

Table 1: How Frequently do Features Appear in Projects, and Which Features are Supported By Four Major Regex Projects? (Table 4)

rank	code	description	example	brics	hampi	Rex	RE2	nPatterns	% patterns	nProjects	% projects
1	ADD	one-or-more repetition	<code>z+</code>	●	●	●	●	6,003	44.1	1,204	73.2
2	CG	a capture group	<code>(caught)</code>	●	●	●	●	7,130	52.4	1,194	72.6
3	KLE	zero-or-more repetition	<code>.*</code>	●	●	●	●	6,017	44.3	1,099	66.8
4	CCC	custom character class	<code>[aeiou]</code>	●	●	●	●	4,468	32.9	1,026	62.4
5	ANY	any non-newline char	<code>.</code>	●	●	●	●	4,657	34.3	1,005	61.1
6	RNG	chars within a range	<code>[a-z]</code>	●	●	●	●	2,631	19.3	848	51.6
7	STR	start-of-line	<code>^</code>	○	●	●	●	3,563	26.2	846	51.4
8	END	end-of-line	<code>\$</code>	○	●	●	●	3,169	23.3	827	50.3
9	NCCC	negated CCC	<code>[^qwx]</code>	●	●	●	●	1,935	14.2	776	47.2
10	WSP	<code>\t \n \r \v \f</code> or space	<code>\s</code>	○	●	●	●	2,846	20.9	762	46.3
11	OR	logical or	<code>a b</code>	●	●	●	●	2,102	15.5	708	43
12	DEC	any of: 0123456789	<code>\d</code>	○	●	●	●	2,297	16.9	692	42.1
13	WRD	<code>[a-zA-Z0-9_]</code>	<code>\w</code>	○	●	●	●	1,430	10.5	650	39.5
14	QST	zero-or-one repetition	<code>z?</code>	●	●	●	●	1,871	13.8	645	39.2
15	LZY	as few reps as possible	<code>z+?</code>	○	●	○	●	1,300	9.6	605	36.8
16	NCG	group without capturing	<code>a(?:b)c</code>	○	●	○	●	791	5.8	404	24.6
17	PNG	named capture group	<code>(?P&lt;name&gt;x)</code>	○	●	○	●	915	6.7	354	21.5
18	SNG	exactly n repetition	<code>z{8}</code>	●	●	●	●	581	4.3	340	20.7
19	NWSP	any non-whitespace	<code>\S</code>	○	●	●	●	484	3.6	270	16.4
20	DBB	$n \leq x \leq m$ repetition	<code>z{3,8}</code>	●	●	●	●	367	2.7	238	14.5
21	NLKA	sequence doesn't follow	<code>a(?!yz)</code>	○	○	○	○	131	1	183	11.1
22	WNW	word/non-word boundary	<code>\b</code>	○	○	○	●	248	1.8	166	10.1
23	NWRD	non-word chars	<code>\W</code>	○	●	●	●	94	0.7	165	10
24	LWB	at least n repetition	<code>z{15,}</code>	●	●	●	●	91	0.7	158	9.6
25	LKA	matching sequence follows	<code>a(?=bc)</code>	○	○	○	○	112	0.8	158	9.6
26	OPT	options wrapper	<code>(?i)CasE</code>	○	●	○	●	231	1.7	154	9.4
27	NLKB	sequence doesn't precede	<code>(?&lt;!x)yz</code>	○	○	○	○	94	0.7	137	8.3
28	LKB	matching sequence precedes	<code>(?&lt;=a)bc</code>	○	○	○	○	80	0.6	120	7.3
29	ENDZ	absolute end of string	<code>\Z</code>	○	○	○	●	89	0.7	90	5.5
30	BKR	match the $i^{th}$ CG	<code>\1</code>	○	○	○	○	60	0.4	84	5.1
31	NDEC	any non-decimal	<code>\D</code>	○	●	●	●	36	0.3	58	3.5
32	BKRN	references PNG	<code>(?P=name)</code>	○	●	○	○	17	0.1	28	1.7
33	VWSP	matches U+000B	<code>\v</code>	○	○	●	●	13	0.1	15	0.9
34	NWNW	negated WNW	<code>\B</code>	○	○	○	●	4	0	11	0.7