**Pràctica 9**: Expressions Regulars

(REGEX)

# 

# 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.

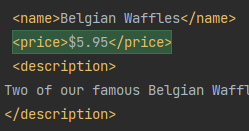
**\* S’ha d’entregar l’enllaç del GIT al moodle.**

# Exercici 1: Analitza documents XML

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

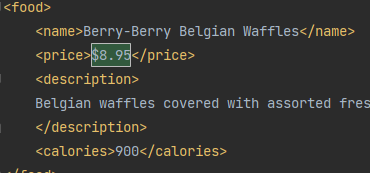
Escriu les expressions regulars que seleccionin els continguts que s’indiquen del fitxer **xml\_for\_regex.xml.** Per cada exercici, trobaràs una captura de pantalla que específica la manera que s’ha de fer la captura de caràcters.

1. Selecciona **les etiquetes <*price*> i el seu contingut.**



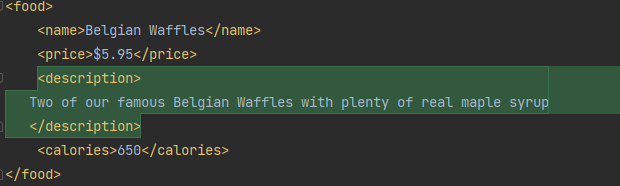
<price>.\*</price>

1. Selecciona **els preus sense l’etiqueta** <*price*>



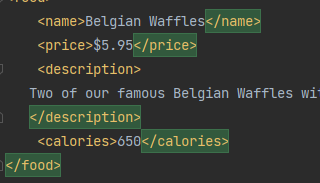
\$\d\*\.?\d+

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



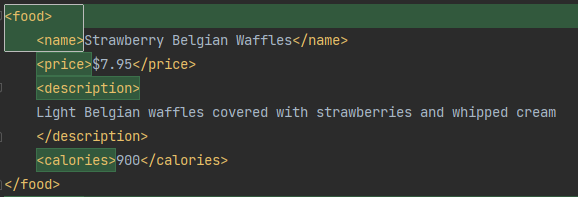
<description>[a-zA-Z\n ]\*</description>

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



</.+>

1. Selecciona totes (i només) les **etiquetes d’obertura.**



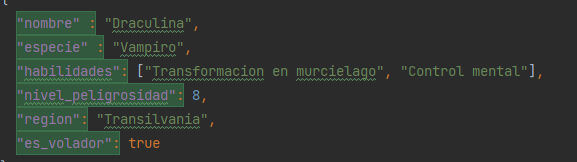
<[^/]\*>

# 

# Exercici 2: Analitza documents JSON

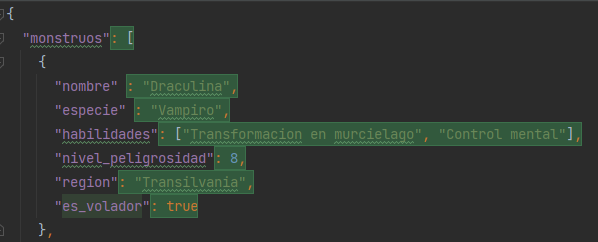
Desenvolupa una expressió regular específica per capturar les cadenes de caràcters indicades en el fitxer **json\_for\_regex.json**. L'expressió regular que utilitzis ha de servir per capturar els *strings* d’aquest document, i no ha de ser genèrica en cap cas.

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



".\*" ?:

1. Selecciona tots[[1]](#footnote-0) els **valors** (*values*) JSON. Pots utilitzar com a referència els dos punts anteriors i la coma, com es mostra a la imatge.



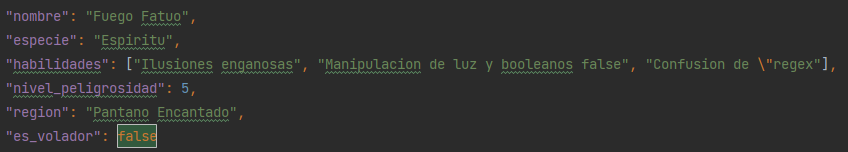
:.\*

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



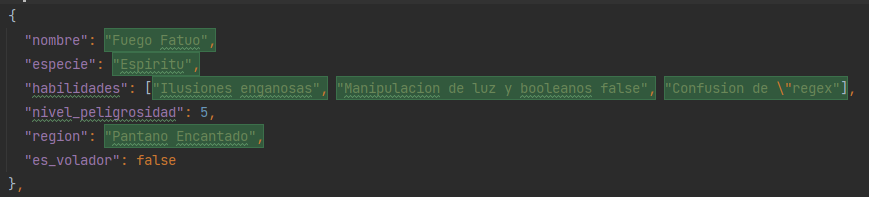
\[.\*\]

1. Selecciona els **booleans**. Compte no seleccionar els strings “true” i “false” dins de *strings.*



(true|false)$

1. Selecciona els ***strings,*** *però* no les *keys* (si t’ajuda, pots seleccionar les comes i els ] tal com es mostra a la imatge)



"([a-zA-Z ]\*|\\")\*"[,\]]

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

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

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

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

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

# 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 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  64 554 08 44  74-554-08-44  +34 645540844  +34945540844 | +34445540844  64554084  +346+45540844  +34-6455--40844  +34 6455 40844 |

**(\+34 ?)?[6789](\d ?-?){8}**

# 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

| **Casos vàlids** | **Casos invàlids** |
| --- | --- |
| 77958643G  00000000X  X7958643A  Y9999999E | 77958643  C7958643Q  X7958643  XX958643F  Z77958643D |

**^[0-9]{8}[A-Z]$|^[XYZ][0-9]{7}[A-Z]$**

# 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

| **Casos vàlids** | **Casos invàlids** |
| --- | --- |
| user2@iticbcn.cat  name.surname@iticbcn.cat  name\_surname@iticbcn.cat  NAME-surname@it-ic.bcn.cat | name.surname@  @iticbcn.cat  çç@iticbcn.cat  name surname@iticbcn.cat |

**^[0-9a-zA-Z-.\_]+@[a-zA-Z-.]+$**

# 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

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

**https?://([0-9a-zA-Z\-\.]+)\*/?$**

# 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

| **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 |

**https?://[a-z\-]+\.(cat|es|edu|org)(/[0-9a-zA-Z\-\_\.\%\?=&]+)\*/?$**

# 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.

| **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) |

# **(C/|Av\.|Pg\.|Rb\.) [a-zA-Z àáèéíòóúÀÁÈÉÍÒÓÚüïÜÏçÇñÑ]+ \d+ ?\d? ?\d?[a-zA-Z ]+ \((Barcelona|Girona|Tarragona|Lleida)\)**

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

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

**^(?=.\*[0-9].\*[0-9])(?=.\*[a-z])(?=.\*[A-Z])(?=.\*[.\_?\\\[\]()]).{8,30}$**

1. Excepte el de la clau “monstruos” [↑](#footnote-ref-0)