ITICBCN
Institut TIC de Barcelona
Consorci d'Educació de Barcelona
Generalitat de Catalunya
Ajuntament de Barcelona

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.**

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

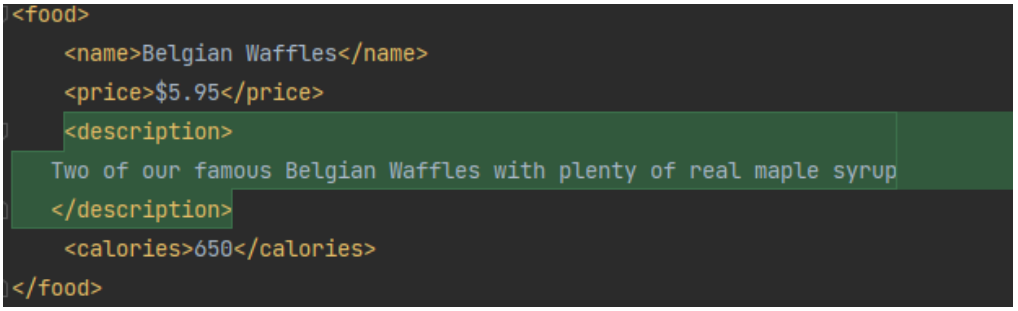
<price>.+</price>

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>
```

[$].+[^</price>]

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

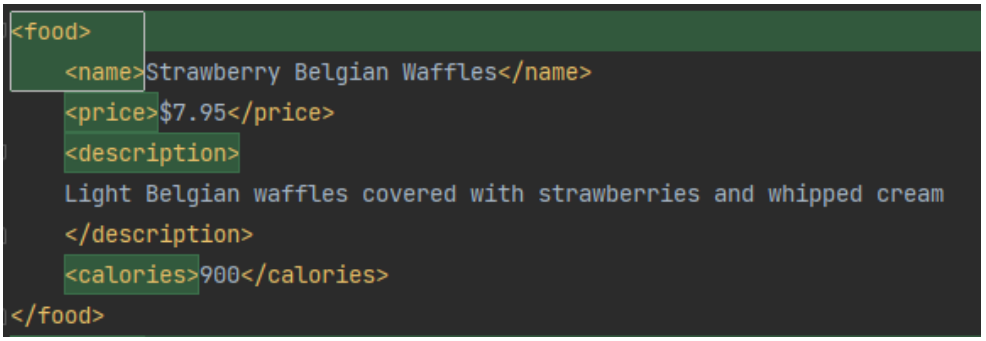


<description>\n.+\n.+

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

</.+

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



<\w*>

Exercici 2: Analitza documents JSON

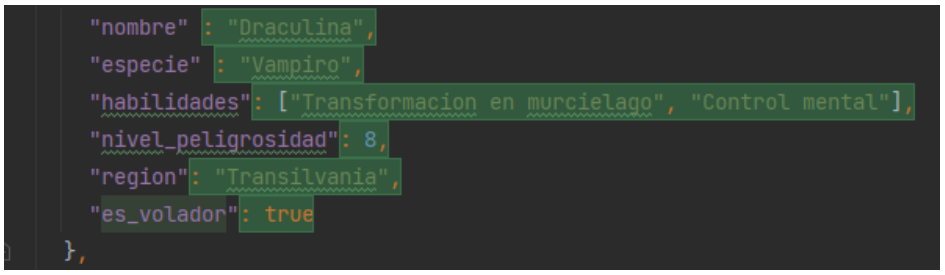
Desenvolupa una expressió regular específica per capturar les cadenes de caràcters indicades en el fitxer **json_for_regex.json**. L'expressió regular que utilitzis ha de servir per capturar els *strings* d'aquest document, i no ha de ser genèrica en cap cas.

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
```

(".*") ?:

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

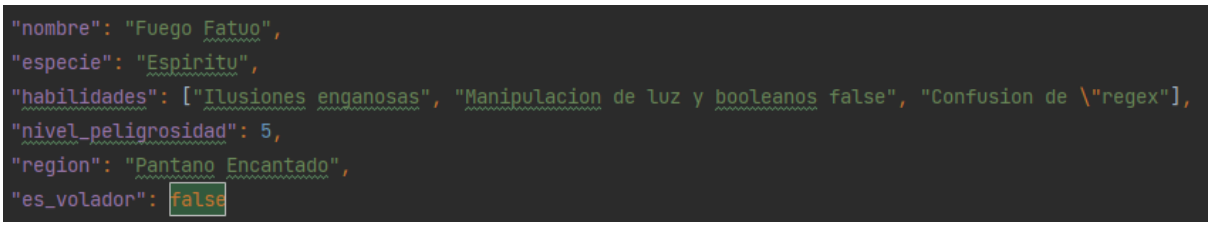


[:].*

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

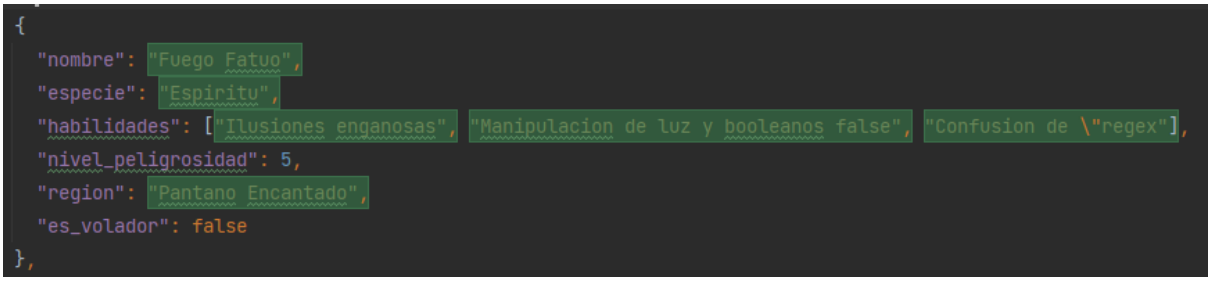
[[].*[^,]

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



(true)|(false)

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



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}
AB45: Podria referir-se a un codi d'identificació únic, com un número de sèrie d'un producte electrònic o una clau d'accés a un sistema.
0000: Podria representar un número de compte bancari o un codi d'identificació d'un objecte, com un número de referència per a una comanda.
0123: pot referir-se a un altre tipus de codi, com ara un número de seguiment de paquet per a enviaments logístics o un identificador de client per a una empresa.

b. [1-2]?\d\d(\.[1-2]?\d?\d){3}
152.62.84.73: Podria referir-se a una adreça IP, que consta de quatre octets, cada un dels quals pot ser un nombre de 1 a 255.
90.67.32.7: També podria representar una adreça IP, tot i que en aquest cas, hi ha menys dígits en els blocs posteriors.
180.230.7.157: pot ser una adreça IP vàlida, amb nombres que varien entre 1 i 255 en cada octet.

c. \d\d[-/]([012]\d)[-/]\d\d\d\d
20/10/2024: Podria referir-se a una data en format "mm/dd/yyyy" (mes/dia/any), on el mes és 10 (octubre), el dia està en el rang de 00 a 29 (20 és vàlid) i l'any és 2024.
28-05-1999: Podria representar una altra data en format "dd-mm-yyyy" (dia-mes-any), on el dia és 28, el mes és 05 (maig) i l'any és 1999.
77/29/7777: pot ser un altre exemple de data.

d. [0123]\d[-/](([012]\d)|[a-z]{3})[-/]\d\d\d\d
30-10-2023: Podria referir-se a una data en format "dd-mm-yyyy" (dia-mes-any), on el dia és 30, el mes és 10 (octubre) i l'any és 2023.
31/jan/2004: Podria representar una data en un format alternatiu, on la primera part pot ser un nombre entre 00 i 29 o tres lletres minúscules (en aquest cas, "Jan" per a gener), seguit de l'any 2004.
28/03/1999: pot ser un altre exemple de data, però en aquest cas, la part del dia (30) no seria vàlida per a febrer ja que està fora del rang vàlid.

e. \w*\.(jpg|png|pdf)
documento.pdf: Podria representar un document en format PDF.
imagen.png: podria referir-se a una altra imatge, aquesta vegada amb extensió "png".
imagen.jpg: Podria referir-se a un fitxer d'imatge amb extensió "jpg".

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)?[(-)]?((6|7)(\d[-]?){8}|(8|9)(\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@iticbcn.cat NAME-surname@it-ic.bcn.cat	çç@iticbcn.cat name surname@iticbcn.cat

[a-zA-Z0-9_.\-]+@[a-zA-Z\.\-]+\.[a-zA-Z]{2,}

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

(http|https):\/\/(www\.)?([a-zA-Z0-9\-.]+)(\/)?$

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?://)([^./]+)\.(es|cat|org|edu)(/[a-zA-Z0-9_-]+)*(\?[\w%&=]*)?/?$

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) Av. Rosselló 32 1 2 Reus (Tarragona)
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\.|C\/|Rb\.|Pg\.)\s+([A-Za-zçñÀ-Üà-ü\s]+)[\d\s]+([A-Za-zçñÀ-Üà-ü\s]+)[(](Barcelona|Girona|Lleida|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 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

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