hela) cells	5
long chain polyunsaturated fatty acids	13	total volatile basic (tvb) nitrogen	5
low density lipoprotein (ldl) cholesterol	13		