

Aborder la question de la distribution dans les champs

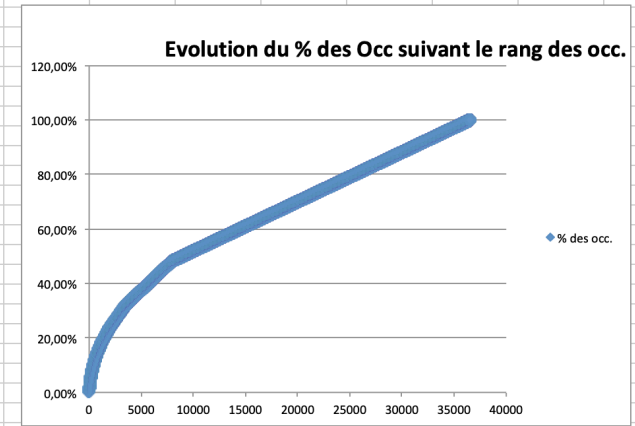
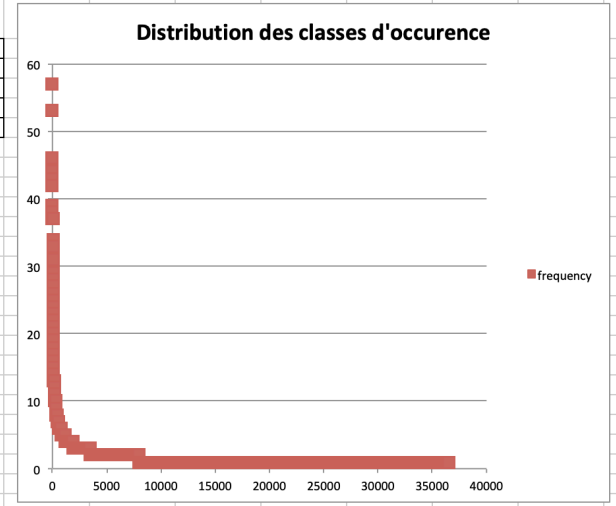
- List builder (régler avec 0 pour tout avoir)
- Travail du fichier dans un tableur
- Objectif: définir les bons seuils de paramétrage du nombre d'items à sélectionner dans les scripts

La volumétrie et la distribution comptent !

Champ: Author

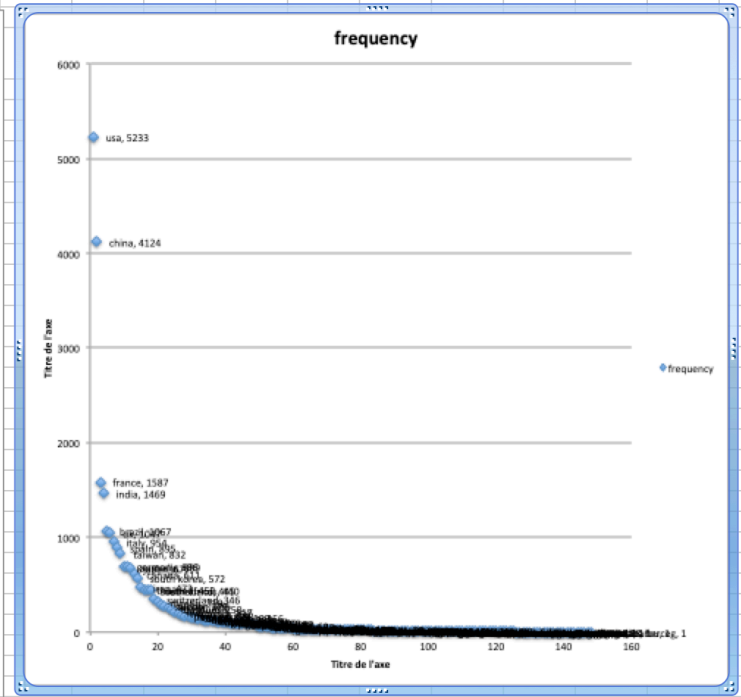
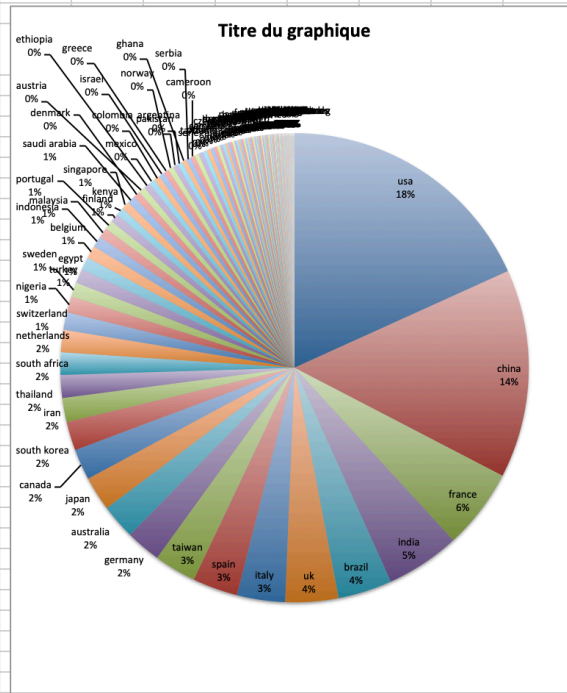
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
	entity	frequency	number of distinct documents	cumul of occur.	% des occ.	Rang des Aut	% des aut														
1	wang, y	57	56	57	0,10%	1	0,003%														
2	li, j	53	52	110	0,20%	2	0,005%														
3	zhang, y	46	46	156	0,28%	3	0,008%														
4	zhang, l	44	42	200	0,36%	4	0,011%														
5	rosenthal, pj	43	43	243	0,44%	5	0,014%														
6	price, rn	42	42	285	0,52%	6	0,016%														
7	smith, pj	42	42	327	0,59%	7	0,019%														
8	zhang, j	39	39	366	0,66%	8	0,022%														
9	liu, y	38	38	404	0,73%	9	0,025%														
10	chibale, k	37	37	441	0,80%	10	0,027%														
11	li, y	37	37	478	0,87%	11	0,030%														
12	wang, j	34	34	512	0,93%	12	0,033%														
13	liu, j	34	34	546	0,99%	13	0,036%														
14	kumar, s	33	33	579	1,05%	14	0,038%														
15	chen, y	31	31	610	1,11%	15	0,041%														
16	nosten, f	29	29	639	1,16%	16	0,044%														
17	egan, tj	28	28	667	1,21%	17	0,047%														
18	wang, xy	28	28	695	1,26%	18	0,049%														
19	white, nj	28	28	723	1,31%	19	0,052%														
20	de kock, c	28	28	751	1,36%	20	0,055%														
21	yang, y	27	26	778	1,41%	21	0,057%														
22	li, x	26	26	804	1,46%	22	0,060%														
23	wang, l	25	25	829	1,50%	23	0,063%														
24	cui, lw	25	25	854	1,55%	24	0,066%														
25	zhao, y	25	25	879	1,59%	25	0,068%														
26	pradines, b	25	25	904	1,64%	26	0,071%														
27	zhou, y	25	25	929	1,68%	27	0,074%														
28	wang, q	25	25	954	1,73%	28	0,077%														
29	zhang, h	25	23	979	1,77%	29	0,079%														
30	raoult, d	24	24	1003	1,82%	30	0,082%														
31	smith, gs	24	24	1027	1,86%	31	0,085%														
32	zhang, x	24	24	1051	1,91%	32	0,088%														
33	sun, y	24	24	1075	1,95%	33	0,090%														
34	wang, h	23	23	1098	1,99%	34	0,093%														
35	gut, j	23	23	1121	2,03%	35	0,096%														
36	mueller, i	23	23	1144	2,07%	36	0,098%														
37	zhang, q	23	23	1167	2,12%	37	0,101%														
38	chen, w	22	22	1189	2,16%	38	0,104%														
39	wang, x	22	22	1211	2,20%	39	0,107%														
40	amaravadi, rk	22	22	1233	2,24%	40	0,109%														
41	kumar, a	22	22	1255	2,27%	41	0,112%														
42	kumar, v	22	22	1277	2,31%	42	0,115%														
43	fidock, da	22	22	1299	2,35%	43	0,118%														
44	anstey, nm	21	21	1320	2,39%	44	0,120%														
45	li, m	21	21	1341	2,43%	45	0,123%														
46	liu, x	21	21	1362	2,47%	46	0,126%														
47	costedoat-chalumeau, n	20	20	1382	2,51%	47	0,129%														
48	lacerda, mvg	20	20	1402	2,54%	48	0,131%														
49	tarnine, i	20	20	1422	2,58%	49	0,134%														

Nb de lignes	36554	moyenne	1,50917845
somme occ	55165		
nb d'auteurs	36553		
sup rang 26	sup 23	Nb d'aut. Par	cumul de Class d'occ;
rang 25	24	36561	
rang 24	31	36558	
rang 23	35	36558	
rang 22	39	36560	
rang 21	45	36557	
rang 20	48	36560	
rang 18	54	36568	
rang 18	68	36564	
rang 17	78	36565	
rang 16	89	36569	
rang 15	104	36565	
rang 14	115	36570	
rang 13	131	36575	
rang 12	152	36593	
rang 11	191	36593	
rang 10	230	36602	
rang 9	278	36631	
rang 8	355	36661	
rang 7	462	36696	
rang 6	604	36754	
rang 5	804	36931	
rang 4	1181	37272	
rang 3	1899	38130	
rang 2	3475	33079	
rang 1	7939	28615	
rang 0	36554		



Champ: Country

	28 691	10 556			
entity	frequency	number of distinct documents	Part de marché des aut	Part de marché des Doc	Ratio E/D
usa	5233	1811	18,2%	17,2%	0,94
china	4124	1438	14,4%	13,6%	0,95
france	1587	386	5,5%	3,7%	0,66
india	1469	602	5,1%	5,7%	1,11
brazil	1067	287	3,7%	2,7%	0,73
uk	1047	508	3,6%	4,8%	1,32
italy	954	302	3,3%	2,9%	0,86
spain	895	231	3,1%	2,2%	0,70
taiwan	832	141	2,9%	1,3%	0,46
germany	696	289	2,4%	2,7%	1,13
australia	689	264	2,4%	2,5%	1,04
japan	677	284	2,4%	2,7%	1,14
canada	611	211	2,1%	2,0%	0,94
south korea	572	216	2,0%	2,0%	1,03
iran	473	116	1,6%	1,1%	0,67
thailand	452	199	1,6%	1,9%	1,20
south africa	445	184	1,6%	1,7%	1,12
netherlands	440	136	1,5%	1,3%	0,84
switzerland	346	180	1,2%	1,7%	1,41
nigeria	319	121	1,1%	1,1%	1,03
sweden	286	128	1,0%	1,2%	1,22
turkey	279	109	1,0%	1,0%	1,06
egypt	255	101	0,9%	1,0%	1,08
belgium	250	116	0,9%	1,1%	1,26
indonesia	217	71	0,8%	0,7%	0,89
malaysia	201	87	0,7%	0,8%	1,18
portugal	182	79	0,6%	0,7%	1,18
finland	173	49	0,6%	0,5%	0,77
kenya	159	79	0,6%	0,7%	1,35
singapore	157	73	0,5%	0,7%	1,26
saudi arabia	156	93	0,5%	0,9%	1,62
denmark	138	64	0,5%	0,6%	1,26
austria	135	59	0,5%	0,6%	1,19
ethiopia	125	52	0,4%	0,5%	1,13
mexico	120	42	0,4%	0,4%	0,95
colombia	117	57	0,4%	0,5%	1,21



Champ: Cited Reference (Top 1000)

	A	B
1	entity (top 1000)	frequency
2	trager w_1976_science	360
3	dondorp am_2009_new engl j med	312
4	gautret p_2020_int j antimicrob ag	292
5	mizushima n_2010_cell	257
6	wang ml_2020_cell res	238
7	ariey f_2014_nature	238
8	klionsky dj_2016_autophagy	223
9	who_2013_world malaria report 2013	205
10	ashley ea_2014_new engl j med	203
11	fidock da_2000_mol cell	198
12	rangwala r_2014_autophagy	193
13	hochberg mc_1997_arthritis rheum	188
14	klionsky dj_2012_autophagy	188
15	huang cl_2020_lancet	186
16	levine b_2008_cell	175
17	noedi h_2008_new engl j med	167
18	mizushima n_2008_nature	165
19	lambros c_1979_j parasitol	164
20	yao xt_2020_clin infect dis	164
21	gao jj_2020_biosci trends	162
22	kabeya y_2000_embo j	161
23	makler mt_1993_am j trop med hyg	151
24	who_2015_guid treatm mal	149
25	liu j_2020_cell discov	148
26	marmor mf_2016_opthalmology	143
27	solomon vr_2009_eur j pharmacol	140
28	mosmann t_1983_j immunol methods	134
29	cao b_2020_new engl j med	133
30	smilkstein m_2004_antimicrob agents ch	129
31	kimura t_2013_cancer res	126
32	marmor mf_2011_opthalmology	125
33	petri m_2012_arthritis rheum-us	123
34	djimde a_2001_new engl j med	122
35	ruiz-irastorza g_2010_ann rheum dis	122
36	mizushima n_2011_cell	121
37	vincent mj_2005_virol j	117
38	zhou f_2020_lancet	117
39	phyo ap_2012_lancet	115
40	pankiv s_2007_j biol chem	115
41	savarino a_2003_lancet infect dis	110
42	sidhu abs_2002_science	109
43	colson p_2020_int j antimicrob ag	109
44	amaravadi rk_2007_j clin invest	107
45	guan w_2020_new engl j med	106
46	yang znj_2011_mol cancer ther	106
47	mizushima n_2007_gene dev	106
48	rosenfeld mr_2014_autophagy	106
49	amaravadi rk_2011_clin cancer res	106
50	price rn_2004_lancet	104
51	yuhita s_2012_nat rev cancer	103

