

Table 1: Frequency of feature appearance in Projects, Files and Patterns, with number of tokens observed and the maximum number of tokens observed in a single pattern.

Rank	Code	Example	% Projects	NProjects	NFiles	NPatterns	NTokens	MaxTokens
1	ADD	z+	73.2	1,204	9,165	6,003	11,136	30
2	CG	(caught)	72.6	1,194	9,559	7,130	12,707	17
3	KLE	.*	66.8	1,099	8,163	6,017	11,620	50
4	CCC	[aeiou]	62.4	1,026	7,648	4,468	8,179	42
5	ANY	.	61.1	1,005	6,277	4,657	7,119	60
6	RNG	[a-z]	51.6	848	5,092	2,631	8,043	50
7	STR	^	51.4	846	5,458	3,563	3,661	12
8	END	\$	50.3	827	5,393	3,169	3,276	12
9	NCCC	[^qwx]	47.2	776	3,947	1,935	2,718	15
10	WSP	\s	46.3	762	4,704	2,846	6,128	32
11	OR	a b	43	708	3,926	2,102	2,606	15
12	DEC	\d	42.1	692	4,198	2,297	4,868	24
13	WRD	\w	39.5	650	2,952	1,430	2,037	13
14	QST	z?	39.2	645	3,707	1,871	3,290	35
15	LZY	z+?	36.8	605	2,221	1,300	1,761	12
16	NCG	a(?:b)c	24.6	404	1,709	791	1,453	28
17	PNG	(?P<name>x)	21.5	354	1,475	915	2,399	16
18	SNG	z{8}	20.7	340	1,267	581	1,159	17
19	NWSP	\S	16.4	270	776	484	676	10
20	DBB	z{3,8}	14.5	238	647	367	573	11
21	NLKA	a(?:!yz)	11.1	183	489	131	148	3
22	WNW	\b	10.1	166	438	248	408	36
23	NWRD	\W	10	165	305	94	149	6
24	LWB	z{15,}	9.6	158	281	91	107	3
25	LKA	a(?:=bc)	9.6	158	358	112	133	4
26	OPT	(?i)CasE	9.4	154	377	231	238	2
27	NLKB	(?<!x)yz	8.3	137	296	94	117	4
28	LKB	(?<=a)bc	7.3	120	255	80	99	4
29	ENDZ	\Z	5.5	90	149	89	89	1
30	BKR	\1	5.1	84	129	60	73	4
31	NDEC	\D	3.5	58	92	36	51	6
32	BKRN	(P?=name)	1.7	28	44	17	19	2
33	VWSP	\v	0.9	15	16	13	14	2
34	NWNW	\B	0.7	11	11	4	5	2