What does it mean to Compile a Regular Expression?

The string containing the expression is passed to the compile method, of a Pattern class, which returns a Pattern instance.

This string, the regular expression, is said to be **compiled into a Pattern**, by Java's regular expression processor.

This compilation process consists of

- Checking the string for syntactical correctness.
- Building an internal representation, a decision tree made up of nodes, and boolean decision points, derived from the various parts of the regular expression.
- Optimizing the pattern, by simplifying the expression and eliminating redundancies. This process increases the efficiency of matching the expression to character sequences.

Advantages of the Pattern Instance

By compiling regular expressions into Pattern objects, you benefit from improved performance and code efficiency, on subsequent matching.

The compiled pattern can be reused across multiple matching operations, saving computational resources, and reducing processing time.

Matcher class advantages

In addition to matching on the entire input, Matcher offers two other operations for **partial matching**.

These methods are **`lookingAt`**, and overloaded versions of the **`find`** method.

Matcher supports **capturing groups** and access to the text within the group, which I haven't yet covered, but which I'll cover shortly.

Matcher can be **reused** for multiple match operations on different String inputs, meaning the pattern doesn't have to be recompiled.

Matcher class disadvantages

An instance of a Matcher class **has state**, which changes as operations are performed on it.

This means the Matcher class is **not thread safe**.

It also means state may need to be reset, before a new String is evaluated.

Greedy and Reluctant Regular Expressions

Greedy regular expressions match as **many** characters as possible.

The expression .*, which is a greedy expression, matches any number of characters, including the empty string.

Reluctant regular expressions, on the other hand, match as **few** characters as possible from the input text.

The regular expression .*? matches any number of characters, but stops at the earliest successful point, where the overall pattern is matched.

The default type of regular expression is greedy.

You can use the ? to be a quantifier modifier, making the regular expression reluctant.