

Lab 11: Data and RegEx

A Data Formats

Characters can be represented in a number of ways. Complete Table 1 using (the first example has already been completed):

<http://asecuritysite.com/Coding/ascii>

<http://asecuritysite.com/Coding/asc>

Table 1: Hex, decimal and binary

Character	ASCII hex	ASCII decimal	ASCII binary
'a'	0x61	97	01100001
'Z'			
	0x42		
			01001000
	0x4A		
			01011110
'~'			

B Base-64 and Hex

Two typical formats for representing binary data is hexadecimal and Base-64. Complete Table 2 using:

<http://asecuritysite.com/Coding/ascii>

Table 2: Text, hex and Base-64

Text	Hex	Base-64
"hello"	0x48656C6C6F	SGVsbG8=
		R29vZGJ5ZQ==
	0x 436973636F	
"Elephant"		

C Regular Expressions

Using the text below and the following Web site:

<http://regex101.com/>

There is not much we can do apart from contacting f.smith@home.net to see if he would like to reboot the server at 192.168.0.1. If he can do this then I will call him on 444.3212.5431. My credit card details are 4321-4444-5412-2310 and 5430-5411-4333-5123 and my name on the card is Fred Smith. I really like the name domain fred@home. Overall our target areas are SW1 7AF and EH105DT. I tested the server last night, and I think the IP address is 10.0.0.1 and there are two MAC addresses which are 01:23:45:67:89:ab or it might be 00.11.22.33.44.55.

The book we will use is "At Home" and it can be bought on amazon.com or google.com, if you search for 978-1-4302-1998-9. My password is:

a1b2c3

Best regards,

Bert.
EH14 1DJ
+44 (960) 000 00 00
1/1/2009

Now see if you can detect:

- (i) Email addresses.
- (ii) IP addresses.
- (iii) Telephone numbers (US style).
- (iv) UK post codes.
- (v) Credit card details.

Now repeat, but this time using the /g modifier on the rule. What effect does it have?

Examples:

Email address: [a-zA-Z0-9._%+-]+@[a-zA-Z0-9._%+-]
IP: [0-9]{1,3}\.[0-9]{1,3}\.[0-9]{1,3}\.[0-9]{1,3}
Telephone: \d{3}[-.]?\d{3}[-.]?\d{4}
Post code: [A-Z]{1,2}[0-9]{1,2}[A-Z]?\s[0-9][A-Z][A-Z]\$
Credit card (Visa): 4\d{3}(\s|-)?\d{4}(\s|-)?\d{4}(\s|-)?\d{4}

The **fred@home** string should trigger the email address regular expression. Now modify it so that it only picks up the valid email address (search on the Internet, if required, for an improved regex for this. What is the updated regex:

Only one credit card is picked-up. Why?

Now modify the regex for credit card detection using the format [345] ... so that it will detect Visa, MasterCard and American Express, with a single regular expression. What is the regex:

What is the difference between:

```
[A-Z]{1,2}[0-9]{1,2}[A-Z]?\s[0-9][A-Z][A-Z]$
```

And

```
[A-Z]{1,2}[0-9]{1,2}[A-Z]?\s[0-9][A-Z][A-Z]
```

Can you generate some examples of where these rules trigger? What modifications did you make to make the first one trigger:

There is a second post code in the text (EH105DT). Can you modify the regex to find this one? What is the update:

Find a UK Post Code matcher from the Internet, and test it so that it only takes valid UK postcodes (and ignores one in the format *ZZ99 ZZZ*).

There are two MAC addresses in the text. Write a regular expression to detect it. What is the regex:

There are two domain names in the text. Write a regular expression to pick these off. What is it:

The following detects a password contains at least one letter, at least one number, and be longer than six characters.

```
(?=.*[0-9]+.*)(?=.*[a-zA-Z]+.*)[0-9a-zA-Z]{6,}
```

Implement it and see if it works. What did it detect?

There is an ISBN number in the text. Write a regular expression to pick this off. What is it:

There is a UK phone number in the text. Write a regular expression to pick this off. What is it:

Write a rule which detects a word with all upper case. What is it:

Write a rule which detects numbers spelt at works (eg “one”, “two” and so on). What is it:

Write a rule which picks off the date. What is it:

D Possible Answers

Email address:

[a-zA-Z0-9._%+-]+@[a-zA-Z0-9._%+-]

([a-zA-Z0-9_-\.]+)@((\[[0-9]{1,3}\.[0-9]{1,3}\.[0-9]{1,3}\.))|((([a-zA-Z0-9_-]+\.)))([a-zA-Z]{2,4}|[0-9]{1,3})(\?)

IP:

[0-9]{1,3}\.[0-9]{1,3}\.[0-9]{1,3}\.[0-9]{1,3}

Telephone:

\d{3}[-.]?\d{3}[-.]?\d{4}

UK Post code:

[A-Z]{1,2}[0-9]{1,2}[A-Z]?\s[0-9][A-Z][A-Z]

[A-Z]{1,2}[0-9]{1,2}[A-Z]?\s?[0-9][A-Z][A-Z]

?((([BEGLMNSWbeglmnsw][0-9][0-9]?)|((([A-PR-UWYZa-pr-uwyz][A-HK-Ya-hk-y][0-9][0-9]?)|((([ENWenw][0-9][A-HJKSTUwa-hjkstuw])|([ENWenw][A-HK-Ya-hk-y][0-9][ABEHMNPRVWXYabehmnprvwxy])))?)?[0-9][ABD-HJLNP-UW-zabd-hjlnp-uw-z]{2}

Credit card (Visa):

4\d{3}(\s|-)?\d{4}(\s|-)?\d{4}(\s|-)?\d{4}
[45]\d{3}(\s|-)?\d{4}(\s|-)?\d{4}(\s|-)?\d{4}

Domain name:

[a-zA-Z\.] + \. (com|net|uk)

[a-zA-Z0-9_-\.] + \. (com|org|net|mi|edu|COM|ORG|NET|MIL|EDU|UK)

MAC

([0-9a-fA-F][0-9a-fA-F]:){5}([0-9a-fA-F][0-9a-fA-F])

([0-9a-fA-F][0-9a-fA-F]::){5}([0-9a-fA-F][0-9a-fA-F])

Password:

(?=.*[0-9]+.*)(?=.*[a-zA-Z]+.*) [0-9a-zA-Z]{6,}