

# Pràctica 9: Expressions Regulars (REGEX)

Noms:

Pablo Fernandez Huaman

Jan Perales Freniche

## 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.

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

`<price>\$\d+\.\d+</price>`

2. Selecciona els preus sense l'etiqueta `<price>`

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

`\$\d+\.\d*`

(?<=>\\$) \d\*\.\d+

Proposta més complicada, però sintàcticament interessant ja que retorna tot els nombres que es trobi davant de ">\$", per tant només buscaria preus amb el símbol i no totes les etiquetes amb dígits decimals com a la primera proposta.

Aquesta sintaxi s'anomena lookahead positiu i l'hem après a:

<https://regexlearn.com/es/learn/regex101>

(pàgina 46/56)

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

```
<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>
```

<description>\s\*.\*\s\*</description>

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

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

</[a-z]+>

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

```
<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>
```

<[^/]\.w\*>

## 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.

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

`".\w*":`

7. Selecciona tots<sup>1</sup> els **valors** (*values*) JSON. Pots utilitzar com a referència els dos punts anteriors i la coma, com es mostra a la imatge.

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

`:\s["\[{}?\..*["\]]?,`

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

```
"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
}
```

`(true|false)$`

---

<sup>1</sup> 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)

```
{
  "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
},
```

(?<=: \[) ".\*"j

## 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}`

AA78 7777 7777 7777 7777

AO23 2535 5646 6473 5364 7573

ZY24 5356 3553 2424 2332 4322

Tipus de dada: IBAN de compte bancari

b.

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

10.0.0.0

192.168.19.0

172.10.0.0

Tipus de dada: rang IPv4

c.

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

30-01-3333

30/12/2012

26-09-1997

Tipus de dada: camp de data)

d.

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

26-nov-1997

23-11-1997

08/oct/1968

Tipus de dada: camp de data pero que també captura formats que tinguin el camp de mes en un string

e.

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

`holamon.jpg`

`coco_trisky.png`

`m4PracticaRegex.pdf`

Tipus de dada: nom d'arxiu i la seva extensió

## Telèfons

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 dígit
- El número comença per 6 o 7 si és mòbil i 8 o 9 si és fix
- Els dígit poden estar seguits o separats per un guionet o espai

Regex

`(\+34)?([\s]?[67]|[\s]?[89])(\d[-\s]?){8}`

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

Regex

`[X-Z]\d{7}[A-Z]\d{8}[A-Z]`

Casos vàlids	Casos invàlids
77958643G 00000000X X7958643A Y9999999E	77958643 C7958643Q X7958643 XX958643F 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

Regex
<code>[\\w\\d\\._-]+@[a-z\\._-]+</code>

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

Regex
<code>https?://[a-z\\._-]+/?</code>

Casos vàlids	Casos invàlids
https://www.educaciodigital.cat/ https://educacio-digital.fr	educacio-digital.es http://educacio-digital.cat/hoola/404

https://www.educacio	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, punts
  - 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 i barra baixes
  - A més, es poden incloure paràmetres, i per això s'han de permetre els símbols ? % & i =
- Pot acabar amb barra

### Regex

[https?://\[a-z-\]+\(cat|es|org|edu\)/\(\[a-z\d\.\\_\?%&=\]/?\)\\*](#)

### Casos vàlids

https://educaciodigital.cat/  
http://educacio-digital.cat/apt1/apt3  
http://educacio-digital.cat/sim.bo-l\_s/me?s?param=1&param2=2

### Casos invàlids

http://educacio-digital.cat//DOBLE  
http://educacio.digital.cat/te\_subdomini  
http://educacio\_digital.cat/te\_barrabaixa\_al\_domini  
https://educacio-digital.fr/fr\_no\_permes  
educacio-digital.es  
https://www.educacio

## 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 dígitos
- 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.



### Regex

`(C/ |Av. |Pg. |Rb. )((\p{L})+\s)+\d{1,} (\d \d)?((\p{L})+\s)+\(((Barcelona|Girona|Tarragona|Lleida))\)`

### Casos vàlids

C/ Diputació 31 1 2 Badalona (Barcelona)  
Av. Girona 42 1 2 Badalona (Barcelona)  
Av. Rossello 35 Arbucies (Girona)  
Rb. Les Rambles 4432 Lleida (Lleida)  
Av. Gran via de les corts catalanes 32 Santa Coloma de Gramanet (Barcelona)

### Casos invàlids

Av. Gran via de les corts catalanes 32 (Barcelona)  
Gran via de les corts catalanes 32 Badalona (Barcelona)  
C/ 32 1 2 Badalona (Barcelona)  
Av. Rosselló 32 1 2 4 Salt (Girona)  
**Av. Rosselló 32 1 2 Reus (Tarragona)**

## 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 dígit**s
- Com a mínim ha d'incloure un dels següents **símbols**: . \_ ? \ [ ] ( )
- La contrasenya ha de tenir entre **8 i 30 caràcters**

### Regex

`(\d*)(\w*)(\.\|\|\.\|.*\|\d\|?)`

### Casos vàlids

12345678aA.\_?  
aA.\_?12345678  
aA\[\]()12345678

### Casos invàlids

123456789  
aA77.\_  
77fghgfAAAAA