

## Regular Expressions – Basics

### Lab RE1 – Gabor

Determine a regular expression which will match only on the indicated strings, or else will find the indicated matches. The form for the submission will be a command line script that takes a single integer as input from 30 to 39, inclusive, and outputs the corresponding regular expression pattern, as shown in class: the pattern is to be delimited by forward slashes and any options should immediately follow the final slash.

- ✓30. Determine whether a string is either 0, 100, or 101.
- ✓31. Determine whether a given string is a binary string (ie. composed only of 0 and 1 characters).
- ✓32. An integer (sub)string refers to a non-empty (sub)string that will convert to an integer but has no leading 0. Zero is represented as the single digit 0. Given a binary integer string, what regular expression determines whether it is even?
- ✓33. What is a regular expression to determine (ie. match) those words in a text that have at least two vowels?
- ✓34. Given a string, determine whether it is a non-negative, even binary integer string.
- ✓35. Determine whether a given string is a binary string containing 110 as a substring.
- ✓36. Match on all strings of length at least two, but at most four.
- ✓37. Validate a social security number entered into a field (ie. recognize ddd-dd-dddd where the d represents digits and where the dash indicates an arbitrary number of spaces with at most one dash). For example, 542786363, 542 786363, and 542 - 78-6263 are all considered valid.
- ✓38. Determine a regular expression to **help** you find the first word of each line of text with a d in it: Match through the end of the first word with a d on each line that has a d.
- 39. Determine whether a string is a binary string that has the same number of 01 substrings as 10 substrings.



## Regular Expressions Basics Summary Table

Basics	Explanation		Options	Explanation
/pattern/	Reg exp delimiters		i	CASELESS
.	Wild card – match any non newline (sans DOTALL)		s	DOTALL - . matches all
	Or (alternation)		m	MULTILINE - ^ matches start of each line; \$ matches end of each line
\	Quotes meta characters		/.../im	Example usage
Char types	Explanation		Char classes	Explanation
\w (\W)	Alphanumeric char or _		[aeiou]	Character class
\s (\S)	Whitespace char		[0-9A-Z_]	Ranges
\d (\D)	Digit (ie. 0-9)		[^0-9]	Negation
Quantifiers	Explanation		Assertions	Explanation
?	0 or 1 of preceding item		^	Start of line (with MULTILINE)
*	0 or more of preceding item		\$	End of line (with MULTILINE)
+	1 or more of preceding item		\b	Word boundary
?	*? +? and ?? are not greedy		\A	Start of string
{4} / {4,}	Exactly 4 / At least 4		\Z	End of string
{3,5}	3, 4, or 5 of preceding item			