

## ACCESO INTELIGENTE A LA INFORMACIÓN WHOOSH - EVALUABLE - GRUPOS 1.1 Y 1.2

Deseamos hacer un sistema de recuperación de información de una web con noticias de cine, para lo cual comenzaremos haciendo scraping sobre las noticias en la web:

<https://www.sensacine.com/noticias/>

SÓLO LAS CUATRO PRIMERAS PÁGINAS

Se desea hacer un buscador sobre distinta información relacionada con las noticias.

Buscamos construir un programa en Tkinter con un MENÚ con dos opciones:

- a) “Datos”, con tres opciones:
  - a. (3 puntos) “Cargar”, que cree esquema e índice en Whoosh que almacene los siguientes datos de cada noticia publicada: categoría, título, enlace a la noticia, breve descripción de la noticia (si la tiene) y fecha de la publicación. Si ya existe un índice creado, que muestre una ventana de confirmación para volverlo a crear. Muestre una ventana de mensajes informando del número de noticias almacenadas en el sistema.
  - b. (1 punto) “Listar”, que muestre una ventana con todas las noticias (título y fecha).
  - c. “Salir”, que cierre la ventana principal.
- b) “Buscar”, con seis opciones:
  - a. (1 punto) “Descripción”, que muestre una ventana con un entry que permita al usuario introducir una frase, y muestre en otra las diez noticias (categoría, título, enlace y fecha) más relevantes que contengan esa frase en la descripción de la noticia.
  - b. (1 punto) “Categoría y Título”, que muestre una ventana con un spinbox, un entry y un botón. En el spinbox se selecciona una de las categorías indexadas y en el entry se introducen una o varias palabras. Cuando se pulse el botón muestre en otra ventana todas las noticias (categoría, título, enlace y fecha) de la categoría seleccionada que contenga/n alguna/s de esa/s palabra/s en el título.
  - c. (1 punto) “Título o Descripción”, que muestre una ventana con dos entry y un botón. En uno se introduce una frase y en el otro una o varias palabras. Cuando se pulse el botón, se muestre otra ventana con todas las noticias (categoría, título, enlace y fecha) que contengan la frase en el título y todas las palabras en la descripción.
  - d. (1 punto) “Fecha”, que muestre una ventana con un entry que permita al usuario introducir un rango de fechas (en formato “5 de Noviembre de 2025 hasta 6 de Noviembre de 2025”), compruebe la corrección del formato, y muestre en otra ventana todas las noticias (categoría, título, enlace y fecha) de las noticias publicadas en ese rango de fechas.
  - e. (1 punto) “Eliminar por Descripción”, que muestre una ventana con un entry que permita al usuario introducir una o varias palabras. Muestre en otra ventana todas las noticias (título) que contienen todas las palabras en el título y su descripción esté vacía. A continuación muestre una ventana de mensaje pidiendo confirmación. Si se confirma, elimina todas las noticias listadas.
  - f. (1 punto) “Título y Fecha”, que muestre una ventana con dos entry y un botón. En un entry se introduce una frase y en el otro una fecha (formato

DDMMAAAA). Tras pulsar el botón, se comprueba el formato de la fecha, y se muestran las cinco noticias (categoría, título, enlace y fecha) más relevantes, teniendo en cuenta que el peso del campo fecha es el doble que el del campo título.

**IMPORTANTE:** Para acceder a las páginas de esta web podéis definir un acceso como browser, por ejemplo de la siguiente manera:

```
req=urllib.request.Request("https://www.sensacine.com/noticias/",  
headers={'User-Agent': 'Mozilla/5.0'})  
f = urllib.request.urlopen(req)  
s = BeautifulSoup(f, 'Lxml')
```