The Parts of a Regular Expression

A regular expression can be made up of combinations of the following:

- **Literals**. These have no additional meaning and are a one to one match. If you specify the literal "abc", the code will match on the first occurrence of "abc", in your string.
- **Character Classes**. Some of these are predefined, others you can define yourself. The period or dot is an example of a predefined character class.
- **Quantifiers**. These metacharacters identify the number of occurrences of a character class or literal, required to make a match. I used the asterisk, but there are several others I'll review shortly.
- **Boundary matchers**, or anchors. These specify the position in the text, for example at the start of the text or the end.
- **Groups**. These identify and allow for the capturing of subexpressions.

The Parts of a Regular Expression

The table on this slide displays some common metacharacters that fall into these categories.

Type	Examples
Character Classes	**. [abc] [a-g] [A-Z] [0-9] [^abc] \d \s \w**
Quantifiers	*** + ?**
Boundary matchers (or anchors)	**^ $ \b**
Groups	**()**

You can find these examples and more by looking at Java's Pattern Class API

https://docs.oracle.com/en/java/javase/17/docs/api/java.base/java/util/regex/Pattern.html

Character classes defined with square brackets

Characters in square brackets may have a different meaning.

Any character except ^, -,] or \ is a **literal**, when it's in the **square brackets**.

As an example, a dot in square brackets will represent a literal period, and not a meta character to match any character.

.	Any character except ^, -,] or \ is a **literal**, when it's in the **square brackets**.
[.]	Means a single period

Quantifiers

There are six different quantifiers you can use in your regular expressions.

Quantifier	Meaning	Pattern Example	Match Examples
*	pattern appears zero or more times	b*	empty string, b, bb, bbb
+	pattern appears one or more times	b+	b, bb, bbb
?	pattern appears zero or one time	colou?r	color, colour
{ n }	pattern must appear exactly n times	b{3}	bbb
{ n, }	pattern must appear at least n times	b{2,}	bb, bbb, bbbb
{ n, m }	pattern must appear at least n but not more than m times	b{3, 4}	bbb, bbbb

Boundary Matchers

There are three common boundary matchers or anchors.

metacharacter	Meaning	Pattern String	Match Notes
^	matches to start of text	"^."	Matches first character in a string
$	matches to end of text	".$"	Matches last character in a string
\b	matches to word	"\\b"	Matches first word in a string.