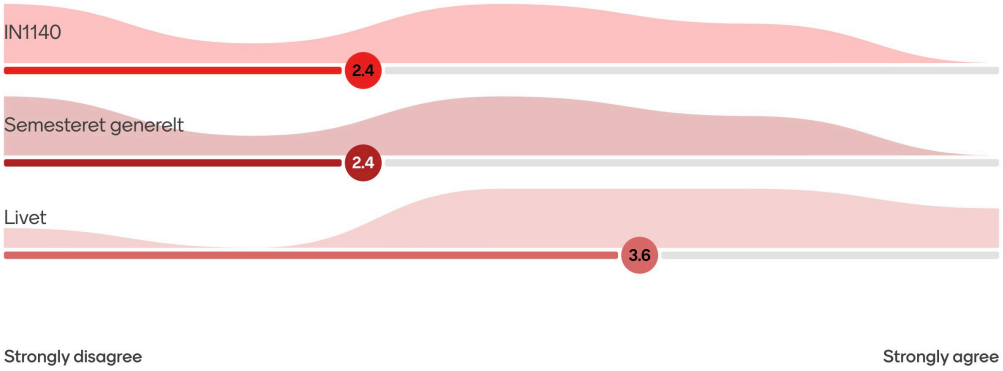
IN1140 Gruppetime #4

Regulære uttrykk og oblig 1b

Dagens gruppetime

- Mini menti
- Regulære uttrykk
- Oblig 1b

"Jeg har kontroll på..."



Non-capturing group!

- (?:)
- Del av pensum
- Praktisk for samensatte uttrykk
- For eksempel, ukedager

Eksempel med regex101.com



Uten non-capturing group



Med non-capturing group

Cheatography

Regular Expressions Cheat Sheet

by Dave Child (DaveChild) via cheatography.com/1/cs/5/

Anchors

^	Start of string, or start of line in multi-line pattern
\A	Start of string
$	End of string, or end of line in multi-line pattern
\Z	End of string
\b	Word boundary
\B	Not word boundary
\<	Start of word
\>	End of word

Character Classes

\c	Control character
\s	White space
\S	Not white space
\d	Digit
\D	Not digit
\w	Word
\W	Not word
\x	Hexadecimal digit
\O	Octal digit

POSIX

Assertions

?=	Lookahead assertion
?!	Negative lookahead
?<=	Lookbehind assertion
?!= or ?<!	Negative lookbehind
?>	Once-only Subexpression
?()	Condition [if then]
?()|	Condition [if then else]
?#	Comment

Quantifiers

*	0 or more	{3}	Exactly 3
+	1 or more	{3,}	3 or more
?	0 or 1	{3,5}	3, 4 or 5

Add a ? to a quantifier to make it ungreedy.

Escape Sequences

\	Escape following character
\Q	Begin literal sequence
\E	End literal sequence

"Escaping" is a way of treating characters which have a special meaning in regular expressions literally, rather than as special characters.

Groups and Ranges

.	Any character except new line (\n)
(a|b)	a or b
(...)	Group
(?:...)	Passive (non-capturing) group
[abc]	Range (a or b or c)
[^abc]	Not (a or b or c)
[a-q]	Lower case letter from a to q
[A-Q]	Upper case letter from A to Q
[0-7]	Digit from 0 to 7
\x	Group/subpattern number "x"

Ranges are inclusive.

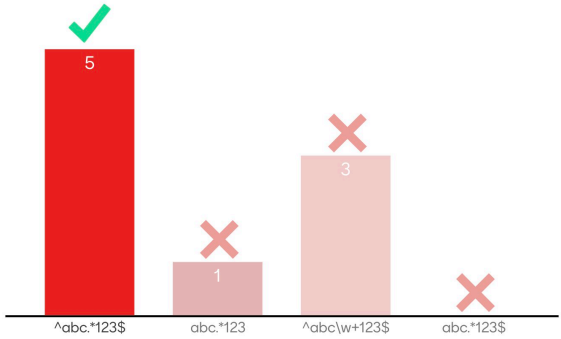
Pattern Modifiers

g	Global match
i *	Case-insensitive
m *	Multiple lines
s *	Treat string as single line
x *	Allow comments and whitespace in pattern
e *	Evaluate replacement
U *	Ungreedy pattern

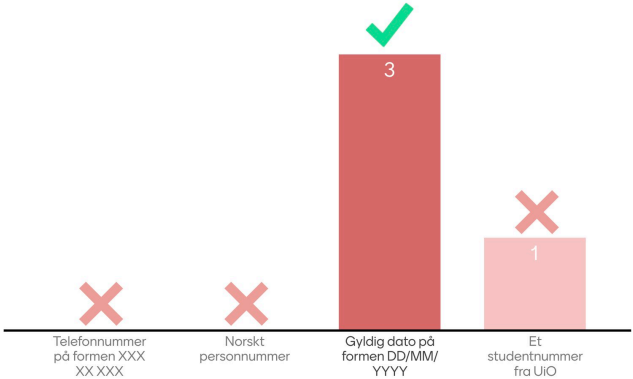
* PCRE modifier

Litt hjelp!

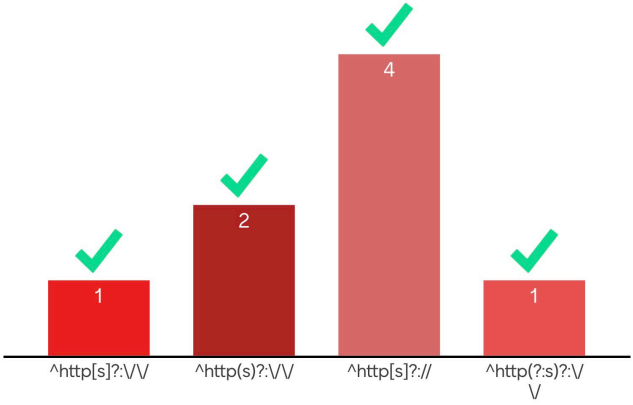
Hvilket uttrykk matcher en tekststreng som begynner med "abc" og slutter med "123", uansett hva som står imellom?



Hva matcher uttrykket ^(?:0[1-9]|[12][0-9]|3[01])/(?:0[1-9]|1[0-2])/(?:\d{4})$



Hvilket regulært uttrykk matcher en gyldig URL som starter med "http://" eller "https://"?



Oblig 1b!