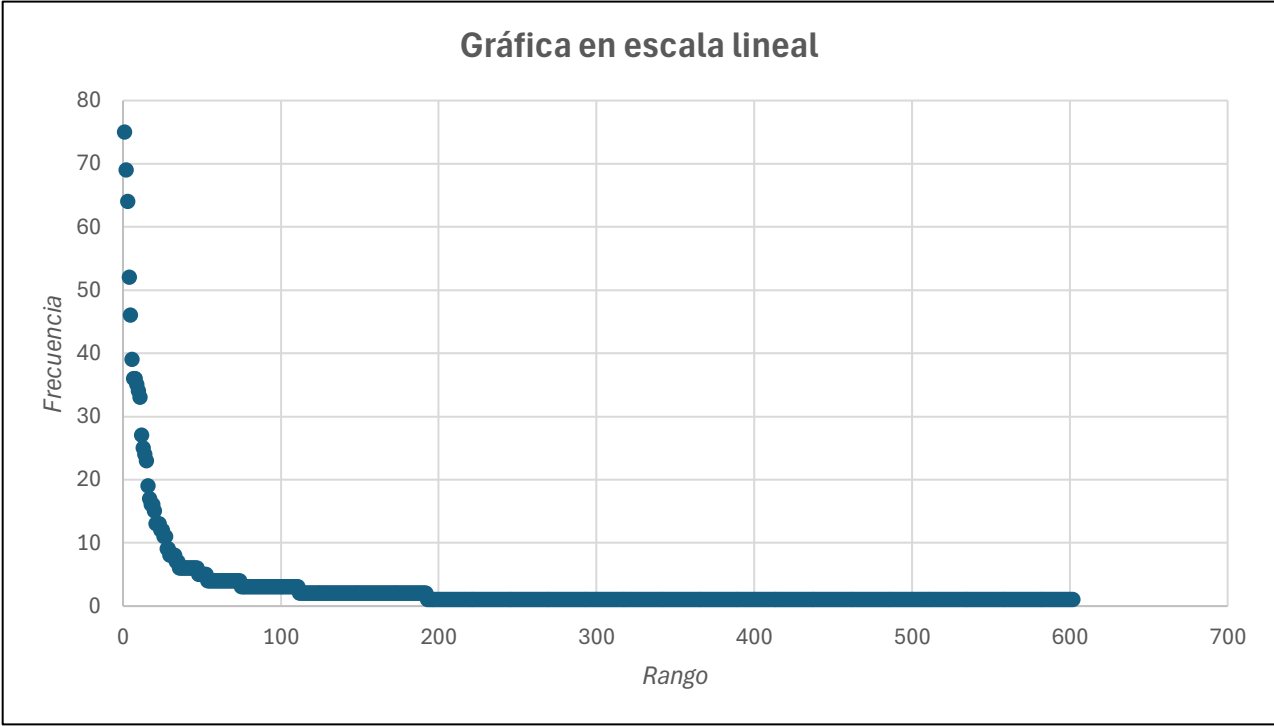


TALLER N. 04: Ley de Zipf

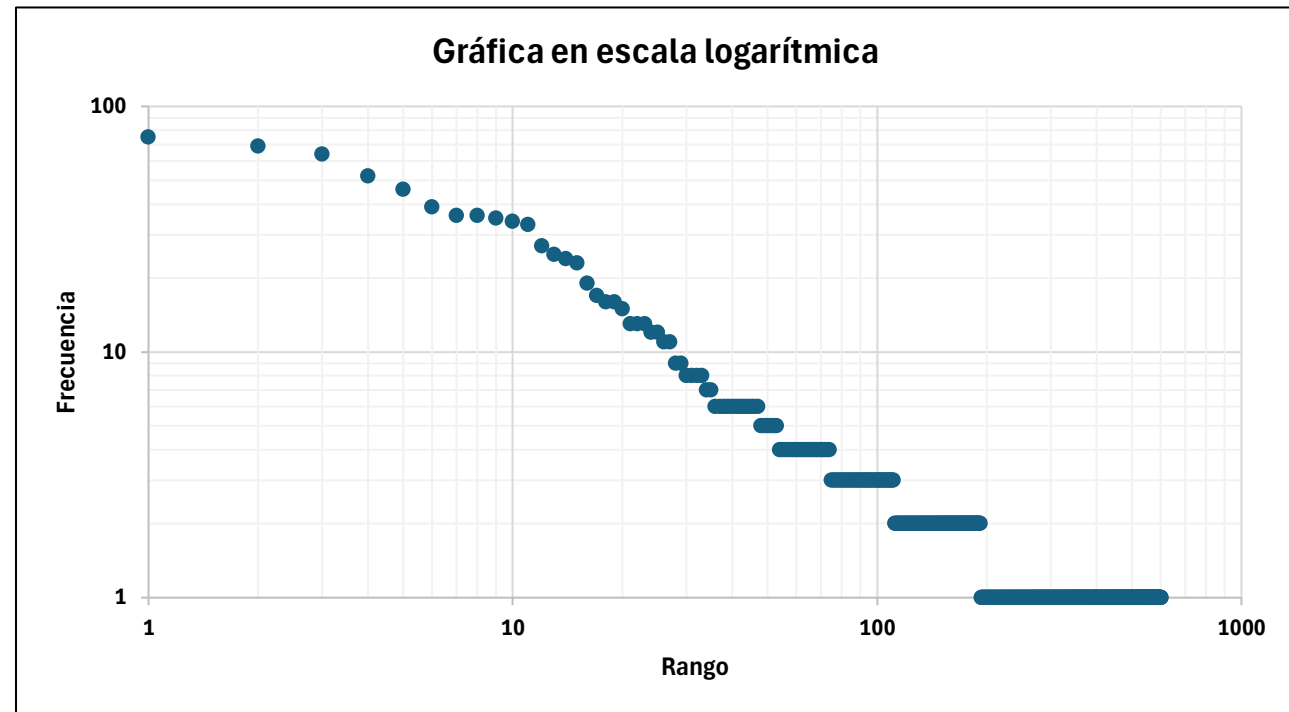
Integrantes: Guaña Karen, Ichau Jeniffer, Pérez Erick

Palabra	Frecuencia	Rango
de	75	1
que	69	2
la	64	3
el	52	4
y	46	5
se	39	6
dursley	36	7
no	36	8
los	35	9
en	34	10
a	33	11
un	27	12
su	25	13
señor	24	14
una	23	15
con	19	16
había	17	17
al	16	18
lo	16	19
las	15	20
estaba	13	21
señora	13	22
pero	13	23
era	12	24
por	12	25
le	11	26
mientras	11	27



TALLER N. 04: Ley de Zipf

potter	9	28
para	9	29
lechuzas	8	30
del	8	31
algo	8	32
gato	8	33
aquel	7	34
tenía	7	35
gente	6	36
día	6	37
dudley	6	38
si	6	39
han	6	40
hermana	6	41
como	6	42
eso	6	43
taladros	6	44
muy	6	45
cuando	6	46
noche	6	47
tenían	5	48
niño	5	49
porque	5	50
más	5	51
capa	5	52
sí	5	53
aquella	4	54
él	4	55
sobre	4	56
qué	4	57
hijo	4	58



Las gráficas muestran que frecuencia de las palabras más usadas con respecto a las de las menos comunes va disminuyendo rápidamente a medida que aumenta su posición en el ranking. Al observar la gráfica en escala logarítmica, se aprecia una línea recta, lo cual confirma la Ley de Zipf. Esta ley establece que la frecuencia de una palabra es inversamente proporcional a su posición, siguiendo una distribución de potencia.

TALLER N. 04: Ley de Zipf

PREGUNTAS

¿Cómo afecta la alta frecuencia de palabras comunes (como stopwords) en la eficiencia de un sistema de búsqueda?

La alta frecuencia de palabras comunes, como stopwords, puede impactar negativamente la eficiencia de un sistema de búsqueda. Estas palabras tienden a aparecer en muchos documentos, lo que puede generar una gran cantidad de resultados irrelevantes. Además, al tener un peso excesivo en el cálculo de la relevancia, pueden hacer que documentos menos relevantes aparezcan en las primeras posiciones de los resultados de búsqueda.

¿Qué técnicas se pueden utilizar para manejar palabras extremadamente frecuentes y palabras raras en sistemas de Recuperación de Información?

Se podría utilizar el filtrado de stopwords para eliminar términos comunes del índice de búsqueda, el uso de la ponderación TF-IDF para atenuar el efecto de las palabras comunes y resaltar términos más informativos y la indexación inversa para indexar solo términos significativos.

¿Cómo se pueden usar estos hallazgos para mejorar la precisión y el recall de un motor de búsqueda?

Para mejorar la precisión en los resultados de búsqueda, podemos quitar las stopwords y ajustar los pesos de las palabras usando técnicas como TF-IDF. Esto ayuda a reducir los falsos positivos y a obtener resultados más relevantes. Además, para mejorar el recall, podemos aplicar técnicas como el stemming, que nos permiten recuperar documentos con diferentes formas de las palabras buscadas, aumentando la cantidad de documentos relevantes encontrados.

ni	4	59
ser	4	60
visto	4	61
cuenta	4	62
hasta	4	63
casa	4	64
dijo	4	65
tiempo	4	66
vio	4	67
tan	4	68
es	4	69
personas	4	70
casi	4	71
sin	4	72
eran	4	73
tenido	4	74
llamado	3	75
otra	3	76
esquina	3	77
hombre	3	78
drive	3	79
hoy	3	80
nada	3	81
estaban	3	82
fuera	3	83
así	3	84
mañana	3	85
volvió	3	86
llevaba	3	87
entró	3	88
vez	3	89

TALLER N. 04: Ley de Zipf

mayor	3	90
debía	3	91
mirada	3	92
calle	3	93
nunca	3	94
coche	3	95
alejó	3	96
también	3	97
seguro	3	98
trató	3	99
privet	3	100
sus	3	101
normal	3	102
quedó	3	103
harry	3	104
habitual	3	105
número	3	106
4	3	107
aunque	3	108
después	3	109
poco	3	110
gran	3	111
momento	2	112
ventana	2	113
volando	2	114
luz	2	115
pensando	2	116
atigrado	2	117
lechuza	2	118
cabeza	2	119
fugaces	2	120

TALLER N. 04: Ley de Zipf

cogió	2	121
jim	2	122
pudo	2	123
estrellas	2	124
dio	2	125
durante	2	126
ciudad	2	127
plano	2	128
entre	2	129
decía	2	130
unas	2	131
todas	2	132
podía	2	133
oficina	2	134
hubiera	2	135
voz	2	136
puerta	2	137
concentrarse	2	138
aves	2	139
todavía	2	140
siquiera	2	141
alta	2	142
dejó	2	143
apresuró	2	144
cinco	2	145
humor	2	146
hacia	2	147
dirigirse	2	148
panadería	2	149
grupo	2	150
tarde	2	151

TALLER N. 04: Ley de Zipf

usted	2	152
ha	2	153
fue	2	154
bueno	2	155
iba	2	156
esposa	2	157
esperaba	2	158
decirle	2	159
conducta	2	160
sólo	2	161
extraña	2	162
celebrar	2	163
antes	2	164
jóvenes	2	165
nueva	2	166
unos	2	167
cerca	2	168
darse	2	169
dos	2	170
ocurrió	2	171
siempre	2	172
vecinos	2	173
cielo	2	174
lugar	2	175
veían	2	176
toda	2	177
todo	2	178
querían	2	179
pasaba	2	180
grunnings	2	181
mejor	2	182

TALLER N. 04: Ley de Zipf

acera	2	183
ya	2	184
extraños	2	185
misterioso	2	186
pequeño	2	187
desde	2	188
hacia	2	189
cuello	2	190
o	2	191
temor	2	192
muggles	1	193
empresa	1	194
ido	1	195
este	1	196
deberían	1	197
finalmente	1	198
llamada	1	199
anciano	1	200
sabe	1	201
quien	1	202
alegrarse	1	203
hay	1	204
molestarme	1	205
puede	1	206
fabricaba	1	207
querido	1	208
mi	1	209
disculpe	1	210
feliz	1	211
alegremente	1	212
abrazó	1	213

TALLER N. 04: Ley de Zipf

completament	1	214
mismo	1	215
jardín	1	216
pared	1	217
sentado	1	218
encontrado	1	219
mejoró	1	220
primero	1	221
camino	1	222
tales	1	223
imaginación	1	224
aprobaba	1	225
tonterías	1	226
deseado	1	227
suyas	1	228
imaginaciones	1	229
fueran	1	230
deseando	1	231
subir	1	232
desconcertado	1	233
importaba	1	234
muggle	1	235
director	1	236
pasaban	1	237
abrazado	1	238
helado	1	239
desconocido	1	240
corpulento	1	241
atención	1	242
mención	1	243
ante	1	244

TALLER N. 04: Ley de Zipf

mucho	1	245
trastornaba	1	246
preocupar	1	247
sentido	1	248
harold	1	249
harvey	1	250
llamarse	1	251
podría	1	252
llamara	1	253
sobrino	1	254
pensándolo	1	255
llamaban	1	256
muchísimas	1	257
inmenso	1	258
especial	1	259
delgada	1	260
apellido	1	261
estúpido	1	262
comportando	1	263
pensaba	1	264
bigotes	1	265
atusó	1	266
aparato	1	267
rubia	1	268
idea	1	269
cambió	1	270
cualquier	1	271
reprochárselo	1	272
llamaba	1	273
bigote	1	274
chillona	1	275

TALLER N. 04: Ley de Zipf

líneas	1	276
sonrisa	1	277
amplia	1	278
iluminó	1	279
rostro	1	280
contrario	1	281
empujón	1	282
disgustado	1	283
parecía	1	284
violeta	1	285
segundos	1	286
suelo	1	287
caía	1	288
tambaleaba	1	289
viejo	1	290
diminuto	1	291
gruñó	1	292
perdón	1	293
rollizo	1	294
chocó	1	295
preocupado	1	296
punto	1	297
edificio	1	298
costó	1	299
modos	1	300
todos	1	301
pues	1	302
movió	1	303
idénticas	1	304
apartados	1	305
tal	1	306

TALLER N. 04: Ley de Zipf

vivían	1	307
orgullosos	1	308
chaparrón	1	309
tuvieron	1	310
ayer	1	311
prometí	1	312
lluvia	1	313
decirme	1	314
telefoneado	1	315
dundee	1	316
yorkshire	1	317
kent	1	318
lugares	1	319
alrededor	1	320
espectadore	1	321
actitud	1	322
sé	1	323
meteorólogo	1	324
ted	1	325
esta	1	326
lluvias	1	327
habrá	1	328
pronóstico	1	329
mcguffin	1	330
decir	1	331
nuevo	1	332
ahora	1	333
comenzado	1	334
hogueras	1	335
semana	1	336
viene	1	337

TALLER N. 04: Ley de Zipf

molesta	1	338
pareció	1	339
esperado	1	340
tu	1	341
últimamente	1	342
sabido	1	343
has	1	344
querida	1	345
petunia	1	346
eh	1	347
nerviosismo	1	348
garganta	1	349
aclaró	1	350
bien	1	351
aquello	1	352
té	1	353
tazas	1	354
comedor	1	355
cuchicheo	1	356
rumor	1	357
bretaña	1	358
sillón	1	359
congelado	1	360
lluviosa	1	361
prometerles	1	362
puedo	1	363
señores	1	364
irónica	1	365
mueca	1	366
permitió	1	367
último	1	368

TALLER N. 04: Ley de Zipf

ver	1	369
salón	1	370
acostaron	1	371
normalidad	1	372
comportarse	1	373
haré	1	374
frase	1	375
aprendido	1	376
contó	1	377
hija	1	378
contigua	1	379
problemas	1	380
informó	1	381
comían	1	382
esperaría	1	383
encontrar	1	384
relacionadas	1	385
decidido	1	386
seguía	1	387
calmarse	1	388
extraño	1	389
aquélla	1	390
preguntó	1	391
severa	1	392
dirigió	1	393
marcar	1	394
ojos	1	395
informativo	1	396
observadores	1	397
locutor	1	398
pájaros	1	399

TALLER N. 04: Ley de Zipf

sueño	1	400
horarios	1	401
cambiado	1	402
causa	1	403
explicar	1	404
incapaces	1	405
son	1	406
expertos	1	407
sol	1	408
salida	1	409
direcciones	1	410
estas	1	411
vuelo	1	412
avisos	1	413
cientos	1	414
producido	1	415
normales	1	416
verlas	1	417
difícil	1	418
cazan	1	419
habitualmente	1	420
pese	1	421
nación	1	422
informado	1	423
fortunadament	1	424
partes	1	425
últimas	1	426
números	1	427
echó	1	428
terminado	1	429
razón	1	430

TALLER N. 04: Ley de Zipf

meneó	1	431
planos	1	432
rótulos	1	433
leer	1	434
saben	1	435
gatos	1	436
sabían	1	437
habían	1	438
buena	1	439
rótulo	1	440
leyendo	1	441
felino	1	442
retrovisor	1	443
apareciesen	1	444
espejo	1	445
observó	1	446
mantener	1	447
subía	1	448
vuelta	1	449
daba	1	450
alejados	1	451
devolvió	1	452
éste	1	453
contempló	1	454
parpadeó	1	455
óptica	1	456
pensamientos	1	457
pensó	1	458
sidó	1	459
vestida	1	460
completo	1	461

TALLER N. 04: Ley de Zipf

llevaban	1	462
conjuntos	1	463
ah	1	464
ridícula	1	465
ropa	1	466
inútil	1	467
soportaba	1	468
opuesto	1	469
individuos	1	470
pudiera	1	471
forma	1	472
imaginar	1	473
pedidos	1	474
cantidad	1	475
advertir	1	476
dejar	1	477
matutino	1	478
embotellamiento	1	479
mente	1	480
apartó	1	481
estremecían	1	482
afueras	1	483
pensar	1	484
conseguir	1	485
dirían	1	486
ilusión	1	487
haber	1	488
teléfono	1	489
maletín	1	490
tendrían	1	491
tunante	1	492

TALLER N. 04: Ley de Zipf

paredes	1	493
contra	1	494
cereales	1	495
arrojando	1	496
berrinche	1	497
beso	1	498
despedirse	1	499
región	1	500
mejilla	1	501
besó	1	502
canturreaba	1	503
dientes	1	504
media	1	505
ocho	1	506
ponía	1	507
corbata	1	508
parda	1	509
sosa	1	510
ir	1	511
trabajo	1	512
ninguno	1	513
parloteaba	1	514
silla	1	515
ruidoso	1	516
misteriosos	1	517
salía	1	518
juntara	1	519
nubes	1	520
aquél	1	521
estado	1	522
ningún	1	523

TALLER N. 04: Ley de Zipf

nuestra	1	524
historia	1	525
comienza	1	526
mirar	1	527
despertaron	1	528
martes	1	529
luego	1	530
cubierto	1	531
segundo	1	532
grises	1	533
metió	1	534
amenazaban	1	535
mirando	1	536
tormenta	1	537
raro	1	538
sucedía	1	539
indicio	1	540
primer	1	541
percibió	1	542
nublado	1	543
llegar	1	544
sugiriera	1	545
contecimiento	1	546
supuso	1	547
moda	1	548
marido	1	549
enfrente	1	550
agitación	1	551
susurraba	1	552
nervioso	1	553
ponían	1	554

TALLER N. 04: Ley de Zipf

sabía	1	555
enfadado	1	556
miró	1	557
pasar	1	558
lado	1	559
parte	1	560
pasó	1	561
olvidado	1	562
estirándolo	1	563
regresaba	1	564
encima	1	565
piernas	1	566
estirar	1	567
decidió	1	568
comida	1	569
hora	1	570
valla	1	571
jardines	1	572
buen	1	573
estuvo	1	574
gritar	1	575
importantes	1	576
hucha	1	577
donut	1	578
tamborileó	1	579
largo	1	580
molestaran	1	581
quería	1	582
secretaria	1	583
gritos	1	584
correr	1	585

TALLER N. 04: Ley de Zipf

instalaba	1	586
cruzar	1	587
doble	1	588
contuvo	1	589
decirles	1	590
quisiera	1	591
murmuraban	1	592
invadió	1	593
gigante	1	594
petrificado	1	595
resultaba	1	596
útil	1	597
oído	1	598
he	1	599
conversación	1	600
palabras	1	601
pocas	1	602
oír	1	603
alcanzó	1	604
papel	1	605
bolsa	1	606
telefónicas	1	607
llamadas	1	608
hizo	1	609
vestía	1	610
avanzó	1	611
tráfico	1	612
supiera	1	613
colecta	1	614
evidente	1	615
publicitaria	1	616

TALLER N. 04: Ley de Zipf

tontería	1	617
alguna	1	618
entonces	1	619
valor	1	620
esmeralda	1	621
verde	1	622
incluso	1	623
espiar	1	624
uno	1	625
vamos	1	626
desconocidos	1	627
enfureció	1	628
excitados	1	629
cuchicheaban	1	630
años	1	631
tanto	1	632
posó	1	633
volante	1	634
fingía	1	635
dedos	1	636
minutos	1	637
soportado	1	638
llegó	1	639
aparcamiento	1	640
gritó	1	641
perfectamente	1	642
tuvo	1	643
embargo	1	644
ellos	1	645
aquellas	1	646
mayoría	1	647

TALLER N. 04: Ley de Zipf

tras	1	648
desfilaban	1	649
abierta	1	650
boca	1	651
señalaban	1	652
pleno	1	653
volaban	1	654
costado	1	655
habría	1	656
secreto	1	657
hecho	1	658
piso	1	659
noveno	1	660
descubriesen	1	661
espaldas	1	662
sentaba	1	663
habrían	1	664
nuevamente	1	665
enfadada	1	666