

Pràctica 9: Expressions Regulars (REGEX)

Pràctica realitzada per :

Sergio prieto y Osama namri

Dam1B

Lliuraments

Els resultats d'aquesta part de la pràctica s'hauran d'entregar en format PDF i l'entrega pot ser a través de GIT* o el moodle.

* S'ha d'entregar l'enllaç del GIT al moodle.

Exercici 1: Analitza documents XML

Clona el repositori <https://github.com/pauitic/practica9>

Escriu les expressions regulars que seleccionin els continguts que s'indiquen del fitxer **xml_for_regex.xml**. Per cada exercici, trobaràs una captura de pantalla que especifica la manera que s'ha de fer la captura de caràcters.

1. Selecciona les etiquetes **<price>** i el seu contingut.
respuesta 1: `<price>.7.95</price>`

```
<name>Belgian Waffles</name>
<price>$5.95</price>
<description>
Two of our famous Belgian Waffl
</description>
```

2. Selecciona els preus sense l'etiqueta **<price>**
respuesta : `1 [$]...`

```
<food>
  <name>Berry-Berry Belgian Waffles</name>
  <price>$8.95</price>
  <description>
  Belgian waffles covered with assorted fres
  </description>
  <calories>900</calories>
```

3. Selecciona les etiquetes **<description>** i el seu contingut. Compte que ara poden haver-hi salts de línia!

Respuesta: `<description>\n.*\n.*>$`

```
<food>
  <name>Belgian Waffles</name>
  <price>$5.95</price>
  <description>
Two of our famous Belgian Waffles with plenty of real maple syrup
  </description>
  <calories>650</calories>
</food>
```

4. Selecciona totes (i només) les **etiquetes de tancament**.

Respuesta : `</.*>`

```
<food>
  <name>Belgian Waffles</name>
  <price>$5.95</price>
  <description>
Two of our famous Belgian Waffles wi
  </description>
  <calories>650</calories>
</food>
```

5. Selecciona totes (i només) les **etiquetes d'obertura**.

Respuestas

`<([^\s/]+)>`

`<([^\s/*])>`

```
<food>
  <name>Strawberry Belgian Waffles</name>
  <price>$7.95</price>
  <description>
Light Belgian waffles covered with strawberries and whipped cream
  </description>
  <calories>900</calories>
</food>
```


Exercici 2: Analitza documents JSON

Desenvolupa una expressió regular específica per capturar les cadenes de caràcters indicades en el fitxer `json_for_regex.json`. L'expressió regular que utilitzis ha de servir per capturar els *strings* d'aquest document, i no ha de ser genèrica en cap cas.

6. Selecciona totes les **keys** del document JSON juntament amb els dos puntets.

Respuesta : `".*:"`

```
{
  "nombre" : "Draculina",
  "especie" : "Vampiro",
  "habilidades" : ["Transformacion en murcielago", "Control mental"],
  "nivel_peligrosidad" : 8,
  "region" : "Transilvania",
  "es_volador" : true
}
```

7. Selecciona tots¹ els **valors** (*values*) JSON. Pots utilitzar com a referència els dos punts anteriors i la coma, com es mostra a la imatge.

respuesta:

Respuestas : `.*,?`

```
{
  "nombre" : "Draculina",
  "especie" : "Vampiro",
  "habilidades" : ["Transformacion en murcielago", "Control mental"],
  "nivel_peligrosidad" : 8,
  "region" : "Transilvania",
  "es_volador" : true
},
```

8. Selecciona les **llistes** de *strings* del document.

Respuesta final `[[\]].*[^,]`

```
"habilidades" : ["Ilusiones enganosas", "Manipulacion de luz", "Confusion de \"regex\""],
```

9. Selecciona els **booleans**. Compte no seleccionar els strings "true" i "false" dins de *strings*.

```
{
  "nombre" : "Fuego Fatuo",
  "especie" : "Espiritu",
  "habilidades" : ["Ilusiones enganosas", "Manipulacion de luz y booleanos false", "Confusion de \"regex\""],
  "nivel_peligrosidad" : 5,
  "region" : "Pantano Encantado",
  "es_volador" : false
}
```

Resputas: `(false|true)$`

¹ Excepte el valor de la clau "monstruos"

10. Selecciona els **strings**, però no les **keys** (si t'ajuda, pots seleccionar les comes i els] tal com es mostra a la imatge)
- ```
(["^"]".*)(["^d"]),$ porahora == > (["^?":]".*)(["^d"]),$
```
- Intento numero  $10^{2^{998}}$  = **("(.+?)"**)

```
{
 "nombre": "Fuego Fatuo",
 "especie": "Espíritu",
 "habilidades": ["Ilusiones enganosas", "Manipulación de luz y booleanos false", "Confusion de \"regex\""],
 "nivel_peligrosidad": 5,
 "region": "Pantano Encantado",
 "es_volador": false
},
```

## Exercici 3: Troba les paraules

A partir de les següents expressions regulars, identifica **tres paraules** que puguin ser capturades per a cada una d'elles. A més, especifica el **tipus de dades** conegudes a les quals podrien referir-se les diferents expressions:

a.

```
[A-Z] [A-Z] \d\d (\d{4}) {5}
```

**Matricules**

JK22 4444 4444 4444 4444 4444  
LO32 4538 3492 4785 3899 0031  
XX00 1111 2222 3333 4444 5555

b.

```
[1-2]? \d\d (\. [1-2]? \d? \d) {3}
```

**IP**

123.123.0.9  
123.123.1.22  
192.299.199

c.

```
\d\d [-/] ([012] \d) [-/] \d\d \d\d
```

**Fechas**

10-10-2000  
10/10/2000

d.

```
[0123] \d [-/] (([012] \d) | [a-z] {3}) [-/] \d\d \d\d
```

**Fecha alfanumerica**

25-03-2024  
25-mar-2024  
01-may-2024 dia del trabajador  
39-iii-0000- does not exist but is valid

e.

`\w*\.(jpg|png|pdf)`

[Documentos](#)

[estasaprobado.pdf](#)

[ponmeundiez.jpg](#)

[10.png](#)

## Telèfons

### Respuesta

`^(?:\+34\s)?(?:6|7|8|9)(?:[\s-]?[0-9]){8}$` funciona

Escriu una expressió regex que validi els telèfons espanyols. Tingues en compte que:

- Pot o no començar amb +34
- El número està format per 9 dígits
- El número comença per 6 o 7 si és mòbil i 8 o 9 si és fix
- Els dígits poden estar seguits o separats per un guionet o espai

| Casos vàlids                                                               | Casos invàlids                                                                 |
|----------------------------------------------------------------------------|--------------------------------------------------------------------------------|
| 645540844<br>64 554 08 44<br>74-554-08-44<br>+34 645540844<br>+34945540844 | +34445540844<br>64554084<br>+346+45540844<br>+34-6455--40844<br>+34 6455 40844 |

## DNI / NIE

Escriu una expressió *regex* pels DNIs i NIE

- Els DNI tenen **8 números** i un **dígit de control** alfabètic
- Els NIE comencen per **X, Y o Z**, tenen **7 nombres** i un dígit de **control** alfabètic

**Respuesta:** `^(?:\d{8}[A-Z]|^[XYZ]\d{7}[A-Z])$` funciona

| Casos vàlids                                     | Casos invàlids                                               |
|--------------------------------------------------|--------------------------------------------------------------|
| 77958643G<br>00000000X<br>X7958643A<br>Y9999999E | 77958643<br>C7958643Q<br>X7958643<br>XX958643F<br>Z77958643D |

# Correus electrònics

Escriu una expressió regex que validi els emails seguint les següents condicions:

- La paraula que precedeix l'arrova "@" pot tenir lletres no accentuades, números, guions, punts i barra baixes
- El domini de la direcció pot tenir lletres, punts i guions

Resposta: `^[a-zA-Z0-9_.-]+@[a-zA-Z.-]+\.[a-zA-Z]{2,}$` funciona

| Casos vàlids                                                                                               | Casos invàlids                                                              |
|------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| user2@iticbcn.cat<br>name.surname@iticbcn.cat<br>name_underscore@iticbcn.cat<br>NAME-surname@it-ic.bcn.cat | name.surname@<br>@iticbcn.cat<br>çç@iticbcn.cat<br>name surname@iticbcn.cat |

# Dominis d'URLs

Escriu una expressió regex que validi els dominis dels URL tenint en compte les següents condicions

- L'URL comença per "http://" o "https://"
- El domini pot tenir lletres, guions, punts
- Pot acabar amb barra

Resposta

`^(https?://)(www\.)?[a-zA-Z]*-[a-zA-Z]*(\.)?(cat|fr)?$`

| Casos vàlids                                                                           | Casos invàlids                                                                                          |
|----------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|
| https://www.educaciodigital.cat/<br>https://educacio-digital.fr<br>http://www.educacio | educacio-digital.es<br>http://educacio-digital.cat/hoola/404<br>http://educacio.digital.cat/nomesDomini |

# URLs completes

Escriu una expressió regex que validi els URL tenint en compte les següents condicions

- L'URL comença per "http://" o "https://"
- El domini pot tenir lletres, guions
  - El domini no pot tenir subdomini
  - El domini ha de pertànyer a .es, .cat, .org o .edu
- La ruta pot tenir lletres i números, guions, punts i barra baixes
  - A més, es poden incloure paràmetres, i per això s'han de permetre els símbols ? % & i =
- Pot acabar amb barra

RESPOSTA

`^https?://([a-zA-Z-])+[-]?[a-zA-Z]*(\.)?(cat/)(.*)*$`



| Casos vàlids                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <a href="https://educaciodigital.cat/">https://educaciodigital.cat/</a><br><a href="http://educacio-digital.cat/apt1/apt3">http://educacio-digital.cat/apt1/apt3</a><br><a href="http://educacio-digital.cat/sim.bo-l_s/me?s?param=1&amp;param2=2">http://educacio-digital.cat/sim.bo-l_s/me?s?param=1&amp;param2=2</a>                                                                                                                                                                                                                       |
| Casos invàlids                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| <a href="http://educacio-digital.cat//DOBLE">http://educacio-digital.cat//DOBLE</a><br><a href="http://educacio.digital.cat/te_subdomini">http://educacio.digital.cat/te_subdomini</a><br><a href="http://educacio_digital.cat/te_barrabaixa_al_domini">http://educacio_digital.cat/te_barrabaixa_al_domini</a><br><a href="https://educacio-digital.fr/fr_no_permes">https://educacio-digital.fr/fr_no_permes</a><br><a href="http://educacio-digital.es">educacio-digital.es</a><br><a href="https://www.educacio">https://www.educacio</a> |

## Adreces

Escriu una expressió regex que validi les adreces que segueixin les següents condicions.

- **Comença** per: C/ Av. Pg. Rb
- Segueix del **nom del carrer** que pot ser una o diverses paraules amb lletres majúscules i minúscules accentuades
- Continua amb el **número de porta** que pot tenir diversos dígits
- Pot tenir **número de pis** i **número de porta**
- Continua amb el **nom de la ciutat**, que pot estar formada per diferents paraules
- Acaba amb la **província** entre parèntesis. Només pot ser Barcelona, Girona, Tarragona o Lleida.

Respuesta

```
^(CV|Av|.Pg|.Rb|.)[a-zA-ZÀ-ÿ\s]+\d+(\s\d\s\d)?[a-zA-ZÀ-ÿ\s]+\((Barcelona|Girona|Tarragona|Lleida)\)$
```

| Casos vàlids                                                                                                                                                                                                                                                                           |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| C/ Diputació 31 1 2 Badalona (Barcelona)<br>Av. Girona 42 1 2 Badalona (Barcelona)<br>Av. Rossello 35 Arbucies (Girona)<br>Rb. Les Rambles 4432 Lleida (Lleida)<br>Av. Gran via de les corts catalanes 32 Santa Coloma de Gramanet (Barcelona)<br>Av. Rosselló 32 1 2 Reus (Tarragona) |
| Casos invàlids                                                                                                                                                                                                                                                                         |
| Av. Gran via de les corts catalanes 32 (Barcelona)<br>Gran via de les corts catalanes 32 Badalona (Barcelona)<br>C/ 32 1 2 Badalona (Barcelona)<br>Av. Rosselló 32 1 2 4 Salt (Girona)                                                                                                 |

# Contrasenyes fortes

Dissenya una expressió regex que validi les contrasenyes fortes.

- Com a mínim ha de tenir una lletra **majúscula** i una **minúscula**
- Com a mínim ha de tenir **dos díigits**
- Com a mínim ha d'incloure un dels següents **símbols**: . \_ ? \ [ ] ( )
- La contrasenya ha de tenir entre **8 i 30 caràcters**

Respuesta

`^(?=.*[a-z])(?=.*[A-Z])(?=.*\d.*\d.*)(?=.*[._?\\[\]()]).{8,30}$`

| Casos vàlids                                      | Casos invàlids                      |
|---------------------------------------------------|-------------------------------------|
| 12345678aA._?<br>aA._?12345678<br>aA\[]()12345678 | 123456789<br>aA77._<br>77fghgfAAAAA |