REGEX REFERENCE SHEET II

Symbol Index

Core Building Blocks

Type	Symbol	Meaning / Example
Any word character	\w	letter, number, or underscore (A–Z, 0–9, _)
Any digit	\d	0–9
Any whitespace	\s	space, tab, newline
Any character	.	matches everything except newline
Specific set	[A-Z]	any uppercase letter
Negated set	[^A-Z]	any character **not** A–Z
Escape special	\. \+ \?	treat . + ? as normal characters

Quantifiers (How Many?)

Pattern	Reads As	Example
*	0 or more	a* → "", "aaa"
+	1 or more	a+ → "a", "aaaa"
?	0 or 1 (optional)	https? → http or https
{n}	exactly *n*	\d{4} → 4 digits
{m,}	at least *m*	[A-Z]{2,} → ≥2 capitals
{m,n}	between *m* and *n*	[A-Z]{2,5} → 2–5 capitals

Anchors & Boundaries

Symbol	Meaning	Example
^	start of string	^The → lines starting with *The*
$	end of string	end$ → lines ending with *end*
\b	word boundary	\bcat\b → matches "cat", not "scatter"
\B	not a boundary	\Bcat\B → inside a word

Grouping & Choice

Pattern	Reads As	Example
(abc)	capture group	remembers the match
(?:abc)	non-capturing group	groups without saving
`a	b`	a **or** b
(?P<name>abc)	named group	access by match.group('name')

Lookarounds (Advanced)

Type	Pattern	Meaning
Positive lookahead	(?=...)	must be followed by …
Negative lookahead	(?!...)	must **not** be followed by …
Positive lookbehind	(?<=...)	must be preceded by …
Negative lookbehind	(?<!...)	must **not** be preceded by …

Common Patterns (Examples)

Goal	Regex	Notes
Email	[\w\.-]+@[\w\.-]+\.\w{2,}	username + @ + domain + TLD
Year	\b\d{4}\b	4-digit number
Hashtag	(?<!\w)#[A-Za-z0-9_]+	start with #, stop at space
URL	https?://\S+	s optional, grab non-spaces
Twitter handle	`(?:(?<=\s)	^)@[A-Za-z0-9_]{1,15}\b`
UK postcode	[A-Z]{1,2}\d[A-Z\d]? \d[A-Z]{2}	simplified rule

Regex Flags (Python)

Flag	Description
re.I or (?i)	ignore case
re.M or (?m)	multi-line (^/$ per line)
re.S or (?s)	dot matches newline
re.X or (?x)	verbose mode (allow comments)