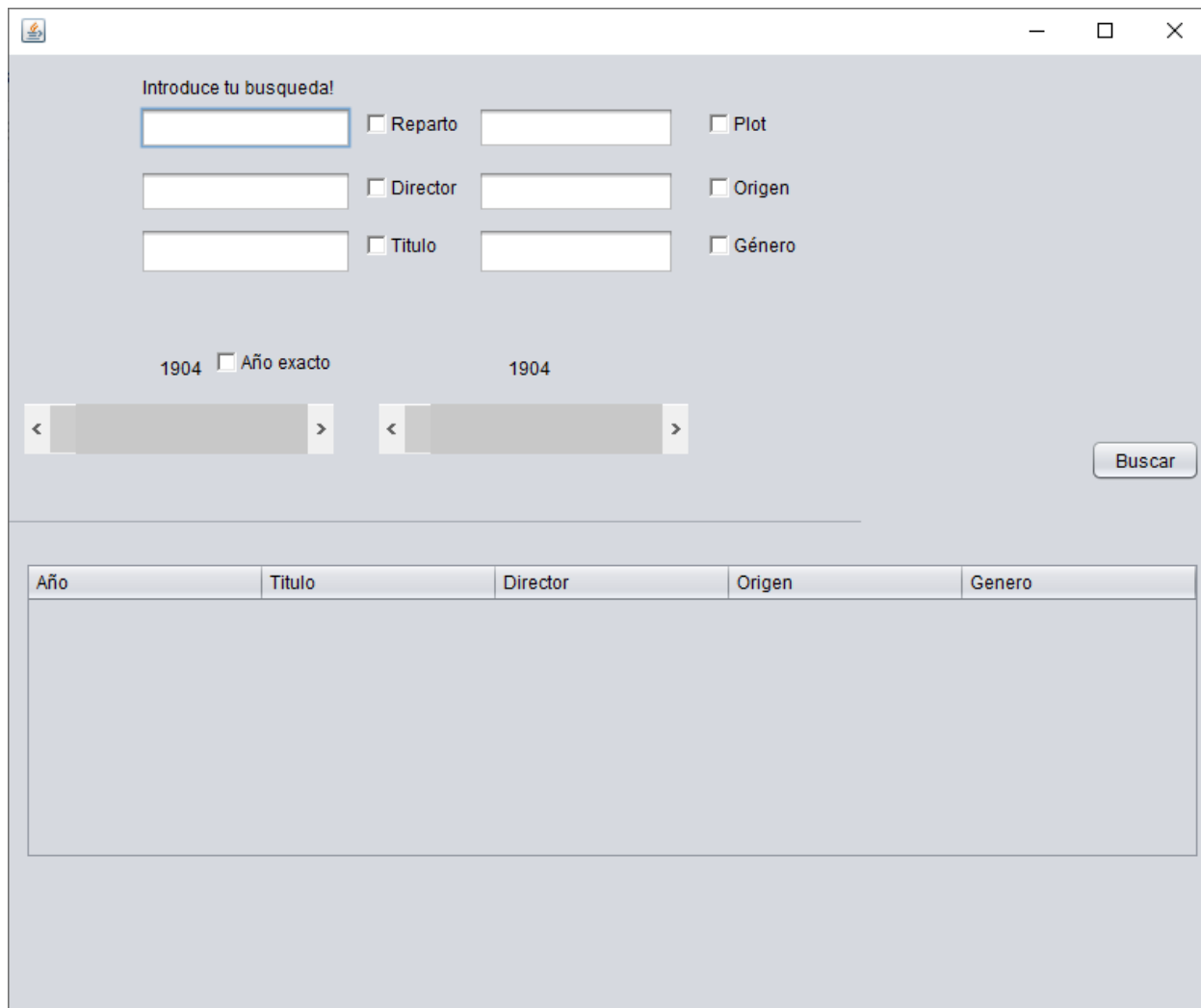


Buscador de Películas

Juan Manuel Herrera García

Recuperación de Información



The screenshot shows a window titled "Buscador de Películas" with a standard Windows-style title bar (minimize, maximize, close buttons). The main area has a light gray background. At the top, it says "Introduce tu búsqueda!". Below this, there are six input fields arranged in a 3x2 grid. Each field has a corresponding checkbox to its left: "Reparto", "Director", "Titulo" in the first column, and "Plot", "Origen", "Género" in the second column. Below the input fields, there are two date range selectors, each showing "1904" and a "Año exacto" checkbox. To the right of these is a "Buscar" button. At the bottom, there is a table with five columns: "Año", "Titulo", "Director", "Origen", and "Genero". The table body is currently empty.

| Año | Titulo | Director | Origen | Genero |
|-----|--------|----------|--------|--------|
|-----|--------|----------|--------|--------|

19/12/2021

Ingeniería Informática UGR

ÍNDICE

1. Introducción
2. Descripción del problema
3. Analizadores
4. Creación del índice
5. Buscadores
6. Interfaz Gráfica
7. Ejemplos

1. INTRODUCCIÓN

En primer lugar, el resultado que queremos conseguir es un buscador de películas, teniendo en cuenta diferentes parámetros de búsqueda, ya sea, director, reparto, año ...

Teniendo en cuenta que podemos seleccionar que obligatoriamente un campo aparezca, que sea opcional o que busque en todos los campos por igual.

Empezamos con un dataset, gracias a este, vamos a obtener la información necesaria de nuestras películas.

2. Descripción del problema

¿A qué problema nos enfrentamos?

Tenemos que pasar de tener un dataset de películas a la creación de un índice que nos sirva para buscar sobre diferentes campos

Para representar los datos y trabajar con ellos lo que he hecho ha sido crear una clase llamada videoclub. Antes de poder almacenar los datos, es necesario hacer uso de expresiones regulares para eliminar algunos errores en el csv, algunos como por ejemplo eliminar espacios en blanco o eliminar redundancias.

3. Analizadores

Una vez representado, procedemos a la creación del índice, para esto, tenemos que usar un analizador, que se encargará de separar las palabras con las que trabajaremos, en mi caso he usado el analizador por defecto.

Lo que hace este será separar las palabras por espacios, hay que tener en cuenta que si tenemos un analizador case sensitive, ya que más adelante en la búsqueda podremos tener problemas.

4. Creación del índice

Una vez hecho este proceso, tenemos que crear el índice, indicando qué campos queremos almacenar en el índice (título, reparto, origen...) asignando a cada uno de estos campos su tipo de atributo. Para el caso de la fecha, es necesario almacenarlo de dos formas diferentes, una para hacer uso más adelante de la búsqueda por rango y por una fecha concreta.

De esta manera, ya sabrá que tipo de información va a recibir y cómo compararlas con los índices. Toda esta tarea la realizamos gracias al IndexWriter de Lucene.

5. Buscadores

Una vez creado el índice, podemos comenzar con las búsquedas.

Para realizar las consultas, haremos uso de nuevo de Lucene, que nos aporta diferentes funciones de búsqueda, los usados en esta práctica son:

- Búsqueda de términos
- Búsqueda booleana
- Búsqueda por rangos
- Búsqueda por 'entero'

Con el primer tipo, podremos realizar una búsqueda por campos, sin tener en cuenta que además deberán aparecer obligatoriamente otros parámetros. Sin embargo, con el segundo se tendrá que tener en cuenta que algunos términos es obligatorio que aparezcan. Los dos últimos apartados hacen referencia a la búsqueda por fecha, usándolo para buscar dentro de un rango de fechas (1904-2017) o en una fecha concreta (ej. 1998).

6. Interfaz Gráfica

Aunque este documento no explica detalladamente el proceso de creación de la práctica, lo he enfocado algo más al uso del mismo. En la defensa podré explicar mejor cómo funciona cada parámetro y no será tan pesado.

En primer lugar, a la hora de iniciar el programa, nos aparecerá la siguiente ventana:

The screenshot shows a graphical user interface for a search application. It features a search bar at the top left, followed by a grid of checkboxes for search criteria: Reparto, Director, Título, Plot, Origen, and Género. Below this is a date range selector with a 'Año exacto' checkbox. To the right is a 'Buscar' button. At the bottom is a table with columns for Año, Título, Director, Origen, and Género. Numbered callouts identify these elements: 1 points to the search bar and criteria checkboxes; 2 points to the 'Año exacto' checkbox; 3 points to the date range input fields; 4 points to the 'Buscar' button; and 5 points to the results table.

| Año | Título | Director | Origen | Género |
|-----|--------|----------|--------|--------|
|-----|--------|----------|--------|--------|

1. Campos de búsqueda por texto

En este apartado, podremos introducir diferentes consultas, cómo se puede ver, al lado de cada campo hay un checkbox, esto indica que queremos que aparezca obligatoriamente.

ACLARACIÓN

En caso de que seleccionemos una opción, y además introducimos otros campos, lo que se hará es una búsqueda ‘MUST’ de aquellos campos que están rellenos y con el checkbox activado, los campos rellenos y con el checkbox **no** activado se hará una búsqueda ‘SHOULD’.

A la hora de que hagamos la búsqueda, hay un pequeño desliz que no he sido capaz de solucionar y es que el rango de fechas inicial está desde 1904-1904, por lo que en principio se tendrá que ampliar, hasta el máximo si queremos buscar todas las películas u otro rango que prefiramos.

2. Año exacto

| Año | Título | Director | Origen | Genero |
|------|---------------------------|-----------------|----------|---------------------|
| 1976 | Across the Great Divide | Stewart Raffill | American | family |
| 1976 | Alice, Sweet Alice | Alfred Sole | American | horror |
| 1976 | All the President's Men | Alan J. Pakula | American | drama, biographical |
| 1976 | Assault on Precinct 13 | John Carpenter | American | action |
| 1976 | At the Earth's Core | Kevin Connor | American | sci-fi |
| 1976 | Baby Blue Marine | John D. Hancock | American | drama |
| 1976 | The Bad News Bears | Michael Ritchie | American | comedy |
| 1976 | The Big Bus | James Frawley | American | comedy |
| 1976 | The Bingo Long Travell... | John Badham | American | comedy, sports |
| 1976 | Black Heat | Al Adamson | American | action |
| 1976 | Black Chamber | Geoffrey Clark | American | drama |

Con esta opción podemos realizar una búsqueda de una película en un año concreto por ejemplo 1976. Una vez seleccionado, solo tenemos que mover la barra izquierda, la derecha no nos servirá para seleccionar la fecha.

3. Barra de rango de fechas

En esta parte, podemos definir el intervalo de los años, no importa que el de la derecha esté por un año inferior al de la izquierda o viceversa, se usará el menor como mínimo y el mayor como máximo.

En el caso de la búsqueda general, no serán funcionales

4. Botón de buscar

Pues un botón de buscar, no hace magia, por debajo coordina el envío de los parámetros que hemos establecido y manda a actualizar la tabla (5) con los resultados

5. Tabla de resultados

En la tabla se mostrarán todos los resultados obtenidos, no he tenido en cuenta mostrar el reparto ni la URL para no tener demasiada información en pantalla.

6. Ejemplos

Voy a poner algunas capturas que pueden servir como guía.

Antes de nada, he intentado solucionar un problema que no deja de perseguirme, a la hora de introducir los nombres, si contienen espacios no los reconoce. Se que de por si Lucene los espacios los entenderá como tokens “palabra_palabra” es decir, como un único token, sin embargo, haciendo uso de la búsqueda booleana, lo que conseguía eran resultados un poco absurdos, quizás por la ponderación de los resultados. Aquí un ejemplo:

Introduce tu búsqueda!

☐ Reparto

☐ Plot

☒ Director

☐ Origen

☐ Título

☐ Género

1904

☐ Año exacto

2017

< >

< >

Buscar

| Año | Título | Director | Origen | Genero |
|------|----------------------|----------------|--------|---------|
| 2008 | Nagaram | Srinivas C. C | Telugu | action |
| 1960 | Naan Kanda Sorgam | C. Pullaiah | Tamil | unknown |
| 1962 | Indra En Selvam | C. Padmanabhan | Tamil | unknown |
| 1978 | Aval Appadithan | C. Rudhraiya | Tamil | unknown |
| 1995 | Murai Maman | Sundar C. | Tamil | unknown |
| 1995 | Ayudha Poojai | C. Sivakumar | Tamil | unknown |
| 1995 | Murai Mappillai | Sundar C. | Tamil | unknown |
| 1996 | Coimbatore Mappillai | C. Ranganathan | Tamil | unknown |
| 1996 | Ullathai Allitha | Sundar C. | Tamil | unknown |
| 1996 | Mettukudi | Sundar C. | Tamil | unknown |
| 1996 | Tata Rida | C. Ranganathan | Tamil | unknown |

Ejemplo de búsqueda donde es obligatorio que aparezca “Cristopher Nolan” y secundario “batman” sin embargo los resultados no tienen nada que ver con la consulta, es por esto por lo que he dejado que solo se busque por una palabra.

Ahora sí, voy con los ejemplos

Introduce tu búsqueda!

bale ☐ Reparto batman ☒ Plot

Nolan ☒ Director ☐ Origen

☐ Título ☐ Género

1904 ☐ Año exacto 2017

< > < >

Buscar

| Año | Título | Director | Origen | Genero |
|------|---------------------------|-------------------|----------|------------------------|
| 2008 | The Dark Knight | Christopher Nolan | British | superhero/action |
| 2005 | Batman Begins | Christopher Nolan | American | superhero |
| 2008 | The Dark Knight | Christopher Nolan | American | superhero |
| 2012 | Dark Knight Rises, The... | Christopher Nolan | American | crime drama, superhero |
| 2012 | The Dark Knight Rises | Christopher Nolan | British | unknown |

Ejemplo 1

Nolan y Batman obligatorio, Bale secundario.

Ejemplo 2

2001 y space, consulta general

Introduce tu búsqueda!

☐ Reparto space ☐ Plot

☐ Director ☐ Origen

2001 ☐ Título ☐ Género

1904 ☐ Año exacto 2017

< > < >

Buscar

| Año | Título | Director | Origen | Genero |
|------|---------------------------|--------------------------|-----------|-----------------------|
| 1968 | 2001: A Space Odyssey | Stanley Kubrick | American | science fiction |
| 1998 | 2001: Do Hazaar Ek | Raj N. Sippy | Bollywood | crime, drama, mystery |
| 2014 | Bodacious Space Pirat... | Tatsuo Sato | Japanese | adventure |
| 2012 | Space Sheriff Gavan: T... | Unknown | Japanese | tokusatsu |
| 1994 | Emmanuelle in Space | Unknown | American | science fiction |
| 1954 | Riders to the Stars | Richard Carlson | American | science fiction |
| 1966 | Way...Way Out | Gordon Douglas | American | comedy |
| 1970 | Space Amoeba | Honda, Ishirōshirō Ho... | Japanese | unknown |
| 1962 | Mouse Into Space | Gene Deitch | American | animated short |
| 1958 | Space Master X-7 | Edward Bernds | American | sci-fi |
| 2013 | Space: First in Space | Paul Berke | Russian | unknown |

Ejemplo 3

2001 y space solo
en 1986

Introduce tu búsqueda!

☐ Reparto ☐ Plot

☐ Director ☐ Origen

☐ Título ☐ Género

1986 ☒ Año exacto 1980

< > < >

Buscar

| Año | Título | Director | Origen | Genero |
|------|----------------------------|---------------------------|----------|------------------------|
| 1986 | SpaceCamp | Harry Winer | American | science fiction |
| 1986 | Howard the Duck | Willard Huyck | American | science fiction comedy |
| 1986 | Star Trek IV: The Voyag... | Leonard Nimoy | American | science fiction |
| 1986 | Little Shop of Horrors | Frank Oz | American | musical comedy |
| 1986 | Robotech: The Movie | Noboru Ishiguro, Carl ... | American | animation |
| 1986 | Robotech: The Movie | Noboru Ishiguro | Japanese | unknown |
| 1986 | Solarbabies | Alan Johnson | American | science fiction |
| 1986 | TerrorVision | Ted Nicolaou | American | comedy horror |
| 1986 | The Transformers: The... | Nelson Shin | American | science fiction |
| 1986 | Captain EO | Francis Ford Coppola | American | science fiction |
| 1986 | Vikings | Reinhold | Teal | unknown |

Ejemplo 4

2001 obligatorio
space secundario
action secundario
desde 1904-1988

Introduce tu búsqueda!

☐ Reparto ☐ Plot

☐ Director ☐ Origen

☒ Título ☐ Género

1904 ☐ Año exacto 1988

< > < >

Buscar

| Año | Título | Director | Origen | Genero |
|------|-----------------------|-----------------|----------|-----------------|
| 1968 | 2001: A Space Odyssey | Stanley Kubrick | American | science fiction |

Introduce tu búsqueda!

☐ Reparto ☐ Plot

☐ Director ☐ Origen

☐ Título ☐ Género

1904 ☐ Año exacto 2017

< > < >

Buscar

| Año | Título | Director | Origen | Genero |
|------|---------------------------|---------------------------|----------|---------------------------|
| 1904 | The Suburbanite | Wallace McCutcheon | American | comedy |
| 1905 | The Little Train Robbery | Edwin Stanton Porter | American | unknown |
| 1905 | The Night Before Chris... | Edwin Stanton Porter | American | unknown |
| 1906 | Dream of a Rarebit Fie... | Wallace McCutcheon a... | American | short |
| 1906 | From Leadville to Aspe... | Francis J. Marion and ... | American | short action/crime wes... |
| 1906 | Kathleen Mavourneen | Edwin S. Porter | American | short film |
| 1907 | Daniel Boone | Wallace McCutcheon a... | American | biographical |
| 1907 | How Brown Saw the B... | Unknown | American | comedy |
| 1907 | Laughing Gas | Edwin Stanton Porter | American | comedy |
| 1908 | The Adventures of Dollie | D. W. Griffith | American | drama |
| 1909 | The Black Viking | D. W. Griffith | American | drama |

Ejemplo 5

Todas las películas de 1904-2017

Ejemplo 6

Películas de 1974

Introduce tu búsqueda!

☐ Reparto ☐ Plot

☐ Director ☐ Origen

☐ Título ☐ Género

1974 ☒ Año exacto 2017

< > < >

Buscar

| Año | Título | Director | Origen | Genero |
|------|----------------------------|---------------------|----------|---------|
| 1974 | 99 and 44/100% Dead | John Frankenheimer | American | unknown |
| 1974 | Airport 1975 | Jack Smight | American | unknown |
| 1974 | Alice Doesn't Live Here... | Martin Scorsese | American | unknown |
| 1974 | All the Kind Strangers | Burt Kennedy | American | unknown |
| 1974 | And Then There Were ... | Peter Collinson | American | unknown |
| 1974 | The Arena | Steve Carver | American | unknown |
| 1974 | Bank Shot | Gower Champion | American | unknown |
| 1974 | Benji | Joe Camp | American | unknown |
| 1974 | Beyond the Door | Ovidio G. Assonitis | American | unknown |
| 1974 | Big Bad Mama | Steve Carver | American | unknown |
| 1974 | Billy Two Hats | Ted Kotcheff | American | unknown |