

Entrega: 8/11/2023

Noms: Mariona Farré Tapias Marc Pérez Guerrero

Activitat 3: Rein

Objectiu del rastrejador:

El nostre rastrejador l'hem anomenat *Vogue Inspector*, és un rastrejador que té com a objectiu rastrejar dins la web de <https://www.vogue.es/> en específic dins d'un article de moda de les tendències d'aquesta tardor, en aquest url:

https://www.vogue.es/galerias/tendencias-mas-arriesgadas-otono-invierno-2023-2024?utm_source=pocket-newtab-es-es

Hem configurat el rastrejador perquè recuperi diferents dades dels articles per tal de, que per exemple, poder consultar quines són les notícies actuals, quines són les novetats o el que impressiona del camp de la moda actualment, etc.

Recuperem els següents camps de l'url principal:

- Títol: Correspon als títols dels diferents articles.
- Article: Correspon a l'article en si, la informació de la notícia (el text explicatiu).
- Url: Amb aquest camp volem obtenir l'enllaç que tenen alguns dels articles que fan referències a altres notícies o explicacions més a fons (contextuals).

Con sujetador a la vista

A medio camino entre el *look* rapero y la tendencia del momento, encontramos la ropa interior a la vista. Gracias al avance del feminismo, convertir las bragas o el sujetador en protagonistas es ya una decisión estilística que muchas amamos. Y Julia Fox también.

Aquí podem veure el títol en gran, el text del article i un exemple d'url amb el text subratllat Julia Fox.

Després de les url dins dels articles obtenim dos camps més:

- Títol: Títol de la informació extra
- Description:

Consisteix en tota la informació sobre aquest títol

Sandalias con pelo: el calzado más excéntrico de 2023 va a estar por todas partes

Ni frío ni calor. No son medias tintas, ni una propuesta que deje indiferente: es la paradoja del calzado que promete ser el más visto de 2023. Con una suela de pelo, las sandalias tipo pelluche ya se han colado en algunas de las colecciones de la temporada. Vayamos juntas por un lado: Gucci y Burberry son las principales autoras de este look. Pamela que incluye bolos y sandalias con pelaje. Que quedan bien en un éxito de ventas no lo dudamos: ahí tenemos los calcetines de pelo con los que debutó Alessandro Michele al frente de la casa italiana, que convirtió en un best seller de la firma. En Londres, Burberry fue otra de las marcas que ha incorporado este calzado en su desfile. En su respectivo desfile, Daniel Lee, que tiene bien calada la debilidad de la moda por los diseños exagerados, incluyó bolos y sandalias con colas de pelo falsas.



Gucci otoño invierno 2023

Burberry otoño invierno 2023

Podem recuperar també les imatges fàcilment, però no ho hem considerat útil pel funcionament del nostre rastrejador, ja que no es poden fer consultes a una imatge.

Creació del rastrejador:

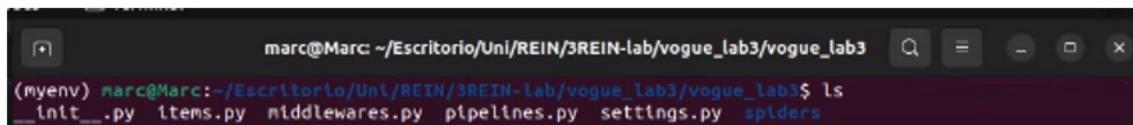
Per començar un rastrejador web, farem servir Scrapy, una biblioteca de Python per poder fer rastreig de qualsevol web indicat.

Per generar l'estructura del rastrejador web, executarem:

```
REIN/LAB3$ scrapy startproject vogue_lab3 vogue_lab3
New Scrapy project 'vogue_lab3', using template directory '/usr/local/lib/python3.8/dist-packages/scrapy/templates/project
', created in:
  /dades/[REIN/LAB3/vogue_lab3

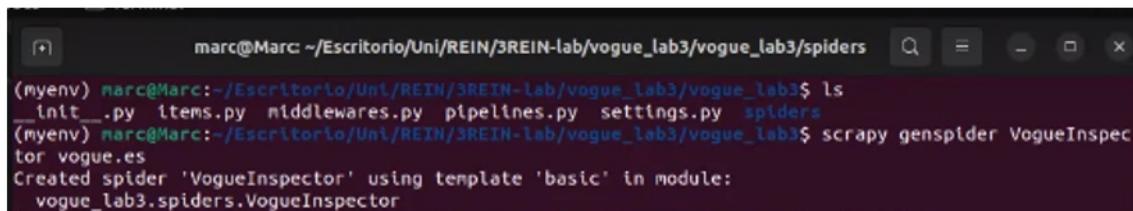
You can start your first spider with:
  cd vogue_lab3
  scrapy genspider example example.com
```

Després podem veure que s'ha creat correctament el projecte fent un `-ls` i veient quins fitxers s'han creat, si surt com ensenyem a continuació s'ha creat correctament:



```
marc@Marc:~/Escritorio/Uni/REIN/3REIN-lab/vogue_lab3/vogue_lab3$ ls
__init__.py  items.py  middlewares.py  pipelines.py  settings.py  spiders
```

Posteriorment hem d'entrar a la carpeta del rastrejador, i des d'allà fer un `scrapy genspider` per crear la nostra *spider*, seguit del nom que li volem posar al rastrejador i la web general que es vol rastrejar (en el nostre cas `vogue.es`)



```
marc@Marc:~/Escritorio/Uni/REIN/3REIN-lab/vogue_lab3/vogue_lab3$ ls
__init__.py  items.py  middlewares.py  pipelines.py  settings.py  spiders
(myenv) marc@Marc:~/Escritorio/Uni/REIN/3REIN-lab/vogue_lab3/vogue_lab3$ scrapy genspider VogueInspector vogue.es
Created spider 'VogueInspector' using template 'basic' in module:
  vogue_lab3.spiders.VogueInspector
```

Si entrem a la carpeta de *spiders*, podem comprovar que s'ha creat correctament:



```
(myenv) marc@Marc:~/Escritorio/Uni/REIN/3REIN-lab/vogue_lab3/vogue_lab3$ cd spiders/
(myenv) marc@Marc:~/Escritorio/Uni/REIN/3REIN-lab/vogue_lab3/vogue_lab3$ ls
__init__.py  __pycache__  VogueInspector.py
```

Al fitxer del rastrejador creat de python (`VogueInspector.py`), podem veure les adreces donades inicialment i hem modificat el start url per el nostre article de vogue de galerias les tendències de tardor:

```

# -*- coding: utf-8 -*-
import scrapy

class VogueInspectorSpider(scrapy.Spider):
    name = 'VogueInspector'
    allowed_domains = ['vogue.es']
    start_urls = ['https://www.vogue.es/galerias/tendencias-mas-arriesgadas-otono-invierno-2023-2024?utm_source=pocket-newtab-es-es']

```

Codi del rastrejador:

Una vegada tenim l'esquelet del nostre rastrejador complet, podem començar a crear el codi per extreure la informació que ens interessi:

Per començar hem fet una comprovació en la funció base de `parse` per veure si hem pogut identificar correctament cada camp d'informació que volem:

```

1 # -*- coding: utf-8 -*-
2 import scrapy
3
4
5 class VogueInspectorSpider(scrapy.Spider):
6     name = 'VogueInspector'
7     allowed_domains = ['vogue.es']
8     start_urls = ['https://www.vogue.es/galerias/tendencias-mas-arriesgadas-otono-invierno-2023-2024?utm_source=pocket-newtab-es-es']
9
10    def parse(self, response):
11        for vogue in response.css('div[class="GallerySlideFigCaption-d0eyTg_dJsBXF"]'):
12            # Crear un diccionario para almacenar los datos
13            doc = {}
14
15            # Seleccionar el elemento h2 dentro de vogue
16            data = vogue.css('h2.tUElRD')
17
18            # Extraer el titulo desde el span dentro del h2 y almacenarlo en 'title'
19            doc['title'] = data.css('span.GallerySlideCaptionHedText-iqj0mm.jwPuvZ::text').extract_first()
20
21            # Extraer el texto de todos los elementos p y almacenarlo en una lista en 'article'
22            doc['article'] = vogue.css('p::text').extract()
23
24            # Extraer las URL de los elementos a dentro de los elementos p y almacenarlas en una lista en
25            'url'
26            doc['url'] = vogue.css('p a::attr(href)').extract()
27
28            # Procesar cada URL individualmente
29            yield doc

```

Si ho executem i ho guardem en un .json ens retorna:

```

[REIN/LAB3/vogue_lab3/vogue_lab3/spiders]$ scrapy crawl VogueInspector -o prova1.json
2023-11-08 11:50:49 [scrapy.utils.log] INFO: Scrapy 2.1.0 started (bot: vogue_lab3)
2023-11-08 11:50:49 [scrapy.utils.log] INFO: Versions: lxml 4.5.1.0, libxml2 2.9.10, cssselect 1.1.0, parsel 1.6.0, w3lib 1.22.0, Twisted 21.7.0, Python 3.8.10 (default, May 26 2023, 14:05:08) - [GCC 9.4.0], pyOpenSSL 19.0.0 (OpenSSL 1.1.1f 31 Mar 2020), cryptography 2.8, Platform Linux-5.15.0-82-generic-x86_64-with-glibc2.29
2023-11-08 11:50:49 [scrapy.utils.log] DEBUG: Using reactor: twisted.internet.epollreactor.EPollReactor
2023-11-08 11:50:49 [scrapy.crawler] INFO: Overridden settings:

```

JSON Path	Dades sense processar	Capçaleres
0:	Jesa Copia Redueix-ho tot Amplia-ho tot	Filtre JSON
0:	title: "Con sujetador a la vista" ▾ article: <ul style="list-style-type: none"> 0: "A medio camino entre el " 1: " rapero y la tendencia del momento, encontramos la ropa interior a la vista. Gracias al avance del feminismo, convertir las bragas o el sujetador protagonistas es ya una decisión estilística que muchas amamos. Y " 2: " también." ▾ url: <ul style="list-style-type: none"> 0: "https://www.vogue.es/moda/nodapedia/personajes/julia-fax/955" 	
1:	title: "Vestidos con 'memes'" ▾ article: <ul style="list-style-type: none"> 0: "El posmodernismo ha llegado para quedarse a la moda. Y ahora es habitual que los " <ul style="list-style-type: none"> 1: ", como el del tabloide estadounidense " 2: ". El feísimo pop ahora se ha vuelto deseable." url: [] 2: <ul style="list-style-type: none"> title: "Sandalias con pelo (y colores neón)" article: <ul style="list-style-type: none"> 0: "Otra de las tendencias más inesperadas de esta temporada son las sandalias de tipo peluche, que ya hemos visto en las pasarelas de " <ul style="list-style-type: none"> 1: ", y que están más " 2: " que nunca. Y si se tiñen de colores neón, mejor que mejor." url: <ul style="list-style-type: none"> 0: "https://www.vogue.es/moda/articulos/sandalias-pelo-tendencia-otono-invierno-2023-2024" 	
3:		

I podem comprovar de forma visual, entrant a la web i comparant les dades extretes amb l'article que surt inicialment:



1 / 15

<

Con sujetador a la vista

A medio camino entre el *look* rapero y la tendencia del momento, encontramos la ropa interior a la vista. Gracias al avance del feminismo, convertir las bragas o el sujetador en protagonistas es ya una decisión estilística que muchas amamos. Y **Julia Fox** también.

```
▼ 0:
  title: "Con sujetador a la vista"
  ▼ article:
    0: "A medio camino entre el "
    ▼ 1:
      " rapero y la tendencia del momento, encontramos la ropa interior a la vista. Gracias al avance del feminismo, convertir las bragas o el sujetador en protagonistas es ya una decisión estilística que muchas amamos. Y "
    2:
      " también."
  ▼ url:
    ▼ 0:
      "https://www.vogue.es/moda/modapedia/personajes/julia-fox/955"
```

Extensió del codi rastrejador:

Ara que tenim la base del rastrejador, hem de fer que trobats en els urls de cada article, entri en aquests i extregui la informació d'aquests sub-urls.

Per després poder fer un índex enllaçant-ho amb l'*Elasticsearch* i a partir d'aquest poder fer consultes amb el *SearchIndex.py* modificat segons les nostres necessitats. A partir del parse basic, creem el parse detail que entrarà en cada url de cada article de vogue que trobi.

```
9  def parse(self, response):
10     for vogue in response.css('div[class="GallerySlideFigCaption-dOeyTg_dJSBXf"]'):
11         # Crear un diccionario para almacenar los datos
12         doc = {}
13
14         # Seleccionar el elemento h2 dentro de vogue
15         data = vogue.css('h2.tUEtRd')
16
17         # Extraer el título desde el span dentro del h2 y almacenarlo en 'title'
18         doc['title'] = data.css('span.GallerySlideCaptionHedText-iqjOMM_jwPuvZ::text').extract_first()
19
20         # Extraer el texto de todos los elementos p y almacenarlo en una lista en 'article'
21         doc['article'] = vogue.css('p::text').extract()
22
23         # Extraer las URL de los elementos a dentro de los elementos p y almacenarlas en una lista en 'url'
24         doc['url'] = vogue.css('p a::attr(href)').extract()
25
26         # Procesar cada URL individualmente
27         for url in doc['url']:
28             yield scrapy.Request(url, callback=self.parse_detail, meta={'title': doc['title'], 'article': doc['article']})
29
30     def parse_detail(self, response):
31         # Crear un diccionario para almacenar datos detallados
32         detail = {}
33
34         # Obtener el título desde los metadatos y almacenarlo en 'TITLE'
35         detail['TITLE'] = response.meta['title']
36
37         # Obtener el artículo desde los metadatos y almacenarlo en 'ARTICLE'
38         detail['ARTICLE'] = response.meta['article']
39
40         # Obtener la URL de la página actual y almacenarla en 'URL'
41         detail['URL'] = response.url
42
43         # Extraer el título de la página y almacenarlo en 'URL_TITLE'
44         detail['URL_TITLE'] = ''.join(response.css('h1.BaseWrap-sc-gjqpdd.BaseText-ewhhUZ.SplitScreenContentHeaderHed-lcUSui_tUEtRd_bwDymH_dFeltga::text').extract())
45
46         # Extraer el texto dentro de los elementos p y concatenarlo en 'DESCRIPTION'
47         detail['DESCRIPTION'] = ''.join(response.css('p *::text').extract())
48
49         # Devolver el diccionario 'detail' como resultado del análisis de la página actual
50         yield detail
```

Ho executem fent un *scrapy crawl* del nostre rastrejador i guardant el que ens retorna en un .json.

```
(myenv) marc@Marc:~/Escritorio/Uni/REIN/5REIN-Lab/vogue_lab3/vogue_lab3/spiders$ scrapy crawl VogueInspector -o vogue.json
2023-10-24 21:43:32 [scrapy.utils.log] INFO: Scrapy 2.11.0 started (bot: vogue_lab3)
2023-10-24 21:43:32 [scrapy.utils.log] INFO: Versions: lxml 4.9.3.0, libxml2 2.10.3, cssselect 1.2.0, parsel 1.8.1, w3lib 2.1.2,
Twisted 22.10.0, Python 3.10.12 (main, Jun 11 2023, 05:26:28) [GCC 11.4.0], pyOpenSSL 23.2.0 (OpenSSL 3.1.4 24 Oct 2023), cryptography 41.0.5, Platform Linux-6.2.0-35-generic-x86_64-with-glibc2.35
2023-10-24 21:43:32 [scrapy.addons] INFO: Enabled addons:
[]
2023-10-24 21:43:32 [asyncio] DEBUG: Using selector: EpollSelector
2023-10-24 21:43:32 [scrapy.utils.log] DEBUG: Using reactor: twisted.internet.asyncioreactor.AsyncioSelectorReactor
2023-10-24 21:43:32 [scrapy.utils.log] DEBUG: Using asyncio event loop: asyncio.unix_events._UnixSelectorEventLoop
2023-10-24 21:43:32 [scrapy.extensions.telnet] INFO: Telnet Password: F41917abb9cc61a9
2023-10-24 21:43:32 [scrapy.middleware] INFO: Enabled extensions:
['scrapy.extensions.corestats.CoreStats',
'scrapy.extensions.telnet.TelnetConsole',
'scrapy.extensions.memusage.MemoryUsage',
'scrapy.extensions.feedexport.FeedExporter',
'scrapy.extensions.logstats.LogStats']
2023-10-24 21:43:32 [scrapy.crawler] INFO: Overridden settings:
{'BOT_NAME': 'vogue_lab3'},

{
'finish_time': datetime.datetime(2023, 10, 24, 19, 43, 33, 919590, tzinfo=datetime.timezone.utc),
'httpcompression/response_bytes': 4639244,
'httpcompression/response_count': 8,
'item_scraped_count': 7,
'log_count/DEBUG': 21,
'log_count/INFO': 11,
'memusage/max': 61272064,
'memusage/startup': 61272964,
'request_depth_max': 1,
'response_received_count': 9,
'robotstxt/request_count': 1,
'robotstxt/response_count': 1,
'robotstxt/response_status_count/200': 1,
'scheduler/dequeued': 8,
'scheduler/dequeued/memory': 8,
'scheduler/enqueued': 8,
'scheduler/enqueued/memory': 8,
'start_time': datetime.datetime(2023, 10, 24, 19, 43, 32, 900281, tzinfo=datetime.timezone.utc)}
marcperezg:33 [scrapy.core.engine] INFO: Spider closed (finished)
```

El .json que ens retornarà sera així:

JSON Dades sense processar Capçaleres

Desa Copia Redueix-ho tot Amplia-ho tot | Filtre JSON

```
{
  "0": {
    "TITLE": null,
    "ARTICLE": [
      {
        "0": {
          "URL": "https://www.vogue.es/moda/articulos/simone-ashley-vestido-conner-ives",
          "URL_TITLE": "Simone Ashley tiene nueva marca favorita (y sabemos cuál es)",
          "DESCRIPTION": "To revisit this article, visit My Profile, then View saved stories To revisit this article, visit My Profile, then View saved stories Por Emily Chan Simone Ashley se ha ganado una legión de seguidores por su interpretación de la ardiente Kate Sharma en la segunda temporada de Los Bridgerton, y sus looks también se están grabando rápidamente en el público fiel. Tras su debut en la Met Gala vestida como la última de una larga lista de estrellas que brindan su apoyo al diseñador londinense Conner Ives. El miércoles por la noche, para un cóctel organizado por el British Fashion Council, Vogue , Ashley optó por un vestido amarillo bordado de la colección de otoño/invierno 2022 de Ives , confecionado a partir de mantones de manila reciclados y con detalles de flecos . Fue una de sus primeras fans (la cantante contrató a Ives cuando aún estudiaba en la Central Saint Martins para que formara parte de su equipo de diseño en Fenty Maison) y Kylie Jenner también es fan. La firma diseños que se convierten en objeto de deseo. Y que ahora amamos."
        }
      }
    ],
    "1": {
      "TITLE": "Todes con falda",
      "ARTICLE": [
        {
          "0": {
            "URL": "https://www.vogue.es/articulos/david-beckham-sarong-metresexual-moda-futbol-masculino",
            "URL_TITLE": "El día que David Beckham posó con un sarong, o cómo el matrimonio de Victoria revolucionó la historia del fútbol masculino",
            "DESCRIPTION": "To revisit this article, visit My Profile, then View saved stories To revisit this article, visit My Profile, then View saved stories Por Alexandra Lores Hubo un tiempo en el que había rumores de que David Beckham iba a ser el próximo Beckham. En 2023, esa separación se hacía más y más imperceptible. Las marcas que llegan a un mayor sector de la población aún siguen haciendo esta diferencia, pero las emergentes –más nicho y una persona indefinida. Una realidad que no debemos dar por hecho, como ninguna otra que implique un avance social; máxime, después de recordar el episodio del sarong , protagonizado por David"
          }
        }
      ]
    }
  }
}
```

Ara per poder enllaçar-ho directament amb l'*elastic search* hem de canviar varies coses dels fitxers del rastrejador:

En el fitxer *settings.py* descomentar les següents línies de codi:

```
# Configure item pipelines
# See https://docs.scrapy.org/en/latest/topics/item-pipeline.html
ITEM_PIPELINES = {
    'vogue_lab3.pipelines.VogueLab3ElasticPipeline': 300,
}
```

En el fitxer de *pipelines.py* canviar el nom per al necessari per ElasticSearch (*VogueLab3Elasticpipeline*) i copiar el codi donat per les *pipelines*:

```
14 # Define your item pipelines here
15 #
16 # Don't forget to add your pipeline to the ITEM_PIPELINES setting
17 # See: http://doc.scrapy.org/en/latest/topics/item-pipeline.html
18
19 from elasticsearch import Elasticsearch
20
21 from elasticsearch.exceptions import NotFoundError
22 from elasticsearch_dsl import Index, analyzer, tokenizer
23
24 class VogueLab3ElasticPipeline(object):
25
26     collection_name = 'scrapy'
27
28     def __init__(self):
29         self.elastic_uri = 'http://localhost:9200/'
30         self.elastic_db = 'scrapy'
31
32     def open_spider(self, spider):
33
34         self.client = Elasticsearch()
35         try:
36             # Drop index if it exists
37             ind = Index(self.elastic_db, using=self.client)
38             ind.delete()
39         except NotFoundError:
40             pass
41         # then create it
42         ind.create()
43         ind.close()
44         # Configure tokenizer
45         my_analyzer = analyzer('default', type='custom',
46                                tokenizer=tokenizer('standard'),
47                                filter=[lowercase, 'asciifolding'])
48         ind.analyzer(my_analyzer)
49         ind.save()
50         ind.open()
51
52     def close_spider(self, spider):
53         self.client.close()
54
55     def process_item(self, item, spider):
56         self.client.index(index=self.elastic_db, doc_type='TFG', body=dict(item))
57
58         return item
```

Finalment, executar-ho fent un *scrapy crawl* del rastrejador sense un .json a on guardar el resultat:

```
[REIN/LAB3/vogue_lab3]$ scrapy crawl VogueInspector
[2023-11-08 11:41:26 [scrapy.utils.log] INFO: Scrapy 2.1.0 started (bot: vogue_lab3)
[2023-11-08 11:41:26 [scrapy.utils.log] INFO: Versions: lxml 4.5.1.0, libxml2 2.9.10, cssselect 1.1.0, parsel 1.6.0
, w3lib 1.22.0, Twisted 21.7.0, Python 3.8.10 (default, May 26 2023, 14:05:08) - [GCC 9.4.0], pyOpenSSL 19.0.0 (OpenSSL 1.1.1f 31 Mar 2020), cryptography 2.8, Platform Linux-5.15.0-82-generic-x86_64-with-glibc2.29
[2023-11-08 11:41:26 [scrapy.utils.log] DEBUG: Using reactor: twisted.internet.epollreactor.EPollReactor
```

Ara ja s'haurpa creat l'índex des del nostre rastrejador, i podem fer consultes modificant el fitxer *SearchIndex.py* pels nostres valors extrets:

```
if __name__ == '__main__':
    parser = argparse.ArgumentParser()
    parser.add_argument('--index', default=None, required=True, help='Index to search')
    parser.add_argument('--query', default=None, nargs=argparse.REMAINDER, help='Lucene query')

    args = parser.parse_args()

    index = args.index
    if args.query:
        query = ' '.join(args.query)

    try:
        client = Elasticsearch()
        s = Search(using=client, index=index)

        q = Q('query_string', query=query)
        s = s.query(q)
        response = s[0:10].execute()
        #print("response:", response)

        for r in response: # display the information that has been indexed for each item
            print('TITLE= %s' % r['TITLE'])
            print('ARTICLE= %s' % r['ARTICLE'])
            print('URL:= %s' % r['URL'])
            print('URL:TITLE= %s' % r['URL_TITLE'])
            print('DESCRIPTION_= %s' % r['DESCRIPTION'])
            print('-----')

            print ('%d Documents'% response.hits.total.value)
    except NotFoundError:
        print('Index %s does not exist' % index)
```

Comprovacions:

Per poder comprovar que hem creat correctament l'index a base del nostre rastrejador, podem executar el **Searchindex** modificat per fer consultes amb diferents termes i veure quins documents ens retorna i si aquests son correctes:

Buscant en TITOL, si posem una frase buida ens retorna zero documents:

```
/REIN/LAB3/vogue_lab3$ python3 SearchIndex_vogue.py --index scrapy --query TITLE  
0 Documents
```

Buscant en TITOL, si posem un títol en específic, ens retorna el document amb el títol igual:

```
/REIN/LAB3/vogue_lab3$ python3 SearchIndex_vogue.py --index scrapy --query TITLE:Todes con falda  
TITLE= Todes con falda  
ARTICLE= ['Los avances sociales también han permitido que los hombres se atrevan con las faldas sin miedo al qué dirán. Antes que ellos, ', ' de una tendencia que ahora cada vez se visibiliza (y normaliza) más.'][...]  
URL:= https://www.vogue.es/articulos/david-beckham-sarong-metosexual-moda-futbol-masculino  
URL:TITLE= El día que David Beckham posó con un sarong, o cómo el marido de Victoria revolucionó la historia del fútbol masculino  
DESCRIPTION= To revisit this article, visit My Profile, then View saved stories To revisit this article, visit My Profile, then View saved stories Por Alexandra Lores Hubo un tiempo en el que había ropa de mujer y ropa de hombre . En 2023, esa separación se va haciendo más y más imperceptible. Las marcas que llegan a un mayor sector de la población aún siguen haciendo esta diferencia, pero las emergentes -más nicho y vanguardistas- disienten para una persona indefinida. Una realidad que no debemos dar por hecho, como ninguna otra que implique un avance social; máxime, después de recordar el episodio del sarong , protagonizado por David Beckham . La mayoría conocimos el episodio después de ver la serie documental Beckham , en donde el futbolista relata con todo lujo de detalles sus comienzos en el fútbol. Unos comienzos humildes, marcados por un padre obsesionado con el Manchester United, que germinaron cuando David aún no había cumplido la veintena. Una historia personal, que se diferenciaba de tantas otras, por el carácter del futbolista británico; tendente a la fama y a gustarse el sueldo en cuestión de horas. Como dice Bad Bunny : "porque el dinero se hizo pa' gastarlo". Así las cosas, cuando Beckham comenzó a despuntar, se dio de bruscos con el rechazo frontal de su entrenador. Sir Alex Ferguson se encontraba más cercano a otro tipo de futbolista; el único que se estilaba por aquel entonces: tradicionalmente masculino y centrado en el deporte . Y el técnico consideraba que el gusto del joven por la moda y los coches de lujo lo alejaban de su trabajo. Su periplo como lo que ahora llamaríamos insider o precriptor de tendencias comenzó colaborando con Adidas , pero pronto se acercó a otras marcas como Emporio Armani , protagonizando una campaña para la firma. Sus estilismos de entonces comprendían desde trajes de cuero de Gucci , a juego con Victoria , hasta chándales blancos, bandanas, pantalones muy anchos y cruces gigantes. Looks que lo convirtieron en "el mayor metosexual de Gran Bretaña" . Según la Real Academia Española , este neologismo creado en 1994 por el periodista británico Mark Simpson , hace referencia a un hombre "especialmente heterosexual" , que se preocupa en extremo de su apariencia y dedica mucho tiempo y dinero a sus cuidados físicos ". El periodista cultural Abraham Rivera recuerda que escuchó el término, por primera vez, a comienzos del nuevo milenio. "Lo asociaba a un chico que probablemente fuera heterosexual y que iba muy arreglado, con un aspecto similar al que podían llevar los gays en ese momento ; un poco más moderno de lo normal, con una camiseta de tirantes, un pantalón ancho , un corte de pelo más especial... Era una palabra que se asociaba a la gente que cuidaba mucho su imagen, quizás en exceso" , aclara. Por Audrey Noble  
Por Silvia Suárez Por Mónica Heras Hay en ambas definiciones del término un uso de palabras como 'extremo' o 'exceso'. Porque los hombres que era
```

Buscant en ARTICLE, ens retorna el document on té aquests termes en aquest apartat:

```
/REIN/LAB3/vogue_lab3$ python3 SearchIndex_vogue.py --index scrapy --query ARTICLE:Los avances sociales  
TITLE= Todes con falda  
ARTICLE= ['Los avances sociales también han permitido que los hombres se atrevan con las faldas sin miedo al qué dirán. Antes que ellos, ', ' de una tendencia que ahora cada vez se visibiliza (y normaliza) más.'][...]  
URL:= https://www.vogue.es/articulos/david-beckham-sarong-metosexual-moda-futbol-masculino  
URL:TITLE= El día que David Beckham posó con un sarong, o cómo el marido de Victoria revolucionó la historia del fútbol masculino  
DESCRIPTION= To revisit this article, visit My Profile, then View saved stories To revisit this article, visit My Profile, then View saved stories Por Alexandra Lores Hubo un tiempo en el que había ropa de mujer y ropa de hombre . En 2023, esa separación se va haciendo más y más imperceptible. Las marcas que llegan a un mayor sector de la población aún siguen haciendo esta diferencia, pero las emergentes -más nicho y vanguardistas- disienten para una persona indefinida. Una realidad que no debemos dar por hecho, como ninguna otra que implique un avance social; máxime, después de recordar el episodio del sarong , protagonizado por David Beckham . La mayoría conocimos el episodio después de ver la serie documental Beckham , en donde el futbolista relata con todo lujo de detalles sus comienzos en el fútbol. Unos comienzos humildes, marcados por un padre obsesionado con el Manchester United, que germinaron cuando David aún no había cumplido la veintena. Una historia personal, que se diferenciaba de tantas otras, por el carácter del futbolista británico; tendente a la fama y a gustarse el sueldo en cuestión de horas. Como dice Bad Bunny : "porque el dinero se hizo pa' gastarlo". Así las cosas, cuando Beckham comenzó a despuntar, se dio de bruscos con el rechazo frontal de su entrenador. Sir Alex Ferguson se encontraba más cercano a otro tipo de futbolista; el único que se estilaba por aquel entonces: tradicionalmente masculino y centrado en el deporte . Y el técnico consideraba que el gusto del joven por la moda y los coches de lujo lo alejaban de su trabajo. Su periplo como lo que ahora llamaríamos insider o preceptor de tendencias comenzó colaborando con Adidas , pero pronto se acercó a otras marcas como Emporio Armani , protagonizando una campaña para la firma. Sus estilismos de entonces comprendían desde trajes de cuero de Gucci , a juego con Victoria , hasta chándales blancos, bandanas, pantalones muy anchos y cruces gigantes. Looks que lo convirtieron en "el mayor metosexual de Gran Bretaña" . Según la Real Academia Española , este neologismo creado en 1994 por el periodista británico Mark Simpson , hace referencia a un hombre "especialmente heterosexual" , que se preocupa en extremo de su apariencia y dedica mucho tiempo y dinero a sus cuidados físicos ". El periodista cultural Abraham Rivera recuerda que escuchó
```

Buscant en URL, si posem un terme en específic ens retorna el document on dins del seu url té aquest terme:

```
/REIN/LAB3/vogue_lab3$ python3 SearchIndex_vogue.py --index scrapy --query URL:upcycling  
TITLE= None  
ARTICLE= ['Los avances sociales están intimamente vinculados a las nuevas tendencias imperantes. Una de ellas, el ', ' , que consiste en crear nuevas piezas a partir de telas de otras prendas antiguas, ha dejado de ser una anécdota y se ha convertido en el emblema de enseñas como ', ' , que firma diseños que se convierten en objeto de deseo. Y que ahora ananoso.'][...]  
URL:= https://www.vogue.es/tags/upcycling  
URL:TITLE=   
DESCRIPTION= To revisit this article, visit My Profile, then View saved stories To revisit this article, visit My Profile, then View saved stories Por Alba Correa Por Mayte Salido Por Nuria Lluis Por Ana Gándara Por Nuria Lluis Por Hamish Bowles Por Emily Chan Por Dilvy Bala Por Alba Correa Por Patricia Moreno Por Julia Hobbs Por Sara Hernando Por Liam Freeman Por Nuala Phillips Por Garbine Contente Por Marina Valera Por Nuala Phillips Por Emily Chan Por Julia Hobbs Por Isabel Puentes Por Eva Blanco Medina Por Eva Blanco Medina Por Begona Gómez Urzaliz Por Nuala Phillips Más en Vogue Condé Nast España e 2023 CONDENET IBERICA S.L. Todos los derechos reservados.  
-----  
1 Documents
```

Buscant en URL_TITLE, si posem un url buit ens retorna zero documents:

```
/REIN/LAB3/vogue_lab3$ python3 SearchIndex_vogue.py --index scrapy --query URL_TITLE  
0 Documents
```

Buscant en URL_TITLE, uns termes en específic, ens retorna el document que conté aquests termes dins de url_title:

```
REIN/LAB3/vogue_lab3$ python3 SearchIndex_vogue.py --index scrapy --query URL_TITLE: Sandalias con pelo: el calzado más excéntrico de 2023 va a estar por todas partes
TITLE= Sandalias con pelo (y colores neón)
ARTICLE= ['Otra de las tendencias más inesperadas de esta temporada son las sandalias de tipo peluche, que ya hemos visto en las pasarelas de ', ', y que están más ', ' que nunca. Y si se tiñen de colores neón, mejor que mejor.']
URL:= https://www.vogue.es/moda/articulos/sandalias-pelo-tendencia-otono-invierno-2023-2024
URLTITLE= Sandalias con pelo: el calzado más excéntrico de 2023 va a estar por todas partes
DESCRIPTION_= To revisit this article, visit My Profile, then View saved stories To revisit this article, visit My Profile, then View saved stories Por Nuria Luis Ni frío ni calor. No son medias tintas, ni una propuesta que deje indiferente: es la paradoja del calzado que promete ser el más visto de 2023. Con una suela de pelo, las sandalias tipo peluche ya se han colado en algunas de las colecciones de la temporada. Vayamos por partes: por un lado, Gucci incluyó una suerte de mules con tacón kitten y estribo Horsebit que incorporaban este detalle en color negro. Que puedan ser un éxito de ventas no lo dudamos: ahí tuvimos los zuecos de pelo con los que debutó Alessandro Michele al frente de la casa italiana, que convirtió en un best-seller de la firma. En Londres, Burberry fue otra de las marcas que ha incorporado este calzado en su desfile. En su respectivo debut , Daniel Lee , que tiene bien calada la debilidad de la moda por los diseños exagerados, incluyó bolsos y sandalias con colas de pelo falso. Las sandalias de pelo recogen el testigo de temporadas anteriores. En su colección de primavera verano 2023, Jacquemus también apostó por un calzado 'furry' con pelo en color amarillo que ha compartido hace unos días en Instagram, combinados con unos vaqueros rectos y su bolso Bambole long lancet. La primavera anterior fue Altuzarra quien optó por un zapato abierto con plantilla de pelo falso y tiras de cuerda para combinar con sus looks bohemio s. Por Audrey Noble Por Sílvia Suárez Por Mónica Heras This content can also be viewed on the site it originates from. Por Tatiana Ojea Por M arta Munsuri Por Patricia Moreno Por Estrella Albendea Más en Vogue Condé Nast España © 2023 CONDENET IBERICA S.L. Todos los derechos reservados.
```

Buscant en DESCRIPTION: un terme, ens retorna el document que conté aquest terme en el seu apartat de la descripció:

```
REIN/LAB3/vogue_lab3$ python3 SearchIndex_vogue.py --index scrapy --query DESCRIPTION:stories
TITLE= None
ARTICLE= ['Los avances sociales están intimamente vinculados a las nuevas tendencias imperantes. Una de ellas, el ', ', que consiste en crear nuevas piezas a partir de retales de otras prendas antigüas, ha dejado de ser una anécdota y se ha convertido en el emblema de enseñas como ', ', que firma diseños que se convierten en objeto de deseo. Y que ahora amamos.']
URL:= https://www.vogue.es/tags/upcycling
URLTITLE=
DESCRIPTION_= To revisit this article, visit My Profile, then View saved stories To revisit this article, visit My Profile, then View saved stories Por Alba Correa Por Mayte Salido Por Nuria Luis Por Ana Gándara Por Nuria Luis Por Hamish Bowles Por Emily Chan Por Divya Bala Por Alba Correa Por Patricia Moreno Por Julia Hobbs Por Sara Hernando Por Liam Freeman Por Nuala Phillips Por Garbiñe Contente Por Marina Valera Por Nuala Phillips Por Emily Chan Por Julia Hobbs Por Isabel Puentes Por Eva Blanco Medina Por Eva Blanco Medina Por Begoña Gómez Urzaiz Por Nuala Phillips Más en Vogue Condé Nast España © 2023 CONDENET IBERICA S.L. Todos los derechos reservados.
```

Problemes en la implementació:

Responent a “Expliqueu si us ha funcionat fàcilment o heu tingut alguns problemes en la implementació (si n’heu tingut, expliqueu breument quins) ”

Em tingut problemes amb la manera amb la que recuperem els diferents camps de la web, ja que per exemple quan estem recuperant el text (article) i trobem una url el codi deixava de guardar el article, simplement guardava l'url i parava. És per això que tenim com a resultat l'opció plantejada de codi, el qual soluciona aquest error i recupera correctament tot l'article. A part d'aquest inconvenient, no hem trobat major resistència més que anar inspeccionant la pàgina web de vogue per tal de saber com recuperar cada camp que ens interessa sense haver de posar un a un els XPATH dels articles que volem guardar.